Promoting diagnostic problem representation

Mathieu R Nendaz[1] *& Georges Bordage*[2]

Purpose Problem representation, as mediated by semantic qualifiers (SQs), has been associated with better diagnostic outcomes. The purpose of this study was to assess the effect of training medical students to use semantic abstractions as a means of building problem representations.

Methods Sixty second-year medical students were assigned to either an intervention group ($n = 20$) or a control group ($n = 40$) during 8 months of an Essentials in Clinical Medicine course which used standardized patient-based workshops. Students were trained to transform findings into SQs and to use abstractions to compare and contrast diagnostic hypotheses. Students were assessed using a standardized patient data collection checklist, a post-encounter patient finding questionnaire (PFQ), and case summaries and write-ups.

Results Experimental subjects used over twice as many SQs in their summaries as control group members (1·40 versus 0·63, $P = 0·006$). The correlation between checklist and PFQ scores was higher for the experimental group than for the control group ($r = 0·70$ versus $r = 0·58$, $P \leq 0·001$). There was no difference between groups in either the number of SQs used in write-ups nor in diagnostic accuracy ($P > 0·56$).

Conclusion A short instructional intervention was successful in promoting the use of SQs and enabled students to recall elicited findings better. This intervention did not enhance data interpretation and diagnostic accuracy. Use of SQs may therefore be a necessary tool for efficient problem representation but one that is insufficient when used in isolation. The naturalistic setting used in this study imposed a number of limitations, implying that further research should test whether instructional efforts should also emphasize recognition of key patient findings and knowledge representation.

Keywords education, medical, undergraduate/ *methods; medical history taking/ *standards; clinical competence/ standards; semantics, Chicago.

Medical Education 2002;36:760–766

Introduction

The acquisition of clinical and reasoning skills is an important aspect of medical education and deserves nurturing. However, clerks and doctors often lack interviewing and data interpretation skills.[1,2] As observed over 20 years ago,[3,4] traditional medical education tends to emphasize thorough data collection rather than focused inquiry based on early problem representation. Students learn how to acquire patient data in a thorough, head-to-toe fashion, but seldom have the opportunity to simultaneously incorporate data acquisition and diagnostic reasoning.[1] It has been shown that experienced doctors use a selective data collection approach influenced by their reasoning processes.[3,5] Thus, there is a need to evaluate instructional methods that help students better integrate clinical skills, reasoning skills and medical knowledge early in their education.

Early problem representation has been shown to play a crucial role in problem solving, both in medicine and in other fields. Traditionally, seven basic attributes have been used to characterize the chief complaint and the history of present illness, namely location, quality, severity, chronology, setting, aggravating or alleviating factors, and associated manifestations. Abstraction of these basic attributes (e.g. 'last night' becomes 'acute onset') constitutes one type of problem representation

[1]Department of Internal Medicine and Unit of Research and Development in Medical Education, Faculty of Medicine, University of Geneva, Geneva, Switzerland
[2]Department of Medical Education, College of Medicine, University of Illinois at Chicago, Chicago, USA

Correspondence: Mathieu R Nendaz, Département de Médecine Interne et Unité de Recherche et Développement en Education Médicale, Faculté de Médecine, Université de Genève, 1211 Geneva 14, Switzerland. Tel.: 00 41 22 372 9055; Fax: 00 41 22 372 9116; E-mail: Mathieu.Nendaz@hcuge.ch

Key learning points

Promoting problem representation for medical diagnosis among second-year medical students is a potential step towards improved diagnostic performance.

The use of semantic abstractions may be a necessary but insufficient condition to promote efficient problem representation.

Teaching the use of semantic abstraction should occur at an early stage, during the process of knowledge acquisition and organization, and should focus on eliciting and recognizing key patient findings.

shown to be associated with greater diagnostic accuracy.[6,7] With the attributes in hand, clinicians or medical students can transform two or three of the attributes into more abstract qualities, called semantic qualifiers (SQs) (e.g. 'three times before' becomes 'recurrent'), and use the set of qualifiers to build a global sense or representation of the problem before tackling possible diagnostic solutions (e.g. this is an 'acute – recurrent – large joint' pain that could be related to gout or septic arthritis). This is in turn associated with better diagnostic accuracy.

The purpose of the present study was to measure the effect of an instructional intervention on diagnostic argumentation and diagnostic accuracy during case write-ups. The instructional intervention was based on structural semantic theory and was aimed at promoting problem representation and prototypical differential diagnosis. It was hypothesized that the students in the intervention group would elicit more basic attributes, recall the basic attributes better, construct better problem representations, compare and contrast diagnostic hypotheses better, and have greater diagnostic accuracy than students in the control group.

Methods

Subjects and setting

Second-year medical students at the University of Illinois at Chicago attend a series of workshops as part of their Essentials of Clinical Medicine (ECM) course. The students are assigned to groups of between nine and 12 students, based on personal preferences. At the time of the study, 146 students were assigned to 14 groups. From the 14 groups, two were assigned by the course co-ordinator to a preceptor (MRN) who promoted problem representation and prototypical differential diagnoses during the small group sessions. Group assignment was based on the logistic scheduling of preceptors. These groups contained 11 students (two female and nine male) and nine students (four female and five male), respectively and constituted the experimental subjects (20 students). For each experimental subject, two matched control subjects were selected from among the remaining students in the class (40 controls), for the reasons discussed below. The subjects were matched according to gender, repetition of the first year of medical school, Cognitive Index (a score incorporating a weighted Medical College Admission Test (MCAT) score) and pre-medical school grade point average) and grade point average (GPA) at the end of the first year of medical school.

Four existing ECM workshops, held over an 8-month period (September to May), were used for the present study. The goals of the workshops were to develop skills in eliciting the chief complaint, history of present illness, and complete history, to prepare case write-ups, and to formulate problem lists and differential diagnoses. Each of the first three workshops consisted of two 3-hour sessions. During the first session of each workshop, every student interviewed a standardized patient (SP) and collected data pertaining to the chief complaint. The session ended with a group discussion and comments by the preceptor, based on excerpts from videotaped student interviews. For the second session, held about 4 days later, each student wrote a summary of the chief complaint and related data and provided a problem list and differential diagnosis. The student-generated materials were discussed with the other students and the preceptor. During the fourth and last workshop, each student conducted a focused 30-minute interview and physical examination of an SP. Two SP cases were used, both of which shared the same chief complaint but were related to different final diagnoses. After the encounter, each student completed a data interpretation questionnaire (15 min) and a patient findings questionnaire (10 min).

The prospective, matched-controlled, blinded subject study was designed to take advantage of the existing naturalistic organization of the curriculum. The first three workshops were used for instructional purposes and the fourth was used to collect outcome data. Institutional Review Board (IRB) ethical approval was requested and granted.

Instructional procedures and materials

The goals and sequence of the ECM workshops, as established by the course directors, were not modified for the present study. The preceptor's prerogative to provide comments and feedback to the experimental students about promoting problem representation and prototypical differential diagnosis was their only distinguishing feature. One of the investigators (MRN) served as the preceptor for the first three workshops for both experimental groups.

During the first three workshops, students were trained to:

1 elicit the basic attributes of the chief complaint;

2 summarize the case in an abstract form using SQs, and

3 use the abstract summary (problem representation) to compare and contrast two or three prototypical diagnostic hypotheses.

Videotapes and students' comments from their encounters with the SPs were used to illustrate the approach, emphasizing the acquisition of the *basic attributes* of the chief complaint and the transformation of certain attributes into more abstract terms, the *basic semantic attributes*. Students were instructed to use the semantic attributes to compare and contrast a limited number of diagnostic hypotheses (e.g. septic arthritis versus gout as the cause of an acute, recurrent, large joint arthritis). During the case discussions, the students were given feedback on how to transform the findings and how to use the ensuing abstractions to compare and contrast diagnoses. Individualized written comments were also provided for each case write-up to reinforce the reasoning approach.

To add practice opportunities between workshops and to reinforce the approach, three written cases were successively emailed to each student. The preceptor followed up the completion of these with personalized written feedback. Finally, 2 weeks before the fourth workshop, the subjects in the experimental groups received a personal email containing a summary of the reasoning approach presented during the workshops along with a clinical example using the process and an article on elaborated knowledge structures.[8]

Instrumentation and outcome measurements

Outcome data were collected during the fourth workshop, after each student saw one of the two SPs. A checklist of the minimal history and physical examination findings was established for each SP and reviewed by the case developers and SP trainers. After each encounter, the SP completed a scan sheet (the Data Acquisition Checklist), indicating whether each finding had been elicited by the student. This instrument captured the number and percentage of findings acquired.

The patient findings questionnaire (PFQ)[9] consisted of 20 multiple-choice questions asking each student to recall the findings gathered during the SP encounter. The PFQs were reviewed by the course directors, the case developers, and the course co-ordinators for accuracy. Students reported their answers on scan sheets. The scores captured the number and percentage of basic attributes elicited and recalled by the students. Students were advised not to include findings that they had not elicited during the encounter because a correction for guessing would be applied.

The data interpretation questionnaire asked the students to:

1 summarize the case in one sentence;

2 give two leading working diagnoses, and

3 write up the case, comparing and contrasting the leading working diagnoses.

The summaries were used to measure problem representation and were scored according to the number and percentage of attribute-related SQs (i.e. the number and percentage of basic attributes transformed into abstract qualities, or basic semantic attributes). Diagnostic accuracy was assessed by classifying each hypothesis according to four categories:

1 accurate final diagnosis;

2 part of the differential diagnosis;

3 related but vague diagnosis, and

4 unlikely diagnosis.

The use of semantic attributes to compare and contrast the diagnostic hypotheses in the write-ups was assessed by counting the proportion of arguments in which the students compared and contrasted diagnoses by actively using SQs.

At the beginning of the data interpretation questionnaire, students were asked whether they had heard about the process of transforming basic attributes into more abstract qualifiers, in order to assess potential contamination among control students and to act as a possible prompt for experimental students. To assess the prompting effect, the question was given to all experimental subjects and to half the control subjects (20 students). The other 20 control subjects received a filler question, not relevant to the study. Three groups of 20 subjects were thus created:

1 a prompted experimental group;

2 a prompted control group, and

3 an unprompted control group.

To assess student opinion on the ECM course and workshops, the experimental group was compared to the rest of the class concerning satisfaction, learning, feedback, and preceptor effectiveness.

Scoring procedures and statistical analyses

Each student was assigned an identification number by a course co-ordinator not involved in the study, thus maintaining confidentiality and blindness. The answers supplied by the data interpretation questionnaires were coded by three doctors. Global rater consistency, assessed on a training set of 10 subjects not involved in the study, ranged from 0·80 to 0·96, but correlation for individual items was more unstable (0·30–0·98). For this reason, the correctness of the reported diagnoses, the number of SQs, and the overall semantic competence of the subject was systematically assessed by all raters and established by consensus whenever there was disagreement.

Experimental subjects were compared to control subjects on five dependent variables: basic attributes *collected* (checklist scores), basic attributes *recalled* (PFQ scores), basic attributes *abstracted* in the summary and in the case write-up (number of SQs), and diagnostic accuracy. Correlation coefficients between checklist scores and PFQ scores were computed to assess the relationship between the data acquired during the encounter and the data recalled during the case write-up. Student *t*-tests were used for continuous variables following assessment of normality and homogeneity of variances. Data transformations were performed whenever necessary. Bonferroni correction for multiple comparisons was applied. Pearson's chi-square and Fisher's exact tests were used for categorical variables. The significance level was set at 0·05 and two-tailed tests were performed. Under the conditions of the study, and based on previous studies on semantic competence,[7] there was an estimated power of 80% to detect a 2·5% difference between continuous variables, with a standard deviation of 2·2%.

Results

After matching, there was no difference among the three groups for any of the potentially confounding variables, namely, gender, repeating the first year of

Table 1 Mean scores for data acquisition, case summaries and write-ups

	Control group	Experimental group	t	d.f.	*P**
Data collection and recall					
Checklist percent score (SD)	55·0 (11·9)	60·5 (12·5)	1·60	58	0·20
PFQ† percent score (SD)	65·3 (14·7)	64·0 (12·1)	0·34	58	0·73
Data interpretation					
Case summaries					
Mean (SD) SQs					
Total SQs (SD)	0·63 (0·77)	1·40 (1·10)	3·10	58	0·006
Distinct SQs (SD)	0·63 (0·77)	1·40 (1·10)	3·10	58	0·006
BSA‡ (SD)	0·60 (0·74)	1·35 (1·04)	3·20	58	0·006
Percent BSA (SD)	11 (13·0)	24 (18·0)	3·30	58	0·006
Case 1					
Total SQs (SD)	0·75 (0·85)	1·78 (1·09)	2·75	27	0·03
Distinct SQs (SD)	0·75 (0·85)	1·78 (1·09)	2·75	27	0·03
BSA (SD)	0·75 (0·85)	1·78 (1·09)	2·75	27	0·03
Percent BSA (SD)	13 (14·0)	30 (18·0)	2·75	27	0·01
Case 2					
Total SQs (SD)	0·50 (0·69)	1·09 (1·04)	1·90	29	0·18
Distinct SQs (SD)	0·50 (0·69)	1·09 (1·04)	1·90	29	0·18
BSA (SD)	0·45 (0·60)	1·00 (0·89)	2·04	29	0·15
Percent BSA (SD)	9 (12·0)	20 (17·0)	2·04	29	0·15
Write-ups					
Mean (SD) SQs					
Total SQs	4·93 (3·90)	4·45 (3·05)	0·47	58	0·63
Distinct SQs	3·70 (2·61)	3·30 (2·36)	0·57	58	0·56
BSA	2·75 (1·56)	2·60 (1·70)	0·34	58	0·73
Percent (SD) semantic arguments	23 (23·0)	33 (47·0)	0·74	58	0·45

* = Corrected for multiple comparisons (Bonferroni).
† = Patient Findings Questionnaire.
‡ = Basic semantic attributes.

medical school, Cognitive Index, and GPA at the end of the first year of medical school (all *P*-values greater than 0·51). There were no differences between the prompted and unprompted control groups for any of the dependent variables. Consequently, subjects were combined into 20 experimental and 40 control subjects for subsequent analyses.

Subjects in the experimental group used more SQs to summarize the case than those in the control group, representing a more than two-fold increase in total and distinct SQs, as well as in the number and percentage of basic semantic attributes (Table 1). No differences were found between groups in the use of SQs in the case write-up section of the data interpretation questionnaire (comparing and contrasting diagnostic hypotheses), nor in the proportion of arguments with which students compared and contrasted diagnoses by actively using SQs (semantic arguments).

Cronbach alphas for the data acquisition checklist and PFQ scores were 0·60 and 0·61 respectively, both of which were considered acceptable. The mean checklist (data acquisition) per cent scores and PFQ (data recall) per cent scores were not statistically different between the experimental and control groups (Table 1). Correlation coefficients between checklist scores and PFQ scores were significantly higher among experimental subjects than among controls, with $r = 0{\cdot}70$ compared to $r = 0{\cdot}58$, $P \leq 0{\cdot}001$.

The number and percentage of subjects who mentioned the correct final diagnosis in their leading working diagnoses were not statistically different between groups, with 20 subjects (50%) in the control group and nine (45%) in the experimental group (Fisher's exact test, $P = 0{\cdot}78$). There was a large case effect on diagnostic accuracy for both the control and experimental groups, respectively, at 85% compared to 89% (Fisher's exact test, $P = 0{\cdot}99$) for Case 1, and 15% compared to 9% (Fisher's exact test, $P = 0{\cdot}99$) for Case 2.

The workshop evaluations from the experimental students were similar to those of the class as a whole. A total of 64% of the experimental students had a very good or excellent opinion of the workshops overall (the same as for the class as a whole); 18% enjoyed the workshops far more than other experiences (compared to 14% for the class as a whole). Experimental subjects felt they had learned far more (19% versus 14%) or more (57% versus 52%) than during other educational experiences. A total of 92% of the experimental group felt they received adequate feedback (compared to 76% of the whole class) and 95% considered the preceptor to be effective (compared to 89% of the whole class).

Discussion

The results of the present study show that a short educational intervention was successful in enabling students to use early abstract transformations of clinical findings in their written summaries, a variable assumed to reflect problem representation. Students in the experimental group used more total and distinct SQs and more basic attributes in their summaries than did those in the control group. In addition, there was a higher correlation between checklist scores (data acquisition) and PFQ scores (data recall) among experimental subjects, suggesting that students in the experimental group were also able to better recall encounter findings than were control subjects. However, the number of SQs used in the long case write-ups and diagnostic accuracy were similar in both groups. In summary, students in the experimental group used more abstractions for problem formulation, and recalled relevant information acquired during the encounter better, but failed to use the abstractions for data interpretation and diagnostic evaluation. The meaning of these results will be discussed according to two main issues, namely, the efficiency of the intervention and the appropriateness of the underlying psychological theory.

Several psychological obstacles make it difficult to teach and assess problem representation. Firstly, although it was assumed that the case summaries were a reflection of problem representation, it is not certain whether the increased use of SQs in the experimental students' summaries necessarily implied truly better problem representation. Students' use of SQs may imply that they were merely trying to apply what they had learned during the workshops, without fully recognizing the importance of the findings they were transforming. Secondly, because reasoning is tightly linked to case content,[3] students' uneven content knowledge probably played some part in their difficulty in eliciting and recognizing key clinical data and in using SQs to further assess the cases. This is suggested by the large case effect on the degree of transformation of findings into SQs and on diagnostic accuracy, as well as by reports from the students about their lack of experience with similar chief complaints. Thirdly, because of differences in learning styles, abstraction as a mediator of problem representation may not be the preferred medium of representation for all students. Some students may be more comfortable with visual or diagrammatic representations than with abstractions. Fourthly, the instructional intervention focused more on reporting findings than on solving problems. Consequently, the students might have considered the intervention to be merely a method-driven strategy rather than a problem-solving one. This distinction is important because there is a lack of evidence that teaching reasoning processes separately from content is effective.[3] This might have prevented some students from internalizing semantic abstractions as a necessary and useful condition for learning and application for practice.[10] Semantic qualifiers may therefore represent a necessary tool for correct diagnosis of a patient's condition, but one which is insufficient when used in isolation. The acquisition, recognition, and transformation of selected, pertinent pieces of information into SQs must be combined for optimal performance. Thus, additional instruction would appear necessary if we are to truly teach students to acquire better problem representations. This should aim to enhance immediate recognition of key findings that can be abstracted and to improve integration of content knowledge. Moreover, since problem representation should play an important role, not only at the end of an encounter when interpreting the findings gathered, but during the encounter as a means of better acquiring relevant patient data, training for problem representation should occur in a dynamic fashion from the very beginning of the encounter and throughout, and should begin with the chief complaint.[1]

The current results also shed some light on the underlying cognitive theory used for the educational intervention. Although greater use of SQs has been correlated with better diagnostic outcomes, this may not be a causal relationship but simply an index of better and more organized specific knowledge allowing for better diagnostic competencies. If the latter is true, then teaching this ability independently from content organization may not promote effective problem representation. Again, this stresses the importance of early and integrated instructional efforts where students are given the opportunity to acquire knowledge and build retrieval pathways that are useful in representing and solving problems.[10]

As in other similar studies,[11] the naturalistic teaching setting in which this study was conducted imposed a number of limitations and restrictions. The statistical power of some analyses was lower than initial estimates based on existing studies of semantic competence. This was due to larger standard deviations than expected and consequently makes the interpretation of negative results more difficult. Secondly, due to the ECM course structure, the intervention was limited to three workshops spread over an 8-month period. Because of these organizational constraints, additional efforts were made to increase the intensity of the intervention through exercises, electronic reminders, and dissemination of formal articles on problem representation. Also, the period of time between the last instructional workshops (workshop 3) and the assessment of outcomes (workshop 4) was long (up to 100 days). During that period, the students had other concurrent activities, such as clinical workshops and hospital *practicum*, that did not necessarily reinforce problem representation and prototypical differential diagnoses. Consequently, while the students may have applied the jargon acquired (the SQs) during the workshops, they did not use it to gain insight into specific problems and to compare and contrast diagnoses. This phenomenon has been observed in students going from the pre-clinical to clinical years.[12] Thirdly, given the limited number of teachers trained in cognitive models of clinical reasoning (such as problem representation and SQs) and the constraints of the course structure and organization, only two student groups were solicited and only two cases could be used for outcome assessment.

According to students, the new instructional method met the criteria for successful learning and compared favourably with other instructional activities in terms of satisfaction, learning, feedback, and effectiveness. The structured approach to clinical reasoning fostered during the instructional intervention probably contributed to the positive evaluations. Students had the opportunity to engage in concrete steps that made the reasoning process more conscious and the pertinence of content knowledge more directly related to its use in practice.

Conclusion

There is a need to implement and evaluate educational strategies based on psychological models and theories. Previous studies in medicine and other fields have underlined the importance of enhancing problem representation in order to gain deeper understanding of problems and to improve diagnostic performance. Semantic abstractions as mediators of problem representation have been associated with better diagnostic outcomes. Results from the present study show that a short instructional intervention can enable students to use SQs and recall patient findings better. However, the use of SQs did not enhance students' data interpretation and diagnostic accuracy, thus questioning the hypothesis that merely using SQs for summarizing a case leads to better problem representations. Teaching problem representation should probably occur at an early stage, namely during the process of knowledge acquisition and organization, and should focus on eliciting and recognizing key patient findings. Finally, while successful diagnosticians do use more SQs in their case discussions, this may not represent an independent factor that explains diagnostic competence, but an index of more organized knowledge that would be more difficult to teach. The naturalistic setting used in this study imposed a number of limitations and restrictions that call for further research to test whether instructional efforts should also emphasize the recognition of key patient findings and knowledge representation.

Contributors

This study was designed and data collected, interpreted and analysed by both authors. The manuscript was written by MRN and successively reviewed by GB.

Acknowledgements

We wish to express our sincere appreciation to the students who so willingly participated in this study and to the course directors and co-ordinators, case developers and standardized patients for their time and dedication.

Funding

This study was conducted while MRN was a fellow at the Department of Medical Education, College of Medicine at the University of Illinois at Chicago. His fellowship was made possible by grants from the University Hospital of Geneva, Switzerland and the Elie Safra Fund.

References

1 Goss JR. Teaching clinical reasoning to second-year medical students. *Acad Med* 1996;**71**:349–52.

2 Solomon DJ, Speer AJ, Perkowski LC, DiPette DJ. Evaluating problem solving based on the use of history findings in a standardized-patient examination. *Acad Med* 1994;**69**: 754–7.

3 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning*. Cambridge MA. Harvard: Harvard University Press; 1978.

4 Kassirer J, Gorry G. Clinical problem solving: a behavioural analysis. *Ann Intern Med* 1978;**89**:245–55.

5 Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med* 1982;**5**:49–55.

6 Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. *Acad Med* 1991;**66**:S70–2.

7 Chang RW, Bordage G, Connell KJ. The importance of early problem representation during case presentations. *Acad Med* 1998;**73**:S109–11.

8 Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med* 1994;**69**:883–5.

9 Williams RG, McLaughlin MA, Eulenberg B, Hurm M, Nendaz MR. The Patient Findings Questionnaire: one solution to an important standardized patient examination problem. *Acad Med* 1999;**74**:1118–24.

10 Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Acad Med* 1996;**71**:988–1001.

11 Oddone EZ, Waugh RA, Samsa G, Corey R, Feussner JR. Teaching cardiovascular examination skills: results from a randomized controlled trial. *Am J Med* 1993;**95**:389–96.

12 Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: studies of reliability and concurrent validity. *Acad Med* 1997;**72** (**10** Suppl.):S37–S39.

Received 31 *January* 2001; *editorial comments to authors* 29 *May* 2001; *accepted for publication* 26 *March* 2002