Critical Thinking in Graduate Medical Education
A Role for Concept Mapping Assessment?

Daniel C. West, MD
J. Richard Pomeroy, PhD
Jeanny K. Park, MD
Elise A. Gerstenberger, MD
Jonathan Sandoval, PhD

During the past few decades medical curricula have increasingly reflected the belief that students should understand concepts, as opposed to rote memorization of facts or procedural algorithms. Many medical schools have adopted alternative teaching and learning strategies that challenge students to become self-directed learners who continuously expand the limits of their knowledge and seek evidence for clinical decision making. The root of these ideas lies in the constructivist theory of learning, which states that meaningful learning occurs when students assimilate new knowledge into a preexisting conceptual framework. This evolving conceptual framework supports expertise in critical thinking and clinical problem solving. Despite its critical importance in the education of future physicians, tools to evaluate the quality of students’ conceptual frameworks are limited because of the difficulty in designing and administering objective measures.

In this article, we describe an application of concept mapping assessment (CMA) that provides a “window” into learners’ organization of knowledge at various points in training. Concept mapping is a technique developed by Joseph Novak in the 1970s in which students create a 2-dimensional diagram outlining their understanding of the relationships between and among important concepts within a given subject area. Unlike most conventional ex-

Context Tools to assess the evolving conceptual framework of physicians-in-training are limited, despite their critical importance to physicians’ evolving clinical expertise. Concept mapping assessment (CMA) enables teachers to view students’ organization of their knowledge at various points in training.

Objective To assess whether CMA reflects expected differences and changes in the conceptual framework of resident physicians, whether concept maps can be scored reliably, and how well CMA scores relate to the results of standard in-training examination.

Design, Setting, and Participants A group of 21 resident physicians (9 first-year and 12 second- and third-year residents) from a university-based pediatric training program underwent concept map training, drew a preinstruction concept map about seizures, completed an education course on seizures, and then drew a postinstruction map. Maps were scored independently by 3 raters using a standardized method. The study was conducted in May and June 1999.

Main Outcome Measures Preinstruction map total scores and subscores in 4 categories compared with postinstruction map scores; map scores of second- and third-year residents compared with first-year residents; and interrater correlation of map scores.

Results Total CMA scores increased after instruction from a mean (SD) preinstruction map score of 429 (119) to a mean postinstruction map score of 516 (196) (P = .03). Second- and third-year residents scored significantly higher than first-year residents before instruction (mean [SD] score of 472 [116] vs 371 [102], respectively; P = .04), but not after instruction (mean [SD] scores, 561 [203] vs 456 [179], respectively; P = .16). Second- and third-year residents had greater preinstruction map complexity as measured by cross-link score (P = .01) than first-year residents. The CMA score had a weak to no correlation with the American Board of Pediatrics In-training Examination score (r = 0.10-0.54). Interrater correlation of map scoring ranged from weak to moderate for the preinstruction map (r = 0.51-0.69) and moderate to strong for the postinstruction map (r = 0.74-0.88).

Conclusions Our data provide preliminary evidence that concept mapping assessment reflects expected differences and change in the conceptual framework of resident physicians. Concept mapping assessment and standardized testing may measure different cognitive domains.

©2000 American Medical Association. All rights reserved.
CONCEPT MAPPING ASSESSMENT

A Contri... CMA may provide unique insights into how an individual organizes his or her knowledge or comes to think about a problem. While concept mapping has been used frequently as a learning tool, especially in high school and college science education, it has only rarely been used as an assessment method, primarily because there is limited information regarding its validity and reliability.12,15-18

Therefore, we undertook this preliminary study to assess whether CMA can measure predicted differences and expected changes in the conceptual framework of resident physicians. In addition, we sought to determine whether concept maps could be scored in a reliable way.

METHODS
Study Subjects and Design
All pediatric residents (33 total) at our institution were eligible to participate in this study, which was approved by our institutional human subjects committee. Written informed consent was obtained. A total of 21 residents enrolled: 9 first-year residents, 8 second-year residents, and 4 third-year residents. Each participant underwent concept map training followed by drawing a preinstruction concept map about the topic seizures. They then participated in a 3-session seizure education course followed by drawing a postinstruction map about seizures.

All study subjects participated in a standardized concept map training exercise based on previously described methods.22 Within 2 to 3 weeks of completing map training, participants were given 30 minutes to independently, and without study aids, draw their concept of seizures on a blank sheet of paper (preinstruction map). Residents then participated in a 3-session (1 hour per session) seizure education course consisting of case-based instruction. One week after completing the course, residents drew a postinstruction concept map about seizures under the same conditions as before.

Concept Maps
We used a modification of the “hierarchical technique” of concept mapping in which a concept map consists of 5 components: concepts, concept links, hierarchies, cross-links, and examples (see Figure 1 for an example concept map of water).14,19

A concept is defined as a perceived regularity in events or objects designated by a label and, when depicted on a map, enclosed with a circle. Creating a concept map involves connecting related concepts using arrow lines with a statement or proposition written above the line describing how the concepts are related.

A concept linked to another concept by an arrow line with a propositional statement, called a concept link, represents the foundational unit of a concept map.20-22 In Figure 1, the concept water connected to living things by the proposition needed by represents a concept link. The expression of the hierarchy between concepts is indicated by the direction of the arrow in the concept link with the most general concepts arranged at the top and more specific concepts arranged hierarchically below. Three levels of hierarchy are represented in Figure 1. The linking of a series of related concept links in a hierarchical fashion represents a domain of knowledge. Cross-links are the method by which the map author indicates how different domains of knowledge are related to each other and are a critical measure of map complexity. In a manner similar to propositions, cross-links are labeled with a phrase that describes the relationship of the domains. In the example map, motion connected to states by the proposition determines defines a cross-link.

Examples are labeled and connected to the related concept in the most subordinate position on the map. In Figure 1, Lake Tahoe is a specific example of a lake. Theoretically, as a learner’s conceptual framework changes, he or she links new propositions sequentially (concept links), general to specific (hierarchy), and across domains (cross-links), thus demonstrating integration of knowledge.22

Scoring Method
Maps were scored independently by 3 different raters, blinded to the identity

Figure 1. Example Concept Map

Example map of water demonstrating the basic components of hierarchical concept maps. Note the cross-link (bold arrow) between the concepts motion and states. Reproduced with permission from Cambridge University Press.15

©2000 American Medical Association. All rights reserved.
of the map author, and using a modification of the structural scoring method previously described. Each rater was a board-certified pediatrician who underwent a 30-minute concept map scoring training exercise immediately prior to grading the maps. Maps were scored in 4 categories: concept links, hierarchy, cross-links, and examples. The value of each scoring category represents the depth of thinking required to form the category. Each valid concept link was given 2 points, and each level of hierarchy was given 5 points. A cross-link was given 10 points, while an example counted for 1 point. Invalid links or concepts were given 0 points. A total score and subscore for each category (concept links, cross-links, hierarchy, and examples) were generated for each map. Standard total scores and standard subscores consisting of the sum of individual rater total scores and subscores, respectively, were assigned to each map.

Data Analysis
Statistical analysis was done using STATA, version 6.0 (College Station, Tex). The correlation of each rater's total scores and subscores for each map was tested using Spearman rank correlation. All other data analyses were done using standard total scores and standard subscores. We compared preinstruction map to postinstruction map scores using the Wilcoxon signed rank test. To assess whether there was a difference in map scores based on level of training or experience, we combined the more experienced second- and third-year residents (n=12) into 1 group and compared their scores with those of the first-year (n=9) residents using the 2-sample Wilcoxon rank sum test. To assess the relationship of concept map scores to standard written examinations, we compared the most recent American Board of Pediatrics Intraining Examination (ITE) scores of second- and third-year residents to first-year residents using the 2-sample Wilcoxon rank sum test. First- and second-year residents took the examination within 1 month of completing this study, while third-year residents took the examination 10 months prior to completing the study. We then tested the relationship of concept map scores to ITE scores using Spearman rank correlation.

RESULTS
All 21 participants completed all phases of the study including concept map training, the seizure education course, and both preinstruction and postinstruction concept maps.

Concept Map Scores Based on Level of Training
Figure 2 shows a complete low-scoring concept map from a first-year resident, while Figure 3 shows a high-scoring map from a resident in the second- and third-year group. Note the complexity, with numerous cross-links in the map from the second- and third-year group compared with the linear simplicity of the first-year resident's map. Table 1 summarizes the mean standard total and mean standard subscores of study participants by level of training. The second- and third-year residents had significantly higher total (P=.04) and cross-link scores on the preinstruction map (P=.01) compared with first-year residents. There was no significant difference between concept link, hierarchy, and example scores based on level of training. The postinstruction maps demonstrated that while the total and subscores for second- and third-year residents remained higher, these differences did not reach significance. Of particular note is that the significant difference in total score and cross-link score seen on the preinstruction maps was no longer apparent after instruction (P=.16 and .15, respectively).

Concept Map Scores After Educational Intervention
Total scores increased significantly after the educational intervention (P=.03). In addition, cross-link (P=.02) and concept link (P=.01) subscores also increased significantly. Subscores for hierarchy and examples both decreased slightly, but this change did not reach significance. Preinstruction and postinstruction maps from 1 subject are

Figure 2. Low-Scoring Concept Map

Reproduction of a hand-written concept map of seizures by a first-year resident. Note the absence of cross-links and the consistent use of 2 to 3 levels of hierarchy, resulting in a low score.
Figure 3. High-Scoring Concept Map

Reproduction of a hand-written concept map of seizures by a resident in the second- and third-year group. Note the complexity of the map as manifested by frequent cross-links and the consistent use of 5 to 7 levels of hierarchy, resulting in a higher map score.

Table 1. Comparison of Total Concept Map and Subscores in Second- and Third-Year vs First-Year Residents

<table>
<thead>
<tr>
<th>Component</th>
<th>Second- and Third-Year Residents</th>
<th>First-Year Residents</th>
<th>P     Value†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preinstruction Map</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>472 (116)</td>
<td>371 (102)</td>
<td>.04</td>
</tr>
<tr>
<td>Cross-link</td>
<td>161 (87)</td>
<td>71 (46)</td>
<td>.01</td>
</tr>
<tr>
<td>Concept link</td>
<td>182 (83)</td>
<td>173 (58)</td>
<td>.86</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>68 (25)</td>
<td>59 (11)</td>
<td>.74</td>
</tr>
<tr>
<td>Examples</td>
<td>64 (19)</td>
<td>67 (27)</td>
<td>.80</td>
</tr>
<tr>
<td><strong>Postinstruction Map</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>561 (203)</td>
<td>456 (179)</td>
<td>.16</td>
</tr>
<tr>
<td>Cross-link</td>
<td>215 (136)</td>
<td>131 (111)</td>
<td>.15</td>
</tr>
<tr>
<td>Concept link</td>
<td>233 (78)</td>
<td>211 (78)</td>
<td>.24</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>62 (20)</td>
<td>61 (25)</td>
<td>.59</td>
</tr>
<tr>
<td>Examples</td>
<td>59 (19)</td>
<td>54 (27)</td>
<td>.75</td>
</tr>
</tbody>
</table>

*Data are presented as mean (SD) standard scores. See "Scoring Method" subsection for explanation of scoring method.
†For comparison by 2-sample Wilcoxon rank sum test.

shown in Figure 4. The resident added a new conceptual domain for prevention and education as well as new cross-links to increased intracranial pressure. The vague concept cerebral irritation included as a focal point in the preinstruction map was omitted in the postinstruction map. Table 2 summarizes the mean standard score and subscores of all subjects before (preinstruction map) and after (postinstruction map) the seizure education course. Total scores increased significantly after the educational intervention (P = .03). In addition, cross-link (P = .02) and concept link (P = .01) subscores also increased significantly. Subscores for hierarchy and examples both decreased slightly, but this change did not reach significance.

Concept Map Scores and ITE

The mean (SD) ITE score of second- and third-year residents was 358 (135) and of first-year residents, 243 (94). The ITE scores of second- and third-year residents were significantly higher than those of first-year residents (P = .05). There was weak to no correlation between CMA scores (total and subscores) and ITE scores (r = 0.10-0.54). The strongest correlation was between ITE score and the example score on the preinstruction (r = 0.54; P = .01) and postinstruction map total score (r = 0.53; P = .01). The correlation between ITE scores and postinstruction map total scores, preinstruction map cross-link scores, and postinstruction map cross-link scores was r = 0.29, 0.34, and 0.20, respectively.

Interrater Correlation of Concept Map Scores

The interrater correlation was weak to moderate (r = 0.51-0.69) for the preinstruction map and moderate to strong for the postinstruction map (r = 0.74-0.88). The interrater correlation for subscores demonstrated an almost identical range to that seen in the analysis of total scores.
COMMENT
We found that concept map scores of residents increased significantly following completion of an educational program after which conceptual framework would be expected to change. This change is perhaps best illustrated by qualitative comparison of the preinstruction and postinstruction maps (eg, Figure 4), which demonstrates the incorporation of additional concepts into the postintervention map, further differentiation of existing concepts, and a change in cross-linking between concepts. Quantitatively, these changes are reflected in a statistically significant increase in total map score and concept link and cross-link subscores. Despite the small sample size, we also found that CMA scores of residents with more training and expertise were significantly higher than those with less training. This difference was the result of greater map complexity, as reflected in the significantly higher cross-link scores of second- and third-year residents. Thus, more experienced residents appeared to organize their knowledge differently, expressing links between concepts that were not observed by less experienced residents. It is important to note that this difference narrowed, as one might anticipate, after all residents participated in the same educational intervention. While not conclusive, these findings provide important preliminary evidence supporting our hypothesis that CMA is a valid measure of conceptual framework and the expected change in that framework with new learning.

We also found that while ITE scores of second- and third-year residents were significantly higher than those of first-year residents, these scores did not correlate well with CMA scores. Thus, residents who scored highly on CMA were not the same residents who scored highly on the ITE examination. This finding is consistent with other reports, which note that CMA scores of college students do not correlate with conventional measures of learning, such as final course grades or scores on standardized tests.16 While there may be other explanations, the absence of a positive correlation suggests that CMA measures a different knowledge characteristic than do multiple choice examinations.

The issue of scoring reliability is critically important to any future use of CMA. We found that interrater reliability was moderate to strong and, notably, was greater on postinstruction maps. This level of scoring reliability is similar to that reported by others and suggests that gains in scoring experience lead to increased consistency across raters.23 Therefore, rater training would be important to any future applications.

The vast majority of previous research regarding concept mapping has involved high school and college science education. In reviewing this literature, there is clear evidence that concept mapping can be an effective learning tool in a variety of classroom settings—including the education of health care professionals.

| Table 2. Comparison of Total Concept Map and Subscores Before and After Educational Intervention* |
|-----------------|------------------|------------------|---|
| Component       | Pre-instruction Map | Post-instruction Map | P Value† |
| Total           | 429 (119)         | 516 (196)         | 0.03 |
| Cross-link      | 123 (85)          | 179 (130)         | 0.02 |
| Concept link    | 178 (72)          | 223 (83)          | 0.01 |
| Hierarchy       | 65 (20)           | 61 (22)           | 0.44 |
| Examples        | 65 (52)           | 57 (23)           | 0.06 |
| Examples        | 65 (52)           | 57 (23)           | 0.06 |

*Data are presented as mean (SD) standard scores. See "Scoring Method" subsection for explanation of scoring method.
†For comparison by Wilcoxon signed rank test.
professionals. However, evidence supporting the validity of concept mapping as an assessment tool is much more limited, probably because there is no criterion standard measure of conceptual framework to which CMA can be compared. Traditional measures of competence, such as board scores or even newer competency-based assessments (eg, Objective Structured Clinical Examinations), may not represent appropriate comparisons since they shed little light on how a learner thinks about a problem. Thus, other investigators have used the same strategy to build a case for the validity of CMA. For example, in a variety of secondary school and undergraduate settings, CMA has consistently been found to measure conceptual change in situations in which such change would be expected and to detect differences between students with more training in a particular field compared with those with less. In medical education, we identified only a single related study in which Pathfinder networks, a substantially different form of concept mapping, was used to assess student understanding of pulmonary physiology. In that study, the investigators found that, after instruction, student Pathfinder networks became more similar to the networks generated by faculty experts. These findings appear consistent with our study, and while not conclusive, provide additional evidence suggesting the predictive validity of CMA.

We have provided encouraging preliminary evidence supporting the validity of CMA as a way to measure conceptual change and differences among resident physicians. Our findings also indicate that rater training is critical to the use of CMA. Future research must focus on further validating this method and addressing issues of reliability before CMA can be applied on a larger scale. Nevertheless, CMA has the potential to evaluate how students or residents organize and use knowledge in a way that traditional objective tests cannot. It also could provide a tool for educators to identify unique distortions in students’ understanding of content and to identify errors of omission. Finally, CMA might provide insight into why some residents score well on objective written examinations but have difficulty applying this knowledge to clinical situations.

**Funding/Support:** This work was supported in part by grants from the Children’s Miracle Network Research Fund and the Hibbard E. Williams Research Fund at the University of California, Davis.

**Previous Presentation:** Presented in part at the Pediatric Academic Societies Joint Meeting, Boston, Mass, May 15, 2000.

**Acknowledgment:** We thank Luann Cruys, Debbie Barnes, and Sally Colvin for invaluable administrative support without which this study could not have been completed and Anthony F. Phillips, MD, Department of Pediatrics, for critical review of the manuscript. We also are indebted to Constance Bowe, MD, Department of Neurology, University of California, Davis, for her dedication to medical education and developing and teaching the seizure education course. Finally, we thank the pediatric residents at UC Davis for participating in this study and their willingness to explore a new way of thinking about assessment.

**REFERENCES**

31. Bar-Lave BZ. Enhancing meaningful learning in an environmental education program: a case study of class empowered through the use of Novak’s and Gowin’s principles of learning how to learn, concept mapping, interviewing, and educating. Dissertation Ab- stracts Int. 1998;48:2590A.