Structural Knowledge Assessment: Change in Cognitive Structure Due to Playing a Serious Game

P. Wouters

Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands.
P.J.M.Wouters@uu.nl

Introduction

Verbal assessment (e.g., knowledge tests, transfer tests) is often used to determine the effectiveness of complex learning environments such as serious games (computer games used for learning and instruction). However, current thinking in cognitive science emphasizes the importance of knowledge organization in learning complex skills [1]. An adequate organization of knowledge in knowledge structures warrants the capability to integrate new information with existing knowledge making the new information meaningful. We propose that structural assessment can be used to measure the quality of knowledge structures.

Structural assessment

The underlying assumption of the structural assessment approach is that we organize our knowledge in knowledge structures containing the important concepts of a domain and the relations among those concepts [2]. Although several methods exist to measure knowledge structures (e.g., ordered-tree techniques, cluster analysis), we have chosen the Pathfinder approach. Two important reasons for this choice were: (1) it does not force a hierarchical structure on the data but identifies meaningful links between concepts, (2) it is a rather straightforward method with predefined concepts, requiring less introspection and cognitive effort from the assessee. Three steps can be distinguished to implement a structural approach to measure and interpret knowledge structures.

1. Knowledge elicitation

Rating the relation between concept pairs is applied in cognitive psychology and is assumed to capture the underlying knowledge organization. The selected concepts are randomly combined and presented in n(n-1)/2 pairs (n is the number of concepts) to the learner who has to rate these on a 5, 7 or 9 point Likert scale. The pair ratings result in a matrix of proximity values between concepts indicating how closely the concepts are related (see Figure 1, left).

	A	**B**	**C**	**D**	**E**
A	0	1	3	2	3
B	1	0	1	4	6
C	3	1	0	5	5
D	2	4	5	0	4
E	3	6	5	4	0

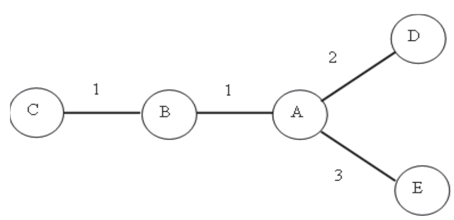

Figure 1. Proximity data (left) and their representation in a PFNet (right).

Pairs A-B and B-C have a distance of 1 and are directly connected because no shorter, indirect path exists (there is no lower link weight sum than 1). A-D has a distance of 2. A direct link connects the concept pair A-D because no shorter path can be found. As an example, consider the potential shorter path in which A-D is connected via A-B and B-D. A-B has a link weighed sum of 1, B-D one of 4 which makes 5 which is not a shorter path. The pairs A-E and A-C have distances of 3. The procedure creates a direct link for the pair A-E, but not for A-C. In the latter case a shorter indirect path from A to C exists via B.

Proceedings of Measuring Behavior 2012 (Utrecht, The Netherlands, August 28-31, 2012)
Eds. A.J. Spink, F. Grieco, O.E. Krips, L.W.S. Loijens, L.P.J.J. Noldus, and P.H. Zimmerman

2. Knowledge representation

The proximity values are difficult to interpret, therefore the elicited knowledge should be represented. The Pathfinder procedure uses a graph-theoretic distance technique (it uses the distance measured by the minimum number of links connecting two nodes in a graph) to represent the proximity matrix in a graphical network structure (a PFNET) [3]. Pathfinder links all concepts and assigns a weigh to each link (based on the ratings). Next, the Pathfinder algorithm removes direct links if there exists a shorter, indirect path that connects both concepts (see Figure 1).

3. Knowledge evaluation

The third step concerns the evaluation of the knowledge representation with a referent knowledge structure (e.g., instructor, expert). For the selection of an appropriate referent, issues such as the agreement on the core concepts in a domain and the number of instructors/experts used to generate the referent structure should be taken into account. In this paper we will use: *similarity* (the number of common links between both networks divided by the total number of links), *coherence* (an indirect measure of similarity by correlating the ratings given for each item in a pair with all of the other concepts) and *graph-theoretic indices of the focal node* (analyze and compare the central concepts).

Methods

Participants and materials

We used the game Code Red: Triage in which advanced (n = 10, M_{age} = 41) and novice players (n = 9, M_{age} = 31) learn in 20 minutes the basic procedure to classify victims according to their injuries. Before and after the game we administered a verbal test (10 mpc items on procedural and declarative knowledge) and structural assessment (based on 13 concepts, 78 pairs, 3 experts were used for the referent structure, 9-point Likert scale).

Results

The verbal test shows that both groups performed better after the game, but the differences before the game between advanced learners and novices had disappeared after the game. A closer look reveals that novices benefit most with regard to procedural knowledge, whereas advanced learners benefit most with regard to declarative knowledge.

The similarity of novices with the referent increased due to the game. This was not the case for advanced learners. The initial larger similarity with the referent for advanced learners -in comparison to novices- disappeared after the game. The coherence scores for novices did not differ from the advanced learners neither before nor after the game. For novices the decrease in coherence after the game was weakly significant, while the decrease was significant for the advanced learners.

Table 1. Results of verbal and structural assessment.

	Novices (N = 9)		Advanced learners (N = 10)	
	Before game	After game	Before game	After game
Knowledge test				
All items	4.00 (1.22)	6.89 (1.76)	7.70 (1.16)	8.30 (1.16)
Declarative items	2.89 (.93)	3.56 (.53)	3.90 (.74)	4.30 (.67)
Procedural items	1.11 (.78)	3.33 (1.50)	3.80 (.92)	4.00 (1.15)
Structural assessment				
Similarity scores	.11 (.07)	.16 (.07)	.24 (.13)	.21 (.08)
Coherence scores	.50 (.12)	.31 (.24)	.49 (.24)	.30 (.20)

	mobility	crt-nail	crt-head	pulse	weather	triage tape	classify	length	breathing	airway	stabilize	circulation	darkness	
mobility		2	4	2	3	3	1	4	2	2	2	3	5	2.8
crt-nail	2		2	1	1	3	1	4	2	2	2	1	3	2.0
crt-head	4	2		2	1	3	3	4	3	2	3	1	1	2.4
pulse	2	1	2		2	1	2	3	1	1	2	1	3	1.8
weather	3	1	1	2		3	2	4	2	2	3	1	2	2.2
triage tape	3	3	3	1	3		2	1	1	2	3	2	4	2.3
classify	1	1	3	2	2	2		3	1	1	1	2	4	1.9
length	4	4	4	3	4	1	3		2	4	4	4	5	3.5
breathing	2	2	3	1	2	1	1	2		2	3	1	3	1.9
airway	2	2	2	1	2	2	1	4	2		1	1	3	1.9
stabilize	2	2	3	2	3	3	1	4	3	1		2	4	2.5
circulation	3	1	1	1	1	2	2	4	1	1	2		2	1.8
darkness	5	3	1	3	2	4	4	5	3	3	4	2		3.3
	5	4	4	3	4	4	4	5	3	4	4	4	5	

median node

center node

Figure 2. Path distance matrix for the averaged expert PFNet.

We computed the three focal nodes: *highest degree node* (the node(s) with the greatest number of links), *median node* (the node(s) with the smallest average distance to all other nodes) and *center node* (the node(s) with the smallest maximum distance to any other node) for all PFnets (see Figure 2 for an example). Based on these nodes we defined the central concepts in the expert's PFNet. Advanced learners, but not novices, show a converging of central concepts with experts after the game.

Conclusion and discussion

The results suggest that structural assessment measures an individual's understanding of a domain at least differently from verbal assessment. While verbal assessment may provide a more nuanced picture regarding declarative and procedural knowledge, structural assessment may add an in-depth understanding of the concepts that are regarded important in a domain. The pattern of results yields at least three questions that should be discussed. First, we failed to find an increase in similarity between the advanced learners and the expert referent after the game. It is possible that there is no 'ideal' referent structure, but that each level of expertise requires a different referent structure (cf. [2], [4]). The second question concerns the high number of links in the PFNets. A possible explanation is that participants have used many extreme relatedness ratings (so many 1s and 9s) or rated many pairs as highly related (8s and 9s). For example, most novices show an increase in extreme relatedness ratings after the training which may have caused many tied values and Pathfinder includes all links when there are tied distances. The third question pertains to the decrease in coherence. It is possible that the lower coherence can be partly attributed to the fatigue and decreasing concentration during the second rating. An alternative explanation may be that the knowledge organization became more differentiated in connected clusters. For example, a cluster around physical body characteristics versus a cluster around environmental, external characteristics. Finally, we propose four guidelines to effectively use structural assessment in serious games:

1. Determine whether the domain allows a referent based structural assessment with similarity (agreement on important concepts in a domain) or that a referent free assessment with coherence is more appropriate.
2. When a referent based structural assessment is chosen, consider carefully which type of referent is most suitable. Especially when different levels of expertise are involved.
3. The concepts should be unambiguous and their number should not be too high (13-20 concepts). Consider to use a subset of a domain when a domain involves more than 20 concepts.
4. Consider the analysis of the graphical knowledge representations to obtain in-depth information about the quality of the knowledge structures.

References

1. Day, E. A., Arthur Jr., W., Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology* **86**, 1022–1033.

2. Kraiger, K., Ford, J. K., Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology* **78**, 311-328.

3. Schvaneveldt, R. W., Durso, F. T., Dearholt, D. W. (1985). *Pathfinder: Scaling with network structures* (Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory). Las Cruses: New Mexico State University.

4. Acton, W. H., Johnson, P. J., Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology* **86**, 303-311.