



Reliability and validity of a Tutorial Group Effectiveness Instrument

Veena S. Singaram, Cees P. M. Van Der Vleuten, Henk Van Berkel & Diana H. J. M. Dolmans

To cite this article: Veena S. Singaram, Cees P. M. Van Der Vleuten, Henk Van Berkel & Diana H. J. M. Dolmans (2010) Reliability and validity of a Tutorial Group Effectiveness Instrument, Medical Teacher, 32:3, e133-e137, DOI: [10.3109/01421590903480105](https://doi.org/10.3109/01421590903480105)

To link to this article: <https://doi.org/10.3109/01421590903480105>



Published online: 10 Mar 2010.



Submit your article to this journal [↗](#)



Article views: 2065



View related articles [↗](#)



Citing articles: 3 View citing articles [↗](#)

WEB PAPER

Reliability and validity of a Tutorial Group Effectiveness Instrument

VEENA S. SINGARAM¹, CEES P. M. VAN DER VLEUTEN², HENK VAN BERKEL² & DIANA H. J. M. DOLMANS²

¹University of KwaZulu-Natal, South Africa, ²University of Maastricht, The Netherlands

Abstract

Background: Tutorial group effectiveness is essential for the success of learning in problem-based learning (PBL). Less effective and dysfunctional groups compromise the quality of students learning in PBL.

Aims: This article aims to report on the reliability and validity of an instrument aimed at measuring tutorial group effectiveness in PBL.

Method: The items within the instrument are clustered around motivational and cognitive factors based on Slavin's theoretical framework. A confirmatory factor analysis (CFA) was carried out to estimate the validity of the instrument. Furthermore, generalizability studies were conducted and alpha coefficients were computed to determine the reliability and homogeneity of each factor.

Results: The CFA indicated that a three-factor model comprising 19 items showed a good fit with the data. Alpha coefficients per factor were high. The findings of the generalizability studies indicated that at least 9–10 student responses are needed in order to obtain reliable data at the tutorial group level.

Conclusion: The instrument validated in this study has the potential to provide faculty and students with diagnostic information and feedback about student behaviors that enhance and hinder tutorial group effectiveness.

Introduction

Group work lies in the heart of problem-based learning (PBL). Ensuring the effectiveness of the small-group tutorial is critical for the success of learning in a PBL program. Tutorial group interactions provide students with opportunities to give and receive explanations, to ask questions, and to discuss disagreements which are assumed to lead to a deep understanding of the subject matter (Visschers-Pleijers et al. 2005). Dolmans and Schmidt (2006) reported that studies dealing with the cognitive effects of PBL demonstrated that the activation of prior knowledge, causal reasoning, and cognitive conflicts lead to conceptual changes. In the long term, group work plays a vital role in developing medical professionalism and team work skills that are essential for effective multi-disciplinary health care teams (Singaram et al. 2008).

Group learning environments such as PBL have the promise of creating effective learning environments, but in reality dysfunctional groups also exist. One of the problems in tutorial group work is referred to as 'ritual' behavior, i.e. students pretend to be actively involved in the group work, whereas they are in fact not actually involved (Dolmans et al. 2005). In some groups, discussions are rather superficial instead of in-depth (Houlden 2001; De Grave et al. 2001, 2002). Quiet and dominant students were also found to hinder student learning (Hendry et al. 2003) as this leads to unbalanced discussions in the group (Virtanen et al. 1999).

Practice points

- The TGEI in this study appeared to reveal valid and reliable factor scores.
- The TGEI can be used to provide students and facilitators with feedback about the student's learning in the PBL group.
- Increased awareness of group functioning is an important aspect of feedback which can lead to an improvement of tutorial group effectiveness which will enhance student learning and academic performance.

Dolmans et al. (2005) in their review of research and debate on PBL highlighted the need for more theoretical based research to understand the factors and conditions under which tutorial group work in PBL is less and more effective.

Slavin (1996) distinguishes two major theoretical perspectives from which small-group PBL learning can be studied. The first perspective is a motivational one. This perspective emphasizes the importance of cohesiveness or team spirit. The second perspective is a cognitive one. A group provides opportunities to interact, discuss, argue and give explanations to each other, and to provide mutual feedback. These cognitive processes are assumed to positively influence student learning. Although groups can be motivating, some groups may have negative effects on students' motivation.

Correspondence: V. S. Singaram, Department of Medical Education, College of Health Sciences, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, PBX Congella 4013, South Africa. Tel: 27312604594; fax: 27312604492; email: singaram@ukzn.ac.za

e.g. when some students do not participate actively. The influence of these factors on tutorial group productivity or success was explored by Dolmans et al. (1998). Their study found a linear relationship between the tutorial group's success and several motivational (i.e. motivation and cohesion) and cognitive (i.e. elaboration and interaction) dimensions. Similar findings were reported by Carlo et al. (2003) who, in addition, noted that students' backgrounds and cultures influence motivation and the cognitive aspects of the small-group tutorial as well. Furthermore, Singaram et al. (2008) highlighted that although diverse students were positive about group learning, attention should be directed toward students who respond negatively to group work as dysfunctional groups have been noticed in practice. Thus, more attention needs to be directed to understanding the array of factors that influence group effectiveness so that dysfunctional groups can be timely diagnosed and appropriately managed.

The need for instruments to diagnose problems, measure the quality of tutorial group interactions, and find ways to improve the functioning of small group has also been noted by Visschers-Pleijers et al. (2005).

The aim of this article is to report on the development, reliability, and validity of an instrument aimed at measuring tutorial group effectiveness in a PBL curriculum.

Method

Setting

A 5-year integrated PBL curriculum replaced the traditional undergraduate medical curriculum in 2001, at the Nelson R. Mandela School of Medicine (NRMSM). PBL modules form a part of the first to third years of study. The student population

at NRMSM is socially and culturally diverse. The majority of the students have a first language other than English (approximately 13 different languages), whilst the language of instruction is English. Students are grouped, taking into account their socio-cultural backgrounds rather than being randomly assigned to PBL groups. The groupings are changed for every new PBL theme/unit. Each group of about 10 students meets twice a week with a facilitator to discuss a case. In the first 2 h session, learning issues are generated which need to be studied during self-study and then reported on in the second 2 h session. One of the students chairs the meeting and the role of chairperson is rotated among the students.

Instrument

A Tutorial Group Effectiveness Instrument (TGEI) was developed based on Slavin's theoretical framework (Slavin 1996). The items within the instrument were clustered around motivational and cognitive factors. One factor focused on the cognitive aspects and two factors focused on the motivational aspects. The instrument was pilot-tested with a group of randomly selected students with diverse backgrounds. This led to the rewording of three items and skipping of one item. Nineteen items underlying factors or aspects of group effectiveness were included, i.e. cognitive, motivational, and demotivational aspects (Table 1). Some items of the instrument were adapted from Dolmans et al. (1998). The students were asked to respond to each item in the instrument based on a 5-point Likert scale ranging from 1 – 'strongly disagree' to 5 – 'strongly agree'. Students were also asked to rate the overall productivity of the tutorial group on a scale from 1 to 5, i.e. 1 – insufficient, 2 – reasonable, 3 – sufficient, 4 – good, and 5 – excellent.

Table 1. Items and factors within the TGEI.

Items and factors	Mean (1–5)	SD	N
Factor 1: Cognitive aspects			
1. During the tutorial, many explanations of the subject content were given by individual students	3.30	1.21	449
2. In the tutorial, group explanations of the subject content were given in own words	3.18	1.21	449
3. Students posed adequate questions to each other to obtain a deeper understanding of the subject matter	3.21	1.18	449
4. Students asked critical questions to check the explanations of content given by other students	3.01	1.14	449
5. In the tutorial group, I learnt much from the contributions of the other group members	3.08	1.24	449
6. In the tutorial group, misconceptions about the subject matter were corrected by other group members	3.39	1.11	449
7. Group members built on each other's arguments	3.19	1.01	449
Factor 2: Motivational aspects			
8. I felt myself as a member of the group responsible for the progress of the group	3.49	1.13	449
9. If I did not prepare well for the tutorial group meeting, I felt uncomfortable in the group	3.62	1.29	449
10. I became more perceptive and sensitive to the needs of the other students within my group during group work	3.43	1.06	449
11. The tutorial group stimulated my self-study activities	3.35	1.23	449
12. The tutorial group had a positive effect on my academic commitments/efforts	3.17	1.21	449
13. My interest in the subject matter increased due to the discussions in the tutorial group	3.30	1.21	449
14. The tutorial group discussion stimulated my group mates to exert maximum effort	2.92	1.14	449
Factor 3: Demotivational aspects			
15. During the course of the tutorial, some group members contributed less to the tutorial group discussion	3.79	1.23	449
16. Some group members intentionally withheld information they had acquired during self-study	3.04	1.24	449
17. I did not contribute as much as to the tutorial group discussion as I could have done	2.86	1.29	449
18. Some group members had a negative effect on the contributions of other group members	2.72	1.32	449
19. Some group members let others do the work	3.44	1.30	449
20. Give a qualification for the overall group productivity (overall score) (1 = insufficient, 2 = reasonable, 3 = sufficient, 4 = good, 5 = excellent)	3.20	0.52	449

Note: Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree and 5 = strongly agree.

Subjects

A total of 483 students responded to the survey with an average response rate of 80%. So in total, 20% of the students did not return the questionnaire. Furthermore, 34 questionnaires with missing information were omitted from the study. Hence, data from 449 students were included in the study (a response rate of 74%). This consisted of first ($n=183$), second ($n=156$), and ($n=110$) third year undergraduate medical students. In total, 52 groups participated in the study. This was made of 20 first year, 19 second year and 13 third year groups. The number of students completing the instrument per group varied between 6 and 11.

Statistical analysis

First of all, mean scores and standard deviations were computed at the item level and at the factor level. It is appropriate to compute mean scores, since research has shown that the Likert-response format produces interval data at the scale level (Carifo & Perla 2008).

A confirmatory factor analysis (CFA) was carried out to assess the adequacy of the three factors underlying the items, to address the construct validity of the instrument. The data were analyzed at the student level ($n=449$) as individual students were asked to give their personal opinion about the group productivity. In the confirmatory factor model, specified in this study, all three factors were correlated. Observed items 1–7 were affected by the first factor, observed variables 8–14 were affected by the second factor, observed variables 15–19 by the third factor. 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 skew and kurtosis values of all data used are smaller than ± 1.5 , or even ± 1.0 , which implies that they are normally distributed. A maximum likelihood estimation (MLE) was used when conducting the CFA. The AMOS program was used to determine whether the data confirmed the three-factor model (Arbuckle 1999).

The coefficient alpha was computed for each factor to determine the internal consistency of each factor. A coefficient of 0.70 or higher was considered as acceptable. In addition, generalizability studies were conducted to estimate the reliability of each factor and to determine how many student responses are needed per group (Crick & Brennan 1983). The analyses were conducted at the individual student level ($n=449$). An all-random students-nested-within-groups design was used, with groups as universe of generalization or object of measurement. In total, 52 groups were involved

who had each been judged by six students or more. This design allows variance component estimation of two sources: (1) differences between groups (G; object of measurement) and (2) differences between students nested within groups and general error (S: G, e) (Shavelson & Webb 1991). Reliability indices (generalizability (G) coefficient and Standard Error of Measurement (SEM)) are reported as function of the number of students completing the questionnaire.

Results

Descriptive statistics

Table 1 contains the items and its mean scores and standard deviations (SDs). The mean scores and SDs for each factor are reported in Table 2, as well as the overall productivity score. As illustrated in this table, the average scores per factor varied between 3.12 and 3.32 (scale 1–5). In the cognitive and demotivational domains, the mean was 3.12 and 3.17 with SD values of 0.81 and 0.82, respectively. In the motivational domain, the mean was 3.32 with an SD of 0.82.

Construct validity

The correlation coefficient between factors 1 and 2 is 0.64, between factors 2 and 3, 0.14 and between factors 1 and 3, 0.19. All correlation coefficients were significant ($p=0.01$). A CFA was conducted. Several statistics were calculated in order to assess whether the empirical data fit with the theoretical proposed model. Unfortunately there is no single best statistic that gives an insight into the fit of the model. The best a researcher can do is to compute several statistics which reflect the fit of the model (Van Berkel & Schmidt 2005). The Chi square divided by the degrees of freedom, i.e. (CMIN/DF) must be less than 3 for correct models. The p -value should be higher than 0.05. The two most important indices are the RMSEA, which should be lower than 0.08 for a good fit and lower than 0.05 for an excellent fit and the CFI which should be 0.90 for a good fit and 0.95 for an excellent fit.

The results of the three-factor model as outlined above, and presented in Table 3, showed the following results: chi-square [149 df] = 452.08, $p=0.000$, a root mean square residual of 0.067 and a CFI of 0.888. The results for a one-factor model are presented in Table 3, and showed the following results: chi-square [153 df] = 965.22, $p=0.000$, a root mean square residual of 0.109 and a CFI of 0.701. The CMIN/DF is 6.3 for the one-factor model. The results for a two-factor model are also presented in Table 3. The statistics for the three-factor model demonstrated a better fit than the statistics for the

Table 2. Number of items, number of students, minimum and maximum score, mean score (scale 1–5), SD, and coefficient alpha per factor and/or the overall group productivity score (scale 1–5).

	N items	N	Minimum	Maximum	Mean	SD	Alpha
F1: Cognitive	7	449	1	5	3.12	0.81	0.82
F2: Motivational	7	449	1	5	3.32	0.82	0.82
F3: Demotivational	5	449	1	5	3.17	0.82	0.64
Overall score	19	449	1	5	3.20	0.52	0.79

Table 3. Fit indices of the one-factor, two-factor and three-factor models.

	N	Chi-square	df	P	CMIN/df	RMSEA	CFI	TLI	PCLOSE
One factor	449	965.22	153	0.000	6.308	0.109	0.701	0.666	0.000
Two factor	449	766.45	152	0.000	5.042	0.095	0.774	0.745	0.000
Three factor	449	452.08	149	0.000	3.034	0.067	0.888	0.872	0.000

Note: N = 449. Method: maximum likelihood.

Table 4. Estimated variance components for the variance associated with groups and students nested within groups are given (in percentages in between brackets) and the generalizability coefficient (G-coefficient) and SEM, as a function of the number of student ratings (N) for the 3 factors (scale 1–5) and the overall score.

	Estimated variance components				
Factor	Group	SG	N	G-coefficient	SEM
F1: Cognitive	0.10 (15%)	0.56 (85%)	8	0.59	0.26
			9	0.62	0.25
			10	0.65	0.24
			11	0.67	0.23
F2: Motivation	0.04 (6%)	0.64 (94%)	8	0.33	0.28
			9	0.36	0.27
			10	0.39	0.25
			11	0.41	0.24
F3: Demotivation	0.07 (10%)	0.60 (90%)	8	0.49	0.27
			9	0.52	0.26
			10	0.55	0.25
			11	0.57	0.23
Overall score	0.28 (22%)	0.99 (78%)	8	0.69	0.35
			9	0.72	0.33
			10	0.74	0.32
			11	0.76	0.30
			15	0.81	0.26

Note: The overall score was a separate item and is not the mean score of all the 19 items, but deals with one overall item (item 20).

one-factor model and the two-factor model. For the three-factor model, the following statistics given in Table 3 meet the criteria: the CMIN/DF is equal to 3 but not below 3, the *p*-value does not differ from zero, but a CFI of 0.89 and an RMSEA of 0.07 indicate a good fit of the three-factor model. In general, the results of the CFA indicate that the three-factor model shows a good fit since the two most important conditions are met. Based on the analysis of the data, it can be concluded that the instrument within the setting of this study appeared to reveal fairly valid factor scores.

Reliability

The coefficient alpha calculated indicated high internal consistency for factors 1 and 2, both 0.82 and a lower internal consistency of 0.64 for factor 3, the demotivational factor. The results of the generalizability studies demonstrated that the variance associated with groups for the overall score is 22% as reported in Table 4. This percentage is the true variance or the variance of interest. The variance associated with groups varies per factor between 6% and 15%. The estimated variance

components were used to estimate reliability indices. Table 4 provides the G-coefficients per factor as a function of the number of student responses per group and the corresponding error of measurement (SEM). The SEM can be used to estimate confidence intervals for individual scores. The SEM should be lower than or equal to 0.25 (0.5/1.96) at the 95% confidence interval, taking into account a practical significance level of 0.5 point on a scale from 1 to 5. Based on this practical significance level of 0.5 points on a scale 1–5, at least nine student responses are required to obtain reliable results for factor 1, and at least 10 student responses for factors 2 and 3. To obtain a reliable G-coefficient of at least 0.70 or higher, at least nine students' responses are needed for the overall score.

Conclusion and discussion

This article focuses on the validity and reliability of an instrument to assess tutorial group effectiveness in PBL environments. All the items were based on Slavin's theoretical framework of collaborative learning that highlights two theoretical perspectives on group learning, one is a cognitive perspective and the other a motivational perspective. The motivational domain indicates the extent to which students motivate, show concern, and help each other learn. The demotivational domain indicated the extent to which non-participation of students affects the group dynamics and hence has a negative effect on student learning in these groups. The cognitive domain is based on the interactions and explanations between peers, which enhances learning.

The results of the CFA indicated that a three-factor model revealed a good fit with the data, i.e. the RMSEA and CFA indicated a good fit. Thus, based on the CFA, it can be concluded that the instrument within the setting of this study appeared to reveal valid factor scores. The findings of the generalizability studies indicated that at the factor level, at least 9–10 student responses are needed in order to obtain reliable data at the factor level. This finding indicates that the instrument is reliable since the tutorial group size within this study varied between 6 and 11.

The finding that the factor scores of the instrument within the setting of this study is valid and reliable, implies that the data collected can be used to measure group effectiveness in PBL tutorials. Having an instrument that discriminates between less and more effective groups equips faculty and students with diagnostic information about group performance and learning. This is important as dysfunctional groups negatively affect tutorial group effectiveness (Virtanen et al. 1999; Houlden 2001; Hendry et al. 2003). Thus, the information obtained at the factor level by using the instrument can create

awareness and an increased understanding of problematic and well-functioning PBL groups. This awareness and evaluation can encourage tutors/facilitators and students to implement relevant strategies and training to improve tutorial group functioning within a PBL curriculum.

In this study, the moderate ratings of the cognitive and motivational factor and the overall tutorial productivity indicate that there is room for improvement in the tutorial group effectiveness within the setting of this study. The same holds for the score on the demotivational factor. A lack of social cohesion in some may be keeping group morale down.

The validated TGEI in this study can be used to provide students with feedback about the functioning of their group during midterm and at the end of the theme. This information would stimulate students to think about their roles and responsibilities as a collaborative learner and also about their peers contributions and attitudes in the small group setting in order to optimize feedback and tutorial group effectiveness. Useful and timely feedback in the small-group PBL could revitalize the group and individuals and encourage the development of essential skills needed by health professionals (Mennin 2007). Thus, the TGEI can be used to provide students and tutors with feedback about the performance of the tutorial group as well as highlight areas of deficiency in group effectiveness.

Although, the CFA and the generalizability study in this article demonstrated that the instrument within the setting of this study revealed valid and reliable factor scores, the data also demonstrated that two factors correlated highly, which indicates that the factors do not discriminate much in terms of group performance. On the contrary, a two-factor solution in which the two high-correlating factors were put together in one factor resulted in a poorer fit as compared to a three-factor solution. Furthermore, the current data were collected in a diverse student population which might limit the generalizability of the findings. Nevertheless, the groups were mixed and well-balanced which might imply that the findings are generalizable toward other PBL curricula. Additional research is needed to assess the use of this instrument in providing insight into the influence of multicultural and multilingual settings on the motivational and cognitive aspects of group effectiveness in PBL. Further validation of the instrument with higher *N* is also recommended.

Acknowledgments

Thanks to Ron Hoogenboom from Maastricht University for statistical support on running and interpreting the generalizability data and to the National Research Foundation from South Africa for the funding received.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

VEENA S. SINGARAM is a program director in the Department of Medical Education and a PhD candidate in the Department of Educational Development and Research at the University of Maastricht.

CEES P. M. VAN DER VLEUTEN, PhD, is a professor of education and chair of the Department of Educational Development and Research at the University of Maastricht.

HENK VAN BERKEL, PhD, is an educational psychologist in the Department of Educational Development and Research at the University of Maastricht.

DIANA HJM DOLMANS, PhD, is an educational psychologist and associate professor in the Department of Educational Development and Research at the University of Maastricht.

References

- Arbuckle JL, Wothke W. 1999. Amos User's Guide, version 4.0. Chicago, IL: Small Waters Corporation.
- Carifio J, Perla R. 2008. Resolving the 50-year debate around using and misusing Likert scales. *Med Educ* 42:1150–1152.
- Carlo MD, Swadi H, Mpofu D. 2003. Medical student perceptions of factors affecting productivity of PBL tutorial groups: Does culture influence the outcome? *Teach Learn Med* 15:59–64.
- Crick JE, Brennan RL. 1983. Manual for Genova: A generalized analysis of variance system. Iowa, IA: American college testing program.
- De Grave WS, Dolmans DHJM, Van Der Vleuten CPM. 2001. Student perceptions about the occurrence of critical incidents in tutorial groups. *Med Teach* 23:49–54.
- De Grave WS, Dolmans DHJM, Van Der Vleuten CPM. 2002. Student perspectives on critical incidents in tutorial groups. *Adv Health Sci Educ* 7:2001–2009.
- Dolmans DHJM, De Grave W, Wolfhagen HAP, Van Der Vleuten CPM. 2005. Problem-based learning: Future challenges for educational practice and research. *Med Educ* 39:732–741.
- Dolmans DHJM, Schmidt HG. 2006. What do we know about cognitive and motivational effects of small group tutorials in problem-based learning? *Adv Health Sci Educ* 11:321–336.
- Dolmans DHJM, Wolfhagen HAP, Van Der Vleuten CPM. 1998. Motivational and cognitive processes influencing tutorial groups. *Acad Med* 73:S22–S24.
- Hendry GH, Ryan G, Harris J. 2003. Group problems in problem-based learning. *Med Teach* 25:609–616.
- Houlden RL, Collier CP, Frid PJ, John SL, Pross H. 2001. Problems identified by tutors in a hybrid problem-based learning curriculum. *Acad Med* 76:81.
- Mennin S. 2007. Small-group problem-based learning as a complex adaptive system. *Teach Teach Educ: Int J Res Stud* 23:303–313.
- Shavelson RJ, Webb NM. 1991. Generalizability theory. A primer. London: Sage.
- Slavin RE. 1996. Research on cooperative learning and achievement: What we know, what we need to know. *Contemp Educ Psychol* 21:43–69.
- Singaram SV, Dolmans DHJM, Lachman N, Van Der Vleuten CPM. 2008. Perceptions of problem-based learning (PBL) group effectiveness in a socially-culturally diverse medical student population. *Educ Health* 21:116.
- Van Berkel H, Schmidt H. 2005. On the additional value of lectures in a problem-based curriculum. *Educ Health* 18:45–61.
- Virtanen PJ, Kosunen EAL, Holmberg-Marttila DMH, Virjo IO. 1999. What happens in PBL tutorial sessions? Analysis of medical students' written accounts. *Med Teach* 21:270–276.
- Visschers-Pleijers AJSF, Dolmans DH JM, Wolfhagen, IHAP, Van Der Vleuten CPM. 2005. Development and validation of a questionnaire to identify learning-oriented group interactions in PBL. *Med Teach* 27:375–381.