Concept mapping assessment in a problem-based medical curriculum

SALAH ELDIN KASSAB1 & SHEREEN HUSSAIN2

1Royal College of Surgeons in Ireland, Medical University of Bahrain, 2Ministry of Health, Kingdom of Bahrain

Abstract

Background: In the problem-based learning (PBL) medical curriculum at the Arabian Gulf University in Bahrain, students construct concept maps related to each case they study in PBL tutorials.

Aim: To evaluate the interrater reliability and predictive validity of concept map scores using a structured assessment tool.

Methods: We examined concept maps of the same cohort of students at the beginning (year 2) and end (year 4) of the pre-clerkship phase, where PBL is the main method of instruction. Concept maps were independently evaluated by five raters based on valid selection of concepts, hierarchical arrangement of concepts, integration, relationship to the context of the problem, and degree of student creativity. A 5-point Likert scale was used to evaluate each criterion. Interrater reliability of the instrument was determined using the intraclass correlation coefficient (ICC) and predictive validity was measured by testing the correlations of concept map scores with summative examination scores.

Results: The ICC of the concept map scores in year 2 was 0.75 (95% CI, 0.67–0.81) and in year 4 was 0.69 (95% CI, 0.59–0.77). Overall concept maps scores of year 4 students were significantly higher compared with year 2 students \((p < 0.001, \text{ effect size } = 0.5)\). The relationship between the students’ scores in concept maps and their scores in summative examination varied from no to mild correlation.

Conclusion: The interrater reliability of concept map scores in this study is good to excellent. However, further studies are required to test the generalizability and validity of assessment using this tool.

Introduction

The development of expertise is not only due to acquiring more knowledge and skills, but also because structuring knowledge is a critical step in the development of novice learners to experts (Boshuizen & Schmidt 1992). The knowledge structure of experts is characterized by elaborate, highly integrated frameworks of related concepts, which facilitate problem-solving and other cognitive activities (Yin et al. 2005). Clinical expertise is achieved by acquiring large amounts of biomedical knowledge structured as concepts linked together in a loosely connected semantic network. The clinical experience then helps the clinician to transform the network-type of knowledge to “illness scripts” that are activated for each disease (Boshuizen & Schmidt 1992).

Concept mapping is a tool which can represent knowledge structure by illustrating the relationships between relevant concepts within a given subject domain. During construction of concept maps, students draw concepts related to a certain domain and link these concepts in a hierarchically organized knowledge framework (Novak & Gowin 1984). Concept mapping is based on the constructivism theory of learning, which indicates that learning is an active process where students actively construct their knowledge (Novak & Gowin 1984). By relating and integrating new knowledge with already existing knowledge structure, learners develop deeper and richer understanding and better use of knowledge (Harris & Alexander 1998).

Problem-based learning (PBL) has been founded on the belief that it produces several cognitive effects essential for training of doctors, including increased retention of knowledge, enhanced integration and application of basic science concepts into clinical contexts with associated development of problem-solving skills, development of self-directed learning skills, and enhancement of students’ intrinsic interest in the subject matter (Dolmans & Schmidt 1996). Concept mapping...
appears to complement PBL in helping medical students to develop organized, integrated knowledge structure, which may help in developing better problem-solving skills.

Although concept mapping could be an important measure of knowledge structure, there is limited information regarding the validity and reliability of concept mapping assessment scores. Concept mapping was introduced in 2004 as part of the PBL undergraduate medical curriculum at the Arabian Gulf University (AGU) in Bahrain. Since then, there was no structured tool for scoring these concept maps. Therefore, this study was designed to develop a structured tool for evaluating the quality of concept maps in the PBL medical curriculum at AGU. In addition, we tried to address the following questions:

1. What is the interrater reliability of concept map assessment scores using this tool?
2. What are the changes in concept map scores as students progress in the PBL curriculum?
3. What is the relationship between the students’s scores in concept maps and their scores in summative examinations (predictive validity)?

Methods

Study context

The College of Medicine and Medical Sciences (CMMS), AGU in Bahrain adopted an integrated problem-based, community-based curriculum since its inception in 1982. The 6-year undergraduate medical curriculum is divided into a phase I (year 1), phase II (years 2, 3, and 4) and phase III (years 5 and 6). The phase II program is organized around nine integrated organ/system units, where PBL is the main strategy of learning. Each week, students study a problem that is designed to cover learning objectives from basic sciences, clinical sciences, and community health domains. At the end of phase II, students undertake a comprehensive summative examination in the different learning domains, which ends with awarding a bachelor degree of basic medical sciences (BSc). The current study included evaluating concept maps of medical students in year 2 (n = 114) during the academic year 2004–2005 and the same cohort when they progressed to year 4 (n = 114) during the academic year 2006–2007. The research protocol was approved by both Curriculum Committee and the Research and Ethics Committee at CMMS, AGU.

Concept mapping process

Students admitted to the AGU receive training on concept mapping as part of the PBL orientation program. In PBL tutorials, students are encouraged by tutors to individually construct a concept map in relation to each problem they study. Students are provided with a study guide for each PBL unit where the guide is designed to serve also as a portfolio for students to construct their concept maps. The learning needs generated by the tutorial group are used as the template for concept map construction in relation to each problem (Figure 1). Concept mapping is part of the continuous evaluation of students in PBL tutorials. In addition, at the end of each PBL unit, students are required to submit concept maps related to any two problems in the unit and assessment scores of these concept maps represent a small percentage of the end-unit summative scores.

Concept mapping assessment

The concept mapping assessment tool is a modified version from previously published literature (Novak & Gowin 1984; Srinivasan et al. 2008). The tool was also reviewed in terms of wording and sampling of the content domain by three faculty members of the Master in Health Professions Education (MHPE) program at AGU. Scoring of each concept map is based on the following five criteria (Table 1):

1. Valid selection of concepts: the degree to which meaningful and valid concepts have been selected from the PBL problem.
2. Hierarchical arrangement of concepts: the degree of arrangement of concepts with more general concepts at the top and more specific below or extending outward.
3. Integration between concepts: the degree to which the map shows meaningful interconnections between the different concepts in the map.
4. Relationship to the context of the PBL problem: the degree to which the concepts are directly related and linked to the context of the problem. The context of the map is the clinical or community problem that the students study during the PBL week.
5. Degree of student creativity: the degree to which the student demonstrates unusual elements that aid communication or stimulate interest without being distracting.

Assessment of concept maps was based on a quality scoring system modified from previously published studies (Ruiz-Primo et al. 2001; West et al. 2002; Srinivasan et al. 2008). Each item of the evaluation form was scored based on a Likert-type scale of 1–5, where 1 is poor and 5 is excellent. An overall score was given based on the total scores of all the five items (out of 25). Each concept map was evaluated independently by five raters who had previous experience for at least 2 years in evaluating concept maps in the PBL program. In addition, these raters have received training about the concept mapping process as part of the faculty development workshops in the college. In addition, the items of the assessment sheet were explained to the raters and instructions for concept map evaluation were available on the evaluation sheet to insure consistency of understanding among them. Each rater scored 228 maps for each year with a total of 456 maps per rater for the 2 years studied. The mean concept map score for each student was calculated by taking the scores of all the raters for each of the two concept maps.

Statistical analysis

The data were entered and analyzed using the Statistical Package for Social Sciences (SPSS) version 17.0 (SPSS, Inc., Chicago, IL, USA). Data were presented as mean ± SD of each
Learning needs from the problem on “convulsions”

- Excitatory and inhibitory neurotransmitters
- Origin of the brain waves.
- Basic principles of EEG
- EEG waves in generalized and focal epilepsies.
- Molecular biology of epilepsy
- Epidemiology of epilepsy
- Classification and differential diagnosis of epilepsy
- Principles of management of seizure disorders
- Anti-epileptic drugs
- Psycho-social impact of epilepsy

Interrater reliability was calculated by using intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) to test for both correlation and agreement among the five raters. Correlation coefficients were interpreted as follows: < 0.4 = poor agreement, 0.4–0.75 = good agreement, and > 0.75 = excellent agreement (Landis & Koch 1977). The relationship between the students' scores in concept maps and their examinations scores was done using Pearson-Product moment-to-moment correlation coefficient. Differences in mean scores of students' concept maps in year 2 and year 4 were analyzed using paired t-test and estimation of the effect size for the differences between the mean scores was calculated, as reported previously (Hojat & Xu 2004). A p-value < 0.05 was considered statistically significant.

**Results**

Interrater reliability of the concept map assessment tool

The total ICC of the concept map scores based on the five raters was 0.75 (95% CI, 0.67–0.81) in year 2 and was 0.69 (95% CI, 0.59–0.77) in year 4 (Table 2). The ICC in the different evaluation criteria ranged from 0.58 (95% CI, 0.37–0.73) to 0.78 (95% CI, 0.71–0.84).

![Figure 1](image-url)  
**Figure 1.** Upper panel: the list of learning needs generated by problem-based tutorial groups studying the "nervous system unit" from a week problem about "epilepsy." Lower panel: a concept map showing the integrated concepts related to the problem on "epilepsy" using learning needs generated in PBL tutorials as a template for map construction.
Table 1. The structured tool for assessment of concept maps in a PBL curriculum.

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
<th>Very good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid selection of concepts</td>
<td>3.20</td>
<td>3.66</td>
<td>4.00</td>
<td>4.30</td>
<td>4.50</td>
</tr>
<tr>
<td>Hierarchical arrangement of concepts</td>
<td>3.00</td>
<td>3.50</td>
<td>4.00</td>
<td>4.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Integration between concepts</td>
<td>3.00</td>
<td>3.50</td>
<td>4.00</td>
<td>4.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Relationship to the context</td>
<td>3.00</td>
<td>3.50</td>
<td>4.00</td>
<td>4.50</td>
<td>5.00</td>
</tr>
<tr>
<td>Degree of student creativity</td>
<td>3.00</td>
<td>3.50</td>
<td>4.00</td>
<td>4.50</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Note: The total mark is the sum of the five sub-scores (out of 25).

Table 2. ICC and 95% CI of the concept map scores of the five raters for years 2 and 4 medical students.

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>ICC Year 2</th>
<th>95% CI</th>
<th>ICC Year 4</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid concepts</td>
<td>0.78</td>
<td>0.71–0.84</td>
<td>0.67</td>
<td>0.56–0.76</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>0.71</td>
<td>0.61–0.78</td>
<td>0.59</td>
<td>0.41–0.67</td>
</tr>
<tr>
<td>Integration</td>
<td>0.69</td>
<td>0.59–0.77</td>
<td>0.53</td>
<td>0.37–0.65</td>
</tr>
<tr>
<td>Relation to the context</td>
<td>0.75</td>
<td>0.66–0.81</td>
<td>0.68</td>
<td>0.58–0.77</td>
</tr>
<tr>
<td>Degree of creativity</td>
<td>0.64</td>
<td>0.52–0.73</td>
<td>0.73</td>
<td>0.65–0.80</td>
</tr>
<tr>
<td>Total score</td>
<td>0.75</td>
<td>0.67–0.81</td>
<td>0.69</td>
<td>0.59–0.77</td>
</tr>
</tbody>
</table>

Notes: Significant differences were tested at the 95% CI. *p < 0.05 in both years.

Table 3. Comparison between concept map scores of students as they progress from Year 2 to year 4 in a PBL curriculum.

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Year 2</th>
<th>Year 4</th>
<th>p-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid concepts</td>
<td>3.83±0.50</td>
<td>4.05±0.48</td>
<td>&lt;0.001</td>
<td>0.5</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>3.62±0.55</td>
<td>3.83±0.54</td>
<td>0.001</td>
<td>0.4</td>
</tr>
<tr>
<td>Integration</td>
<td>3.50±0.54</td>
<td>3.64±0.52</td>
<td>0.016</td>
<td>0.3</td>
</tr>
<tr>
<td>Relation to the context</td>
<td>3.66±0.55</td>
<td>3.85±0.56</td>
<td>0.003</td>
<td>0.3</td>
</tr>
<tr>
<td>Degree of creativity</td>
<td>3.23±0.61</td>
<td>3.55±0.67</td>
<td>&lt;0.001</td>
<td>0.5</td>
</tr>
<tr>
<td>Total</td>
<td>17.87±2.60</td>
<td>19.03±2.44</td>
<td>&lt;0.001</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Notes: The total score of concept map assessment is out of 25. p-value < 0.05 is considered statistically significant and effect size ≥ 0.5 is considered practically significant.

Change in concept map scores as students’ progress in the PBL curriculum

Table 3 shows the differences between overall concept map scores of the students in year 4 compared with their scores in year 2. Overall concept map scores (out of 25 marks) of the students in year 4 were significantly higher compared with their scores in year 2 (19.03±2.44 vs. 17.87±2.60, respectively, p < 0.001, Effect size = 0.46). This difference was statistically significant when comparing all the subscores of the five evaluation criteria.

Table 4. Correlations between concept map scores in years 2 and 4 and BSc summative examination scores.

<table>
<thead>
<tr>
<th>Correlations</th>
<th>MCQ</th>
<th>Integrated short essays</th>
<th>OSCE</th>
<th>OSCE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 2 concept map scores Pearson correlation (r)</td>
<td>0.078</td>
<td>0.240</td>
<td>0.211</td>
<td>0.166</td>
<td>0.287</td>
</tr>
<tr>
<td>p-value (1-tailed)</td>
<td>0.204</td>
<td>0.008*</td>
<td>0.012*</td>
<td>0.039*</td>
<td>0.001*</td>
</tr>
<tr>
<td>Year 4 concept map scores Pearson correlation (r)</td>
<td>0.069</td>
<td>0.107</td>
<td>0.135</td>
<td>0.192</td>
<td>0.164</td>
</tr>
<tr>
<td>p-value (1-tailed)</td>
<td>0.233</td>
<td>0.129</td>
<td>0.076</td>
<td>0.020*</td>
<td>0.040*</td>
</tr>
</tbody>
</table>

Notes: Statistical analysis is conducted by using Pearson-Product moment-to-moment correlation coefficient. MCQ, Multiple Choice Questions, and OSPE, Objective Structured Practical Examination; OSCE, Objective Structured Clinical Examination.

*p < 0.05 and is considered statistically significant.

Relationship between students’ concept map scores and their scores in summative BSc examination

To study the predictive validity of the concept map assessment, we examined the relationship of concept map scores in year 2 and year 4 with final BSc summative examination scores (Table 4). There were no significant correlations between concept map scores and students’ scores in MCQ examination in either year 2 or year 4 students. On the other hand, year 2 concept map scores had significant positive correlation with the students’ scores in OSPE (r = 0.21, p = 0.01), OSCE (r = 0.17, p = 0.04), and integrated short essay questions (r = 0.24, p = 0.01). However, year 4 concept map scores had significant correlation only with OSCE examination scores (r = 0.19, p = 0.02), but the relationships with other BSc examination components were not statistically significant.

Discussion

This study introduces a scoring system for evaluating the quality of concept maps constructed by students in a PBL program. Students construct their maps based on the knowledge gained during the week of the case guided by the learning needs generated in PBL tutorials. The study demonstrates that concept map scores have good to excellent interrater reliability. The progression of students in the PBL curriculum was associated with their ability to generate better-scoring concept maps. However, the concept map scores demonstrated no to mild correlation with summative examination components.

As students progressed from year 2 to year 4 in the PBL program, their concept maps were better in terms of organization of knowledge, understanding of the relationships between concepts, integration between basic and clinical concepts, and creativity. This improvement in concept map scores could be attributed to improved ability of students to draw concept maps as a result of repeated practice (Srinivasan et al. 2008). Alternatively, the PBL process itself could have a
positive effect on the students’ ability to integrate and organize their knowledge structure. Changes in students’ concept maps as a result of educational interventions have been examined by other researchers. Following a focused instruction, students’ maps of the pulmonary physiology concepts became more similar to the concept maps generated by faculty experts (McGaghie et al. 2000, 2004). Concept map scores of residents after an instruction course training about seizures was associated with more incorporation of new concepts into maps, further differentiation of existing concepts, and enhanced cross-linking between domains of knowledge (West et al. 2000). A future study design comparing the knowledge structure, using concept maps, of recent graduates from PBL and non-PBL programs would be interesting.

Students in this study were not provided with the concepts and were given the task to construct maps guided by the learning needs generated in PBL tutorials. This approach has the advantage of capturing the variations in students’ knowledge structure and opens avenues for students’ creativity in constructing the maps. A previous study indicated that scores from maps constructed with selected linking phrases are more reliable than maps constructed with created linking phrases (Yin & Shavelson 2008). Despite the relatively unconstrained concept mapping task, scores generated from the assessment tool in this study have good to excellent interrater reliability. The use of raters who are quite familiar with the PBL process and concept mapping assessment is probably an essential component for obtaining low variability among the raters. We have used a quality scoring system, which takes into account the quality of the propositions and the importance of understanding the subject domain. This is different from the structural scoring system determined by classifying and counting the number of propositions and determining the maximum level of hierarchy. Our finding of the high interrater reliability of this quality scoring system is in agreement with previous reports indicating that this system for concept mapping assessment is reliable while the structural system is not reliable (Ruiz-Primo et al. 2001; Yin et al. 2005; Srinivasan et al. 2008). However, other reports demonstrated that the both scoring systems are reliable, with higher interrater reliability and test–retest reliability for the quality scoring system (West et al. 2000, 2002).

As a measure of the predictive validity of concept map assessment, we examined the correlations between students’ scores in concept maps and their scores in summative BSc examination. There was no correlation between MCQ scores and concept map scores despite the fact that most of the MCQs used in this comprehensive examination are selected to be integrated, context-based (usually a clinical vignette) and usually testing a cognitive domain higher than recall of knowledge. Our finding of the high interrater reliability of this quality scoring system is in agreement with previous reports indicating that this system for concept mapping assessment is reliable while the structural system is not reliable (Ruiz-Primo et al. 2001; Yin et al. 2005; Srinivasan et al. 2008). However, other reports demonstrated that the both scoring systems are reliable, with higher interrater reliability and test–retest reliability for the quality scoring system (West et al. 2000, 2002).

Conclusion

We conclude from this study that the concept mapping scores using the tool provided in this study have a good to excellent interrater reliability. Practicing of concept mapping exercises would help medical students in a PBL program to develop better-scoring maps. It appears that concept map assessment using this tool is measuring a construct that is different from what is measured in summative examination using other assessment tools. Therefore, further studies will be required to test the validity and generalizability of this tool before adopting it in other PBL programs.

Acknowledgments

The authors would like to thank the raters of concept maps Professor Usha Sachdeva, Professor Amer Ansari, Dr Nasir Abdul-Latif, Dr A. Halim Deifalla, and Dr Feisal Subhan. We also appreciate the professional statistical assistance from Mr. Mohammed A. Obeidat.

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

Notes on contributors

SALAH ELDIN KASSAB, MD, PhD, MHPE, is currently a professor of Physiology, Royal College of Surgeon in Ireland, Medical University of Bahrain. He was the assistant dean for Academic Affairs and chairman of the Curriculum Committee, College of Medicine and Medical Sciences, Arabian Gulf University till August 2009.

SHEREEN HUSSAIN, MD, MHPE, is a consultant of Family Medicine, Ministry of Health, Bahrain, and a graduate of the MHPE program at the Arabian Gulf University in Bahrain.

References


