QUANTITATIVE CONCEPT MAPPING IN PULMONARY PHYSIOLOGY: COMPARISON OF STUDENT AND FACULTY KNOWLEDGE STRUCTURES

William C. McGaghie, Donald R. McCrimmon, Gordon Mitchell, Jason A. Thompson, and Michael M. Ravitch

Quantitative concept mapping, in contrast with qualitative approaches, is rigorous scientifically and permits statistical analyses of data about concept learning. This study extends past quantitative research on the structure of student concept learning in pulmonary physiology. Pathfinder scaling is used to derive concept maps for medical and veterinary students and their physiology instructors at Northwestern University and the University of Wisconsin, respectively. The concept maps are evaluated for coherence (internal consistency), student-instructor similarity, and correlation of similarity with final examination scores. Results show that student and instructor concept maps are coherent and that student concept maps become increasingly similar to instructors' concept maps from pre- to postinstruction, but that student-instructor concept map similarity does not correlate with examination performance. Research outcomes are discussed concerning possible sources of variation in student and faculty knowledge structures.


Key words: physiology education; mental model building

Research in the health sciences including studies in physiology (17, 18, 20) and medicine (5, 14, 29) has tried to capture or represent the cognitive structure, the intellectual organization, of concepts acquired by learners at different educational levels. This research is based on two fundamental assumptions. First, the ability to organize, simplify, and store a large body of educational material in a specific field is essential for the development of expertise (8). The coherence of information increases as individuals learn to organize and simplify data and concepts for storage in memory. Meaningful learning involves the ability to structure educational material, not just the ability to acquire material in volume and recall it later on multiple-choice tests. Second, the organization of concepts in memory mediates their recall and use. As the conceptual organization of educational material is simplified, its retrieval from memory becomes automatic and the cognitive workload needed for recall is reduced. Thus, as reported by Nobel Laureate H.A. Simon in his landmark book, Models of Thought (27), the ways in which experts and learners in a field of study intellectually structure its key concepts affect their ability to retrieve and use the material.

Studies in medical education have been at the leading edge of research on knowledge structures. This research has been done using both qualitative and quantitative methods. Qualitative methods include hand-drawn maps, where experts and learners portray
their intuitions about structural relations among concepts in a field of study (6, 16), and the use of “thinking aloud” protocols in the cognitive science tradition (2, 7).

By contrast, quantitative methods including multidimensional scaling (MDS) (5, 14), assessment of semantic structures (3, 4), and concept network analysis using the Pathfinder scaling algorithm (17, 18) have been used to study knowledge structures in medicine, basic biomedical science, and other fields. These quantitative methods are rigorous, can be replicated, and allow for statistical analysis of concept data sets. For example, such analyses permit investigators to systematically study issues including expert-novice differences in concept structure (1, 13, 18), changes in concept structure due to educational interventions (11), correlations between indexes of concept structure and independent measures of educational achievement (1, 13–18), and variation among experts in concept structure (17). Although individual studies within this body of research vary in scope, measurement methods, data analysis procedures, and scientific rigor, they all address the research goal of organizing medical and basic science information with coherence and parsimony.

Pathfinder scaling, grounded in graph theory (24), is one member of a family of indirect quantitative methods for representing the structure of knowledge in an intellectual domain. It is termed indirect because response data come from rating or sorting tasks rather than directly from verbal reports or observations of expert or learner performance (22). Pathfinder uses pairwise judgments of similarity between a set of concepts (e.g., partial pressure, respiratory mechanics) to produce a network, or concept map, that shows the latent structure of a conceptual domain for individuals or groups. Pathfinder networks represent concepts as nodes and relationships between concepts as network links. The algorithm takes proximity values as input and yields a concept map with shortest-path link distances as output (15, 25).

Pathfinder networks of concept structure differ from those derived by MDS, even though both scaling methods rely on the same data type and format, i.e., pairwise judgments of concept similarity. Gonzalvo et al. (Ref. 13, p. 602) state, “Although both MDS and Pathfinder reduce a large amount of proximity data to an interpretable form, they achieve this goal by using different mechanisms that tend to highlight different aspects of the underlying structure.” Specifically, Pathfinder networks are different from MDS results because 1) the theoretical and mathematical algorithms underlying the scaling procedures are not the same, 2) Pathfinder output is in two dimensions compared with N-dimensional MDS solutions, and 3) Pathfinder focuses on local relationships among concepts, whereas MDS captures more global information about the concept space (13).

The similarity of two Pathfinder networks having a common set of nodes can be measured by the index \( PFC^* \), which indicates how many links between concept pairs they have in common (10). For example, consider two networks, each with four nodes (A, B, C, and D). Network 1 has three links (A-B, A-C, and A-D), and network 2 has three links (A-B, A-D, and B-C). The \( PFC^* \) for these two networks would be 0.50, the ratio of the number of links the networks have in common (two: A-B and A-D) to the four links in either network. Another index, coherence, indicates the “internal consistency” of a network by comparing direct and indirect measures of the similarity of each concept pair. For each pair of items, an indirect index of relatedness is computed by correlating the proximities between the items and all other items. Coherence is then computed by correlating the original proximity data with the indirect measures (12).

Our previous Pathfinder scaling research addressed medical student learning of 13 basic concepts in pulmonary physiology (e.g., alveolar ventilation, gas exchange). These studies found that 1) student concept maps derived quantitatively become increasingly similar to faculty concept maps as a consequence of instruction, 2) an index of student-faculty concept similarity correlates with multiple-choice examination performance, and 3) much variation exists among individual faculty experts and faculty groups in the way they intellectually structure the physiology concepts (17, 18).

This study extended our research group’s earlier work on mapping concept structures in pulmonary physiology in two ways. First, this study involved a larger and more diverse sample of research participants than our
earlier research. Second, it applied Pathfinder scaling to a knowledge structure problem involving a slightly different and smaller set of 12 concepts in pulmonary physiology than we used before. The concepts were chemoreceptors, lung gas exchange, ventilation, spinal cord, perfusion, intrapleural pressure, respiratory mechanics, surface tension, resistance, control of breathing, diffusion, and partial pressure. Concept maps were derived from individual medical students and veterinary students before and after a period of focused instruction. Student concept maps were then evaluated for similarity with maps obtained from faculty experts, a research tactic that replicates and extends past research (1, 11, 13, 17, 18). We hypothesized that as a consequence of instruction, students’ concept maps would approximate the structure of concept maps produced by faculty who were teaching the students. Finally, an index of student-faculty concept map similarity was correlated with students’ pulmonary physiology course examination scores. This analysis sought to establish the correspondence of structural learning measured by Pathfinder concept maps and student learning measured by objective examinations.

METHODS

Participants. One hundred fifty-three Northwestern University medical students and seventy-six University of Wisconsin veterinary medical students volunteered to participate in the study during the fall of 1996. They supplied preinstruction and, with some attrition, postinstruction data. This occurred in the context of a first-year physiology course section or a complete course at each institution, respectively.

Measurements. A questionnaire was constructed that presented all possible \( \frac{n(n - 1)}{2} = 66 \) pairs of the 12 concepts randomized for presentation order (left-right) and sequence (1–66). Instructions directed the research participants to provide a judgment about the relatedness of the two concepts in each pair (degree of similarity) on a scale of 0 (completely unrelated) to 9 (highly related). This judgment task was completed on both occasions by most of the medical and veterinary medical students in ~20 min. Students at each institution also completed a final examination composed of objective questions [multiple-choice questions (MCQs) at Northwestern; MCQs, short answer, and computation items at Wisconsin] as a customary measure of knowledge acquisition. This complements, but does not duplicate, the structural knowledge assessments obtained from the pre- and postinstruction concept questionnaire.

Data collection. Northwestern medical students and Wisconsin veterinary medical students completed the questionnaire in class, immediately before and after an instructional unit on pulmonary physiology that lasted three weeks. Students took the final examination during the last class session at each institution. The concept questionnaire was also completed by the two course instructors, each on a single occasion.

Data analysis. Data analysis proceeded in five steps. First, Pathfinder network analyses were performed for each student’s concept similarity judgments, yielding individual pre- and postinstruction concept maps. A coherence index (range from 0.00 to 1.00) was calculated for each individual concept map. Next, mean (with standard deviation) coherence indexes were calculated for each of four student subgroups: 1) Northwestern preinstruction, 2) Northwestern postinstruction, 3) Wisconsin preinstruction, and 4) Wisconsin postinstruction. Second, Pathfinder network analyses were performed individually on the concept similarity judgments from the two course instructors. These analyses yielded faculty referent structures for comparison with student maps. A coherence index was also calculated for each faculty concept map. Third, chance-corrected similarity measures were computed comparing each student concept map with the concept map produced by each student’s course instructor (10). Mean (with standard deviation) chance-corrected similarity indexes were also derived for the four student subgroups. Fourth, ANOVAs were done using the indexes of coherence and similarity as dependent variables. The ANOVAs aimed to detect within-group (Northwestern, Wisconsin) pre-to postinstruction changes and between-group differences on the two occasions. Fifth, student-instructor similarity scores were correlated with student final examination scores as an index of correspondence of structural learning with learning measured by MCQ tests.

RESULTS

Descriptive statistics obtained from the analyses of student data are presented in Table 1. Pre- and
postinstruction outcomes are shown for Northwestern medical students and Wisconsin veterinary medical students concerning 1) response rates, 2) concept map coherence, 3) student-instructor similarity, and 4) similarity-final examination correlations.

Attrition from the study occurred for the volunteer students in both groups. For the Northwestern medical students, the preinstruction to postinstruction loss was from 153 to 115 students, a drop of 24%. For the Wisconsin veterinary medical students, the preinstruction to postinstruction loss was from 76 to 67, a 12% reduction. Students lost to follow-up either were not present or did not complete the concept questionnaire at the course final examination, or else they declined participation when contacted subsequently.

Pathfinder concept networks or maps are shown in Fig. 1 for the Northwestern medical course instructor (A) and postinstruction for a medical student having a highly similar concept structure (B) and a medical student with a very dissimilar concept structure (C). Matching concept maps are given in Fig. 2 for the Wisconsin veterinary medicine course instructor (A) and postinstruction similar (B) and dissimilar students (C), respectively.

The Northwestern instructor’s conceptual structure of the 12 pulmonary physiology concepts, shown in Fig. 1A, has three distinct substructures. A single conceptual link joins the three units, indicating a relationship between each physiological subsystem.

Figure 1B shows a student concept network that is similar to the instructor’s. However, two differences are present. First, only one conceptual link, represented by a broken line, is missing. Partial pressure and perfusion are not conceptually linked for this student, although both belong to a substructure otherwise similar to the instructor’s. Second, several new links are present, each represented by a bold line. Three of these links fall within conceptual structures similar to the professor’s: 1) surface tension-resistance, 2) intrapleural pressure-resistance, and 3) perfusion-diffusion. By contrast, three other links are between concepts in different substructures: 1) respiratory mechanics-chemoreceptors, 2) lung gas exchange-chemoreceptors, and 3) ventilation-perfusion. These links suggest that each conceptual subsystem is less distinct for this student than for the course instructor. This finding concurs with theories of expert versus novice conceptual knowledge, which assert that experts have more parsimonious knowledge structures of a domain than novices and more selectively represent relationships between concepts.

The concept maps shown in Fig. 1, A and B, are easily contrasted with the concept map in Fig. 1C, a representation from a very dissimilar student. A very different structure of the 12 pulmonary physiology concepts is seen for this student. Only 3 of the 12 links are common with the professor’s network, and the three subsystems do not emerge.

Similar findings were observed in the Wisconsin data, in which the instructor’s concept network (Fig. 2A) is in close agreement with a student’s similar network (Fig. 2B) and in sharp contrast with a veterinary medical student’s very dissimilar network (Fig. 2C).

Coherence indexes for the concept maps from the Northwestern and Wisconsin course instructors are 0.68 and 0.57, respectively.

ANOVA demonstrates that the coherence of students’ concept maps at Northwestern increased significantly (P < 0.05) as a consequence of instruction in respiratory physiology (Table 1). However, the analysis did not show a significant increase in coherence for the Wisconsin students. The data also show that Northwestern students’ preinstruction concept map coherence did not differ from that of their Wisconsin

---

**TABLE 1**

Descriptive statistics for student data

<table>
<thead>
<tr>
<th></th>
<th>Northwestern</th>
<th>Wisconsin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preinstruction</td>
<td>Postinstruction</td>
</tr>
<tr>
<td>Response rate</td>
<td>153</td>
<td>115</td>
</tr>
<tr>
<td>Concept map coherence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean SD</td>
<td>0.28 (0.28)</td>
<td>0.35 (0.26)</td>
</tr>
<tr>
<td>Student-instructor similarity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean SD</td>
<td>0.07 (0.09)</td>
<td>0.22 (0.11)</td>
</tr>
<tr>
<td>Similarity-final examination correlations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIG. 1.
Pathfinder concept maps from the Northwestern medical course. A: professor’s concept map; B: postinstruction student’s similar concept map; C: postinstruction student’s dissimilar concept map.
FIG. 2.
Pathfinder concept maps from the University of Wisconsin veterinary medical course. A: professor's concept map; B: postinstruction student's similar concept map; C: postinstruction student's dissimilar concept map.
counterparts. However, the mean postinstruction coherence of Northwestern students’ concept maps is significantly higher than the coherence of Wisconsin students’ concept maps \((P < 0.0001)\).

Different outcomes were obtained for the chance-corrected indexes of student-instructor concept map similarity (Table 1). ANOVA reveals significant increases in pre- to postinstruction student-instructor similarity at both Northwestern \((P < 0.0001)\) and Wisconsin \((P < 0.0001)\). The data also demonstrate that student-instructor concept map similarity was significantly higher at Wisconsin than at Northwestern before instruction \((P < 0.01)\). However, after students received the respiratory physiology instructional units, Northwestern students’ concept maps were significantly more similar to their instructor’s map than those of their Wisconsin peers \((P < 0.05)\).

The chance-corrected similarity index for the concept maps from the Northwestern and Wisconsin course instructors is 0.37 (maximum \(= 1.00)\).

Finally, correlations involving the course final examinations at Northwestern \((KR20\text{ reliability} = 0.79)\) and Wisconsin \((KR20\text{ reliability} = 0.86)\) with the pre- and postinstruction index of student-instructor similarity did not yield statistically significant \((P < 0.05)\) results (Table 1).

**DISCUSSION**

Figures 1A and 2A present the physiology course instructors’ maps of the 12 pulmonary concepts as simple, two-dimensional maps. The maps are expressions of each instructor’s tacit knowledge about the intellectual organization of the concepts \((12)\). The instructors’ concept maps can be used as “advance organizers,” i.e., teaching targets, before educational interventions are started for the medical and veterinary students at Northwestern and Wisconsin, respectively. Future iterations may show that the instructor maps can be used as postinstruction “gold standards,” i.e., benchmarks to gauge student learning. This potential use of faculty concept maps for student evaluation would be similar in purpose but different in format from setting minimum passing standards for course examinations \((21)\).

The quantitative coherence, i.e., internal consistency, of the student and faculty concept maps is informative. Northwestern students achieved a significant improvement in concept map coherence from pre- to postinstruction that was not matched by Wisconsin students. However, neither student group matched the concept map coherence of their respective faculty experts. The differences between student groups could result from some combination of student aptitude \((28)\), student learning efficiency, or teaching time allocation. Specific reasons for the differential increase in concept map coherence for students at the two universities remain to be identified.

Concept maps from the two respiratory physiology instructors, similar professors who have won student teaching awards at major universities, were similar quantitatively \((0.37 \text{ of } 1.00)\), but not identical. This research outcome, coupled with findings from other studies, underscores the idea that experts disagree \((17)\). It is also a reminder that the structure and meaning of scientific knowledge is constructed by individuals rather than existing as an independent, objective entity \((26)\). Among experts, even in a narrow scholarly domain, the structure of knowledge has not only central tendency but much variation \((15, 17)\). There is richness and diversity of understanding about basic biophysiological concepts rather than singular meaning. This suggests that considerable interindividual variation exists in the organization of respiratory concepts without apparent compromises in problem-solving ability. Nevertheless, when the concept maps were shared, the instructors were able to appreciate the logic of each other’s concept organization.

There are several alternative explanations for the observed differences in concept maps. One is that the particular terms chosen do not optimally represent basic concepts in respiratory physiology. The two instructors had worked together to identify terms that spanned the concepts generally taught in pulmonary physiology at this level. However, in trying to span the field, terms (e.g., respiratory mechanics, chemoreceptors, and perfusion) may have been picked that can reasonably be thought of as fairly closely related or only distantly related.
A second possible explanation for the interindividual differences in concept maps is related to the context in which an individual is thinking about pulmonary physiology at the time the concept similarities are rated. For example, chemoreceptors could be considered in terms of the reflex changes in minute ventilation they elicit. Alternatively, chemoreceptors could be considered in terms of how they are stimulated by changes in partial pressure of oxygen or carbon dioxide. As in the case of the two instructors, the resulting maps could differ greatly between individuals on the basis of the relative proximity of minute ventilation and partial pressure.

The quantitative similarity of student-instructor concept maps is also instructive. Students at both Northwestern and Wisconsin achieved highly significant boosts in conceptual similarity with their instructors after receiving the educational interventions. Before instruction, the similarity index significantly favored the Wisconsin program, whereas after instruction the similarity of student-instructor concept maps was significantly higher at Northwestern. It is unlikely that teaching was a major contributor to these differences. The number of lecture hours at each institution was similar, and a comparison of the lecture notes given to the students showed that they were very similar. However, there were substantial differences with respect to when during the academic year that pulmonary physiology was taught. At Wisconsin, pulmonary physiology was presented in September at the beginning of the academic year. By contrast, at Northwestern it was taught in December, following introductory coursework on cellular function, histology, and gross anatomy as well as the physiology of the cardiovascular and renal systems. The greater exposure that Northwestern students had to physiology-relevant material may have contributed to their larger change in response to teaching and the increased coherence these students exhibited with the pulmonary physiology material.

Identification of the sources of interindividual variation in the concept maps is beyond the scope of this article, but two approaches to isolating such sources are possible. First, different concept terms may result in less variability. The instructors worked together to identify reasonable terms, but a trial-and-error process may be needed to optimize the selection of appropriate concept terms. Second, variability may be limited if a particular context could be established in which the terms (and pulmonary function) are to be considered. For example, subjects could be asked to relate the similarity of the terms with respect to their direct impact in determining the decrease in minute ventilation in a patient with severe asthma.

Data from this study do not show a correlation between student-instructor concept map similarity scores and scores from objective examinations measuring similar material. This is so despite results showing that student concept maps in the aggregate are internally consistent and increase significantly in similarity with maps from course professors following instruction. Such findings contrast with research in other academic fields, where strong correlations have been shown between measures of student structural knowledge and customary measures of educational achievement (11, 13, 18). Each approach to student assessment in pulmonary physiology (structural, examination scores) yields very solid data, but the measurements have no common variance. More research is needed to determine whether the absence of correlations in this study is due to unreliable measurements, which is not likely, or whether the measures are probing fundamentally different traits (9).

This report also underscores the need to use a variety of approaches to study student structural learning, including Pathfinder scaling and multidimensional scaling (MDS) (30). A substantial portion of the data presented in this article has been analyzed using MDS and reported elsewhere (19). That companion article portrayed the 12 pulmonary physiology concepts in four dimensions, in contrast with the two dimensions displayed here. However, it reached similar conclusions about the excellent internal consistency of the concept similarity ratings when subjected to MDS analyses. The concept similarity ratings (expressed as MDS dimension weights) also did not correlate with objective examination scores, just as the index of student-instructor concept map similarity used in this article did not correlate with examination scores. Finally, the former article did not address student-instructor concept similarity, especially as a referent structure for student achievement (19).
The limits of this study are obvious. They include potential bias due to student attrition, focus on a narrow intellectual domain in the medical basic sciences, and involvement of a narrow range of students in two similar but not identical health professions. At this point it is premature to conclude how helpful concept mapping may ultimately be in defining exactly how students and instructors store information in physiology and other basic medical sciences (23). There is recognition, however, of the great variation in how this information is stored. More studies are needed to uncover the sources of this variation. Research is especially needed to determine whether variation in information storage is maintained when the research context used to study pulmonary physiology (or another system) is constrained to limit concept linkages.

This research was funded in part by grants from the Washington Square Health Foundation, the Charles E. Culpeper Foundation, and the Bowman C. Lingle Trust.

Earlier versions of this report were presented at the annual meeting of the Association of American Medical Colleges, Washington, DC, July 1999; the annual meeting of the American Educational Research Association, San Diego, CA, April 1998; and a meeting of the International Association of Medical Science Educators, Washington, DC, July 1999.

Address for reprint requests and other correspondence: W. C. McGaghie, Northwestern Univ. Medical School, Office of Medical Education, 3–130 Ward Bldg., 303 E. Chicago Ave., Chicago, IL 60611-3008 (E-mail: wcmc@northwestern.edu).

Received 14 September 1999; accepted in final form 1 March 2000.

References


22. Olson JR and Biolsi KJ. Techniques for representing expert knowledge. In: Toward a General Theory of Expertise, edited


