Effects of Tutor Expertise on Student Performance in Relation to Prior Knowledge and Level of Curricular Structure

Diana H. J. M. Dolmans, PhD, Ineke H. A. P. Wolfhagen, PhD, and Henk G. Schmidt, PhD

ABSTRACT

Purpose. To test whether there are effects of tutor expertise on student performance under conditions of curricular materials that have low or high levels of structure and that are poorly or well matched to students' levels of prior knowledge.

Method. The study was conducted in 1994–95 at the medical school of the University of Limburg. The data set used for analysis included 135 tutorial groups (with ten to 12 students per group), 119 tutors (each running only one group per unit), and 15 units in four curriculum years. The analysis was conducted at the level of tutorial groups since a tutor's level of expertise might differ for distinct units. Tutors were asked to judge their levels of expertise related to the cases discussed, based on which a distinction was made between expert and non-expert tutors. The degrees of structure of curricular materials and students' levels of prior knowledge were rated by the students. Using analyses of variance, students' scores on end-of-unit tests (each with about 150 true–false items) were compared for groups led by expert and non-expert tutors under conditions of low and high levels of structure and low and high levels of prior knowledge.

Results. No difference was found between the test scores of groups led by expert and non-expert tutors. The interaction effects between expertise and structure and expertise and prior knowledge also turned out to be not statistically significant.

Conclusion. The results suggest that expert tutors do not compensate for lack of curricular structure or students' lack of prior knowledge. This finding is not consistent with that of a recent study that expert tutors do compensate for lack of structure and lack of prior knowledge. This discrepancy may be accounted for by a much smaller range within which the structuredness of the curriculum and students' levels of prior knowledge varied in the present study compared with the previous study. An implication might be that faculty should put their efforts into designing structured curricula that are well matched to students' levels of prior knowledge instead of selecting hyper-expert tutors. Acad. Med. 1996;71:1008–1011.

Several explanations have been proposed for contradictory findings in studies investigating whether the tutor in a problem-based curriculum should be an

Dr. Dolmans and Dr. Wolfhagen are educational psychologists, and Dr. Schmidt is professor of educational psychology and health professions education; all with the Department of Educational Development and Research, University of Limburg, Maastricht, The Netherlands.

Correspondence should be addressed to Dr. Dolmans, University of Limburg, Department of Educational Development and Research, PO Box 616, 6200 MD Maastricht, The Netherlands. e-mail: <dolmans @educ.rulimburg.nl>. Reprints are not available.

expert in the subject matter under discussion. Schmidt[1] concludes that when the structure of a unit is low or students lack prior knowledge, the impact of a tutor's expertise on students' performances is greater. Davis et al.,[2] who in an initial study found significant differences between students' performances in groups led by experts and those led by non-experts, could not replicate this finding when more highly focused cases were used. In this respect, Wilkerson[3] concludes that the effect of subject-matter expertise on behavior may be situa-tion-specific rather than tutor-specific.

Our aim in the present study was to investigate whether there are effects of tutor expertise on test scores under conditions of curricular materials that have low or high levels of structure and that are poorly or well matched to students' levels of prior knowledge. When assessing tutor expertise, most studies make use of independent judges' ratings of whether a tutor is an expert or non-expert related to the subject of a particular unit.[1,2] There is, however, a snag here. One of the characteristics of prob-

lem-based learning units is that they cover multiple disciplines. This implies that it is rare for a tutor to be either an expert or a non-expert for a particular unit. A tutor will be an expert for particular cases and a non-expert or semi-expert for other cases included in a single unit. In this study, tutors were therefore asked to rate their levels of expertise related to the subject of each case included in a particular unit on a three-point Likert scale: 1, no expertise at all; 2, sufficient expertise; and 3, much expertise. Subsequently, for each tutor an average expertise score per unit was computed. It is assumed that this measure is more fine-grained than that resulting from asking independent judges to rate tutors as expert or non-expert.

METHOD

Subjects

The study participants consisted of students in 177 tutorial groups, run by 153 tutors, in the academic year 1994–95 at the medical school of the University of Limburg. Each tutor ran only one tutorial group in each of several curricular units. The numbers of students included in the tutorial groups varied between ten and 12. The analysis was conducted at the level of tutorial groups, since a tutor's level of expertise might differ for distinct units. The total number of tutorial groups included in the data set used for analysis was 135 (76.3% of all tutorial groups). The remaining tutorial groups were excluded from the analysis because their data records were incomplete. The final data set consisted of 119 tutors participating in 15 units (each lasting about six weeks) in four curriculum years (with approximately 150 students per year).

Instruments

Students' performances were measured at the end of each unit by tests contain-

ing about 150 true–false items. Each test score was expressed as the percentage of items answered correctly minus the percentage of items answered incorrectly.

At the end of each unit, students filled out a program-evaluation questionnaire that contained statements with a five-point Likert scale ranging from 1, "strongly disagree," to 5, "strongly agree." One item was used to indicate the student's level of prior knowledge for a particular unit: "The unit's subject matter was adapted to my prior knowledge." Two items were used as indicators of the amount of structure of each unit: "The objectives of the unit were clear to me" and "The majority of the cases were presented clearly." The scores on these two items were averaged. Unit objectives and case presentations were at least somewhat independent of tutors' performances because during each unit all students and all tutors were provided with the same course book, which contained a global description of the unit objectives and the texts of the cases.

As indicated above, tutors were asked to rate their subject-matter expertise for each case included in a particular unit on a three-point Likert scale: 1, no expertise at all; 2, sufficient expertise; and 3, much expertise. For each tutor an average self-rating was computed for all cases included in a particular unit. The average score for each tutor was considered to be the average level of expertise of the tutor related to the cases discussed for that unit.

Analysis

The analyses were conducted at the level of tutorial groups. Students' test scores were aggregated at the tutorial-group level because individual test scores might not be independent of each other, since the students worked together in tutorial groups for several hours a week. The average scores for the structuredness of the curricular ma-

terials and the degree to which the curricular materials were adapted to students' levels of prior knowledge were also aggregated at the tutorial-group level. These scores were subsequently dichotomized to facilitate interpretation of the data. A score below the median value was considered low, a score above the median value high. The average score for a tutor's level of expertise was also dichotomized. If the average score of a tutor was below the median value, this tutor was considered a non-expert; otherwise the tutor was considered an expert.

The data were further analyzed using analysis of variance (ANOVA). In the ANOVA, average test scores per tutorial group were standardized by making use of a z-transformation. The formula used was mean test score per tutorial group minus total mean test score for that particular unit, divided by the corresponding standard deviation. The mean of the resulting z-score is equal to zero and the standard deviation is equal to one. The independent variables included in the ANOVA were structuredness, prior knowledge, and tutor expertise. The dependent variable was test scores.

RESULTS

The average percentages of items answered correctly minus the percentages of items answered incorrectly for the 135 tutorial groups varied between 25% and 58%, with a mean score of 40.4% (SD, 6.0). The average score for the structuredness of the curricular materials was 3.76 (SD, 0.48), with a median value of 3.79 and a range of 1.67–4.72 (on a scale of 1–5). The average score for the prior knowledge item was 3.77 (SD, 0.44), with a median value of 3.74 and a range of 2.33–4.75 (on a scale of 1–5). The average score for a tutor's level of expertise was 2.13 (SD, 0.45), with a median value of 2.18 and a range of 1–3 (on a scale of 1–3). Table 1 contains the average percentages of

Table 1

Mean Scores for Students in 135 Tutorial Groups Led by Expert or Non-expert Tutors under Conditions of Low or High Levels of Curricular Structure and Low or High Levels of Students' Prior Knowledge, University of Limburg Medical School, 1994–95*						
	For Groups Led by Non-expert Tutors			For Groups Led by Expert Tutors		
	% Score			% Score		
Condition	Mean	(SD)	No. of Groups	Mean	(SD)	No. of Groups
Low structure	40.0	(5.4)	30	42.3	(7.3)	34
High structure	40.7	(4.9)	36	40.3	(4.9)	35
Low prior knowledge	40.4	(5.0)	31	42.2	(6.8)	30
High prior knowledge	40.4	(5.2)	35	40.6	(5.8)	39
Overall	40.4	(5.1)	66	41.3	(6.2)	69

*At the end of each curricular unit, the students took a test with about 150 true–false items. Each student's score was expressed as the percentage of items answered correctly minus the percentage of items answered incorrectly. Whether tutorial groups were conducted by expert or non-expert tutors was determined by tutors' self-ratings on each case within a particular unit. Whether the curricular materials had low or high levels of structure and whether the units were poorly or well matched to the students' levels of prior knowledge were determined by students' ratings. See text for details.

items answered correctly minus the percentages of items answered incorrectly for different levels of tutor expertise under conditions of low and high structure and low and high prior knowledge. To test whether these scores differed significantly, two two-way ANOVAs were performed using standardized test scores.

The first two-way ANOVA used test scores as the dependent variable and tutor's level of expertise and amount of structure as independent variables. The main effect of expertise on test scores was not statistically significant ($F_{134,1}$ = 0.107, p = .744, MS_e = 0.012). Thus, there was no effect of tutor expertise on test scores. The effect of low versus high structure was also not statistically significant ($F_{134,1}$ = 0.054, p = .817, MS_e = 0.006). This implies that when structure was low, the effect of tutor expertise was not greater than when structure was high. The two-way interaction between tutor's expertise level and amount of structure also proved not to be statistically significant ($F_{134,1}$ = 1.526, p = .219, MS_e = 0.177), which

implies that a tutor's level of expertise did not have a differential effect on test scores under conditions of low versus high structure.

The second two-way ANOVA used test scores as the dependent variable and tutor's level of expertise and amount of prior knowledge as independent variables. The main effect of expertise on test scores was, of course, not statistically significant ($F_{134,1}$ = 0.091, p = .763, MS_e = 0.011). The effect of prior knowledge was also not statistically significant ($F_{134,1}$ = 0.696, p = .406, MS_e = 0.081). This implies that tutorial groups that indicated limited prior knowledge did not differ in their test scores from groups that indicated high amounts of prior knowledge. The two-way interaction between tutor's level of expertise and amount of prior knowledge was also not significant ($F_{134,1}$ = 0.281, p = .597, MS_e = 0.033). Thus, the effect of tutor expertise on test scores did not differ between conditions of low and high prior knowledge. (In addition, two-way ANOVAs were conducted in which curriculum year

and curricular unit were used as covariants to test the effects of curriculum year and unit. These effects were not statistically significant.)

DISCUSSION

The aim of this study was to investigate whether there are effects of tutor expertise on test scores under conditions of curricular materials that have low or high levels of structure and that are poorly or well matched to students' levels of prior knowledge. No statistically significant difference was found between the performances of groups led by expert and non-expert tutors. The interaction effects between a tutor's level of expertise and amount of structure on the one hand and between a tutor's level of expertise and students' levels of prior knowledge on the other also turned out to be not statistically significant. This suggests that a tutor's level of expertise does not compensate for lack of structure or for lack of prior knowledge.

What do these results demonstrate when compared with a recent finding in a study by Schmidt[1] that compared non-expert and expert tutors? In the Schmidt study a differential effect of tutor expertise on student performance was found, as a main effect and in interaction with the structuredness of curricular materials and students' levels of prior knowledge, a finding leading Schmidt to conclude that tutor expertise compensates for lack of structure and lack of prior knowledge. It is assumed by some researchers that an expert tutor compensates for these deficiencies by playing a more directive role in the tutorial group and suggesting more of the topics for discussion.[4] In the present study, we concluded that tutor expertise compensates for neither lack of structure nor prior knowledge. An explanation that presents itself as a possible candidate for explaining the discrepancy between the two studies may be the different levels at which the

data were analyzed. In the Schmidt study, the student was the unit of analysis, whereas in the present study the tutorial group was the unit of analysis. This was done to avoid the possibility that individual test scores were not independent of each other. This means that the data in our study were all aggregated at the tutorial group level, and hence the power of our statistical tests was lower than that in the Schmidt study. However, closer inspection of our data reveals that even if the power levels had been similar, our conclusions must have been the same, namely, that structure of the curriculum has no differential effect on test scores, either as a main effect or in interaction with tutor expertise. These findings can be attributed only to a much smaller range within which the structuredness of the curriculum and students' levels of prior knowledge varied as compared with the curriculum studied by Schmidt.

Did students indeed view the curricula differently in our study and the Schmidt study? The range for structuredness in our study was 1.67–4.72 (mean of 3.76, SD, 0.48) and the range for level of prior knowledge was 2.33–4.75 (mean of 3.77, SD, 0.44). The Schmidt study does not report average scores and standard deviations, but 40% of the scores for the structuredness of the curriculum are lower than or equal to 3 (on a scale of 1–5). If, in our study, a score equal to or lower than 3.44 is considered to be a low level (in our study average scores are used, which implies that decimal places need to be taken into account), 19% of the scores can be considered low, as opposed to 40% in the Schmidt study. The same comparison can be made for level of prior knowledge: 23% of the scores can be considered low in our study as opposed to 51% in the Schmidt study. Thus, the structuredness of the curricular materials in our study varied within a smaller range than that in the Schmidt study, which might explain why the expert tutors in our study did not compensate for lack of structure and lack of prior knowledge, in contrast to the expert tutors in the Schmidt study.

The present study, like previous studies conducted by Schmidt[1] and by Davis et al.,[2] demonstrates that the effect of a tutor's subject-matter expertise on students' performance will indeed, as Wilkerson[3] had concluded, be situation-specific rather than tutor-specific. In a well-structured curriculum, a tutor's level of expertise is less likely to be actually "used" by students or has less surplus value, because students rely on the curriculum or allow the curriculum to do its work. An implication might be that faculty should put their efforts into designing curricula that are well structured and that link up well with students' prior knowledge instead of selecting hyper-expert tutors.

REFERENCES

1. Schmidt HG. Resolving inconsistencies in tutor expertise research: does lack of structure cause students to seek tutor guidance? Acad Med. 1994;69:656–62.
2. Davis WK, Oh MS, Anderson RM, Gruppen L, Nairn R. Influence of a highly focused case on the effect of small-group facilitators' content expertise on students' learning and satisfaction. Acad Med. 1994;69:663–9.
3. Wilkerson L. The next best thing to an answer about tutors' content expertise in PBL. Acad Med. 1994;69:646–8.
4. Silver M, Wilkerson L. Effects of tutors with subject expertise on the problem-based tutorial process. Acad Med. 1991;66:298–300.