Structural Representations in Knowledge Acquisition

Pilar Gonzalvo, José J. Cañas, and María-Teresa Bajo

Multidimensional scaling (MDS) and Pathfinder techniques for assessing changes in the structural representation (RP) of a knowledge (KN) domain as a function of learning were compared and evaluated. Relatedness ratings, collected from university students before and after they studied a textbook, were analyzed with both MDS and Pathfinder procedures. Relatedness ratings were collected from the students' instructors, and the similarity between the instructors' structural RPs and those of their students provided evidence of learning. A conceptual test captured the ability of the techniques to predict concept-to-concept learning. Students' networks and dimensional RPs varied depending on whether they had (a) already studied the textbook and (b) learned the material. Comparisons between student and expert similarity measures indicated that multidimensional scaling and graph theoretic approaches are valid techniques. MDS represented global properties of KN, whereas graph networks captured local conceptual relations.

Learning sometimes consists of the acquisition of totally new knowledge, but most of the learning that occurs in life involves either the incorporation of new facts into prior knowledge or the modification of the organizational structures of it (Rumelhart & Norman, 1981). Early studies provided evidence that knowledge acquisition implies modifications in the way in which knowledge is represented and organized. These studies involved comparisons between experts and novices (Adelson, 1981; Chase & Simon, 1973) as well as among students at different stages of learning (Shavelson, 1972, 1974). Researchers investigating the novice-expert shift have argued that the knowledge representation of experts differs from that of novices not only in that experts represent more and different relations among concepts than do novices but also in that experts organize their knowledge in terms of abstract relational dimensions that are not accessible to novices (Chi, Feltovich, & Glaser, 1981; Larkin, 1981; Voss, Green, Post, & Penner, 1983). Similarly, researchers comparing the representation of students' knowledge at two different learning stages have shown that as students learn a discipline, their knowledge of the structural relations among the main concepts of the discipline becomes more like that of the experts (Goldsmith, Johnson, & Acton, 1991; Schoenfeld & Herrmann, 1982; Shavelson, 1972).

A critical question, then, is how can these organizational properties of knowledge be captured and represented? Educational researchers have been concerned with developing ways of assessing structural understanding of concepts within a knowledge domain. Essay examinations are sometimes used to encourage and assess relational understanding. For example, an essay question in which students are asked to contrast and compare two ideas is an attempt by the researcher to evaluate the students' cognitive representation of the structural relation between those two concepts (Norman & Rumelhart, 1975). However, the utility of essay testing is very limited because various factors other than students' relational knowledge may influence the exam scores that are obtained (Diekoff, 1983). Alternative methods of assessing structural knowledge have included multiple-choice, word-association, graph-construction, and relationship-judgment tests (Diekoff, 1983; Johnson, 1967; Preece, 1976; Shavelson, 1974).

Relation and similarity judgments have also been used within the area of cognitive psychology to assess the content of semantic memory. People's judgments of the relatedness or similarity between the members of pairs of concepts are assumed to capture the underlying organization of semantic knowledge. This procedure typically produces a matrix of proximity values in which each value represents the degree of relationship between a pair of concepts. Theorists assuming dimensional models of semantic memory have made use of multidimensional scaling algorithms (MDS) to reveal the underlying dimensions on which participants base their judgments. MDS locates each concept in an $N$-dimensional space in which the distance between points reflects the psychological proximity of the corresponding concepts. The dimensions defining the space are supposed to represent the main properties along which the concepts within the domain are organized. For example, if people are rating concepts included in the mammal category, the dimensional space may be defined by properties such as size and predictability. Particular concepts (sheep, goat, lion, bear, mouse, cat, etc.) would be located in the space according to their value on
those dimensions. Lion and cat share similar values in the predacity dimension, hence they would be located in the portion of the space representing high values on that dimension. However, lion and cat differ in size and would be located on different portions of space representing the size dimension. The distance between the points lion and cat represents the psychological proximity between them and would be larger than would the distance between lion and bear, which share similar size and predacity values.

On the other hand, theorists assuming network representation of concepts in semantic memory have made use of proximity data to define empirical network structures (Friendly, 1977). Recently, Schvaneveldt’s research group (Schvaneveldt & Durso, 1981; Schvaneveldt, Durso, & Dearholt, 1985) developed a graph-theoretic technique (Pathfinder) that derives network structures from proximity data. The Pathfinder algorithm takes proximity matrices and produces a network in which concepts are represented as nodes (e.g., lion, cat, etc.) and relations between concepts are represented as links between the nodes (e.g., lion—bear). A weight corresponding to the strength of the relationship between two nodes is associated with each link and reflects the distance between the nodes (e.g., lion—bear would have a strong link weight, whereas lion—cat would have a weaker one). Concepts in the network could be directly linked (e.g., lion—bear) or linked through an indirect path (e.g., lion—cat—mouse). The Pathfinder algorithm searches through the nodes to find the closest indirect path between concepts. A link remains in the network only if that link is a minimum length path between the two concepts. Thus, Pathfinder extracts the latent structure in the data by eliminating spurious links.

Empirical evidence has supported the psychological validity of both MDS and Pathfinder representations. Several researchers have shown the ability of MDS solutions to predict categorical judgment times (Rips, Shoben, & Smith, 1973), to predict organization in free recall (Caramazza, Hersh, & Torgerson, 1976), and to capture changes in mental representation as a function of learning (Adelson, 1981; Cañas, Gonzalvo, & Bajo, 1992). Similarly, in many recent studies, researchers have indicated that Pathfinder solutions are able (a) to capture many categorical relations (Schvaneveldt & Durso, 1981) and rhyme structural representations (Rubin, 1990), (b) to predict memory retrieval (Hajo & Cañas, 1992) and memory organization (Cooke, Durso, & Schvaneveldt, 1986), and (c) to capture novice—expert representational shifts (Goldsmith et al., 1991; Schvaneveldt, Durso, Goldsmith, Bree, Cooke, & De Maio, 1985). In addition, both MDS and Pathfinder techniques provide measures to compare the cognitive structures of experts and novices. For MDS researchers accomplish this by computing Euclidean distances between all pairs of concepts in the two structures and correlating them (Brown & Stanner, 1983; Goldsmith et al., 1991). With Pathfinder structures two possible measures can be used: First, graph-theoretic distances (GTDs; number of links between pairs of concepts) can be computed for experts and novices, and then they can be correlated; second, a C value consisting of the degree to which the same node in two graphs is surrounded by a similar set of nodes is computed for each node in the novice and expert networks. Researchers can then average these C values to give an index of graph similarity. This index ranges from complementary networks (0) to identical networks (1)—see Goldsmith et al. (1991).

Therefore, both representational approaches provide valid techniques for representing and assessing structural knowledge (however, see Arabie, 1993, for criticisms of the Pathfinder technique and see Schvaneveldt, Cooke, Dearholt, Durso, & McDonald, 1994, for a rebuttal of them). Schvaneveldt, Durso, and Dearholt (1985) argued that the two representational approaches might convey different information. The goals of MDS are (a) to represent the semantic dimensions underlying a knowledge domain and (b) to arrange the concepts in the dimensional space. MDS uses a least squares technique to determine the position of the concepts in the space. Each rating datum (high or low) exerts the same level of influence on the spatial solution, so that the obtained solution is the best fit to all rating data. On the other hand, Pathfinder determines whether to link items on a pairwise basis. Pathfinder includes a link between two concepts in a network only if it is a minimum-length path between the two concepts. Hence, the network excludes important links between pairs of concepts without any global goodness-of-fit constraint. Therefore, 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. MDS seems to capture global information, whereas networks represent more local relationships. This suggestion is consistent with results showing that MDS is predictive of tasks such as categorization (Rips et al., 1973) and analogy completion (Rumelhart & Abrahamson, 1973), which require global analysis of the concepts, whereas Pathfinder seems to do well at predicting performance on recall tasks such as free recall (Cooke et al., 1986) or paired associates (Branaghan, 1990), in which interconcept relations are crucial. This interpretation leads to the joint use of both techniques for assessing the structural properties of knowledge.

Our primary concern was to evaluate the predictive validity of both MDS and Pathfinder solutions for assessing changes in the structural representation of a specific knowledge domain as a function of learning. Thus, relatedness ratings were collected from students before and after they studied a textbook on the history of psychology (Leahy, 1988) and attended to classes on this topic. To control for the possible effect of experience with the rating task, we collected additional student ratings from a control group of students that rated the history of psychology concepts only after they had studied the book and attended to the classes. Relatedness ratings were also collected from the students’ instructors on the topic, and the similarity between the instructors’ structural representations and those of their students was considered to be evidence of learning.

Because MDS and Pathfinder procedures highlight different structural aspects of knowledge, we expected that the representations derived from the two procedures would differ. Thus, we expected to capture global structural proper-
ties from the MDS procedure and local concept-to-concept interrelations from the Pathfinder algorithm. These two representational aspects seem to be particularly important for knowledge domains, such as "history of psychology," in which the important organizational principles are not clear. Many history of psychology concepts might have different meanings depending on different theoretical points of view. For example, functionalism might be viewed as a school of thought or as a basic assumption of cognitive thought opposed to behaviorism. In the same manner, sexuality might be viewed as a response, as a drive, or as the underlying principle determining most mental life. In addition, the categories and dimensions organizing and relating those concepts are not obvious. Concepts could be organized according to the school of thought that makes most use of them, according to the degree of mentalism that they imply. Therefore, there is no simple rule that could serve to classify all the concepts. In this sense, history of psychology can be considered an ill-structured domain because there is no simple set of features and logical rules that determines the classification of concepts (Medin & Schaffer, 1978). Therefore, a secondary aim of our investigation was to extend the validity of the Pathfinder and MDS techniques to assess learning in an ill-structured knowledge domain. Finally, we were interested in evaluating the validity of these techniques for predicting concept-to-concept learning (as reflected by pretest-posttest changes). Therefore, a conceptual test in which students provided definitions for each of the concepts in the study was introduced before and after the study phase.

Method

Participants and Design

Seventy-two students were selected from an introduction to psychology course taught at the University of Granada. As part of the course requirements the students had to study a history of psychology textbook on which they were then tested. Therefore, students participating in the experiment had the same task as those who were enrolled in the course but not participating in the experiment, with the only difference being that the participants' knowledge was evaluated in a different manner. All students in the course were informed about the experiment and asked for their participation. Those who volunteered completed a short questionnaire asking for their age, grade point average, and whether they had taken any previous psychology courses. Selection was made so that all students had similar ages (M = 18.4, SD = 1.03) and medium to high grade point averages (M = 6.08 on a 10-point scale, SD = 0.88). Because there were many students having had psychology as part of their bachelor curriculum, we classified students as those with and without previous knowledge. Half of the students had some knowledge about psychology, and the other half had none. Half of the previous knowledge students were randomly assigned to a prestudy–poststudy group, whereas the other half were assigned to the poststudy only group. Similarly, half of the students without previous knowledge were assigned to the prestudy–poststudy group, and the other half were assigned to the poststudy only group.

Our aim with the previous knowledge manipulation was to determine whether the Pathfinder and MDS techniques were able to capture differences in structural knowledge depending on whether students had some knowledge about psychology. Therefore, we expected some differences between the previous knowledge groups even in the pretest phase. However, this was not so, and there were no differences between the groups in either the relatedness judgment task or the definition task. Comparisons were made on three of the student–expert similarity measures (student–expert correlation on the original ratings, C values, and angular differences [AD]) and on the grades assigned by two judges to the definitive task. None of the comparisons were statistically significant. Therefore, having had psychology previously did not mean that the students had any superior knowledge about the psychology concepts that we had selected. Therefore, to simplify our analyses we collapsed data from the two knowledge groups. Hence, 36 students composed the prestudy–poststudy group, and 36 composed the poststudy–only group. Data from 3 students had to be discarded from the prestudy–poststudy group because the students did not participate in the poststudy phase of the experiment.

The expert group was composed of psychology professors at the University of Granada who had taught the introductory psychology course during the 2 years preceding the study. With this criterion, only 5 people qualified as "experts." Four experts were selected to perform the rating task and the 5th participated in the concept selection. Although four might seem to be a small number to compose the expert group, some previous studies also using a small number of experts have provided stable referent structures against which to compare students (e.g., Goldsmith et al., 1991).

Materials

Thirty-two psychology concepts were selected from the history of psychology textbook. We selected these concepts so that different schools of psychology were represented and with the restriction that all concepts were clearly defined and explained in the textbook (e.g., psychoanalysis, empiricism, perceptual analysis, intelligence). All possible pairs of concepts (496) were formed for presentation in the rating task. Additionally, a small booklet for the definition task was prepared. All the concepts were typewritten, and a blank space was provided so that the participants could write their definitions.

Procedure

Students in the experiment were told that they would rate the relatedness of pairs of 32 history of psychology concepts and that they were also to provide definitions for the concepts. Participants were also told that their performance on the task would be used to evaluate them for this course requirement (students in the course and not participating in the experiment had to take a multiple-choice test on the topic). Students in the pretest–poststudy group were told that they had to participate in the two phases of the experiment. To prevent the prestudy–poststudy students from studying only the concepts presented during the first phase, we did not inform them about the task to be performed in the second phase. Prestudy–poststudy students were told only that their history of psychology knowledge would be evaluated later.

The first phase of the experiment started approximately during the 2nd week of the academic course, and the second phase started 2 months later. Students were supposed to study the textbook during the time between the two phases. Also during the time between the phases, the course instructor explained some psychology concepts. Students in the poststudy group participated only in the second phase. In each phase, students per-
formed the rating task first and the definition task second. Each student was tested individually.

Students rated each pair of concepts using a 9-point scale ranging from least related (1) to most related (9). Each pair of concepts was centered from left to right on an IBM computer screen. The order of presentation of the pairs was randomized for each student. Students made their responses by pressing the appropriately numbered key on the number keypad. The order of concepts was randomized for each student. Students were told to work quickly and to give the first relatedness judgment that came to mind. On average, students needed approximately 1 hr to complete the rating task. To acquaint the students with the experimental procedure, we introduced 21 practice trials before the actual experimental trials.

Once students finished the rating tasks, they were given a short break (approximately 10 min), and then they received the instructions for the second task. The students were asked to provide definitions for the concepts they had just rated. The concepts were typewritten on white paper, and a space was provided below the typewritten concept for students to write their definitions. The order of concepts was randomized only once and, therefore, all students were presented with the same sequence. However, the students were told to work on the concepts in any order that they wanted. On average, students needed approximately 1 hr to complete the definition task.

Finally, proximity data were obtained from the 4 experts, who performed the rating task in exactly the same manner as the students.

**Results**

We first converted the concept ratings to proximities (dissimilarities) by subtracting each rating from the maximum possible (9). Thus, low-proximity values represented similar concepts and high-proximity values represented dissimilar ones. These proximities were then analyzed using the individual differences multidimensional scaling (INDSCAL) weighted MDS procedure (Carrol & Chang, 1970) and the Pathfinder network algorithm (Schvaneveldt, Durso, & Dearholt, 1985). Experts’ ratings were averaged to perform subsequent analysis on a single set of experts’ ratings. Before averaging and to obtain an index of experts’ similarity, we computed correlations between pairs of experts’ ratings. These correlations ranged from 0.58 to 0.62 (all ps < .01). A three-dimensional INDSCAL solution for each of the students’ ratings and the average rating of the experts yielded an optimal degree of stress and interpretability. Therefore, all subsequent analyses were based on these three dimensions.

Pathfinder networks were obtained with the $q$ and $r$ parameters equal to $n - 1$ (where $n$ = number of concepts) and infinity, respectively. The Pathfinder algorithm organizes data by eliminating links that are not the minimum-length path between two concepts. A path consists of a sequence of nodes and the connecting links. The length of a path is defined by the $r$ parameter, which is based on the Minkowski $r$-metric. The length of a path defined by the $r$ parameter is a function of the weights (calculated from ratings) associated with the links in the path. As $r$ decreases, links are usually added to the network. For example, with $r = 1$, Pathfinder simply uses the sum of the links’ weight to determine the length of a path in the network. When the parameter $r$ is set to infinity, the number of links in the network is maximally reduced. The parameter $q$ defines the maximum number of links in a path and also affects network density. The parameters $r = \infty$ and $q = n - 1$ generate the simplest (least dense) Pathfinder network and require only ordinal assumptions to be made about the distance estimates (Schvaneveldt, Durso, & Dearholt, 1985). As with INDSCAL, the experts’ ratings were averaged to obtain a single reference network against which we could compare the students’ networks.

Students’ definitions were graded by two independent judges who rated the quality of the definitions on a scale ranging from poor (1) to good (5). The two judges were 3rd-year graduate students in the basic experimental psychology program (history of psychology part of the curriculum in basic experimental psychology at the University of Granada). These judges had already finished their graduate courses and had some teaching assignments. We preferred these judges over the experts and ourselves so that the generalizability of the results would be increased. The two judges were provided with the history of psychology textbook and instructed to grade the definitions as if they were answers to questions on an actual exam. The scores were averaged once over concepts and once over students for each of the groups and phases of the experiment, and interjudge reliabilities (correlations) were calculated for the averaged concept scores and for the averaged student scores. These correlations were very high and ranged from 0.7 to 0.9. Therefore, the averaged scores of the two judges were used in the following analyses.

**Assessment of Learning**

We assessed learning by comparing the cognitive structure of the students and the experts before and after the study phase and by comparing the students’ performance on the definition task before and after studying. We computed similarity values by averaging experts’ ratings first and then by comparing each student to this single set of averaged ratings. If learning involves the restructuring of the mental representation of a knowledge domain, then students’ structural representations should change and become more similar to those of the experts as learning progresses. Therefore, the change should be present in both student–expert similarities and student definition scores.

Two similarity measures were obtained from the Pathfinder algorithm: First, GTDs were computed for both students and experts. Given that two nodes are connected in the network (there is a path between them), the nodes can be close (directly linked) or far (two or more links in the path between them). The distance between pairs of nodes is a structural property of the graph (network). Therefore, two graphs might be similar to the extent that distances between pairs of concepts are similar. We calculated these distances by computing the number of links constituting the shortest path between pairs of concepts in the network, and these were represented in a distance vector. A network with $n$ nodes may be represented by a vector of $(n^2 - n)/2$ dis-
tances. The correlation between the distances in two networks provides an index of network similarity (Goldsmith & Davenport, 1991). A high correlation is interpreted as high network similarity, whereas a low correlation is interpreted as the two networks being different. Negative correlations usually would not be found, because they would mean a consistent but contrary type of knowledge for the same set of concepts. Hence, correlations between GTDs of students and experts were calculated for each of the groups and phases.

Second, C values were obtained for each student. As we already mentioned, the C value provides a measure of the degree of similarity between two graphs. The set of nodes that are within Distance 1 (i.e., immediate neighbors) from a particular node is also a structural property of a network. Two networks can be compared by assessing the similarity of their neighbors for corresponding nodes. C is a measure of the degree of shared elements relative to some total pool of elements. That is, the C values provide a measure of the degree to which the nodes in two graphs are surrounded by a similar set of nodes. We accomplished this by examining the intersection of the node neighborhood and the union of the neighborhoods. Overall graph similarity is the mean of the C values obtained for the graph's nodes. This measure ranges from 0 to 1, with higher values indicating greater similarity (Goldsmith & Davenport, 1991). A C value was computed for each node in the student and expert networks. These C values were averaged for each student and provided an index of graph similarity.

ADs between student and expert vectors in the multidimensional space obtained by the INDSCAL procedure were also calculated for each group and phase. The INDSCAL procedure creates a multidimensional space in which novice and expert weights on the dimensions are represented. Hence, a multidimensional solution for each of the students' ratings and the average rating of the experts was obtained. The similarity between experts and novices was then estimated by the AD between the vectors representing them (Schiffman, Reynolds, & Young, 1981). Finally, the similarity of the direct ratings was also calculated by correlating the raw ratings of each of the students and the averaged experts' rating. Table 1 shows the mean definition scores, mean direct ratings (DR), mean C values, mean GTDs, mean angular multidimensional distances for each group and phase of the experiment.

As shown, both students' definition scores and student-expert similarity (as reflected by the correlation between their raw ratings, DR) increased from the pretest phase to the posttest phase and both were similar for the two posttest groups. We performed correlated-sample t tests on the pretest–posttest group that yielded significant differences between the two experimental phases, t(32) = 12.62, p < 0.001, for scores and t(32) = 6.00, p < 0.001, for ratings. Similarly, independent t tests performed for comparison between the pretest phase of the pretest–posttest group and the posttest phase of the posttest-only group were significant, t(67) = 9.50, p < 0.001, for scores and t(67) = 5.00, p < 0.001, for ratings. In addition, as expected, students' scores and ratings in the two post-

<table>
<thead>
<tr>
<th>Group and phase</th>
<th>Score</th>
<th>Rating</th>
<th>C</th>
<th>GTD</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre–post study group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presubtest phase</td>
<td>1.88</td>
<td>.34</td>
<td>.14</td>
<td>.11</td>
<td>.50</td>
</tr>
<tr>
<td>Postsubtest phase</td>
<td>2.56</td>
<td>.44</td>
<td>.21</td>
<td>.23</td>
<td>.29</td>
</tr>
<tr>
<td>Postsubtest only group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presubtest phase</td>
<td>2.49</td>
<td>.42</td>
<td>.21</td>
<td>.21</td>
<td>.28</td>
</tr>
</tbody>
</table>

a Maximum = 5. b Range = high similarity (1) to low similarity (0); c Range = low similarity (0) to high similarity (1); d Range = low similarity (0) to high similarity (1); e Range = high similarity (0) to low similarity (1).

study phases were similar, both ts < 1 in absolute value. Thus, students' performance on the definition task was better after studying the textbook, and this improvement was also present in the students' ratings. Therefore, both student scores and ratings changed with learning.

A more interesting goal was to show that knowledge organization changed as a function of learning. Thus, all student–expert similarity measures increased from the pretest to the posttest phase. These changes are shown in Table 1 and indicate that after studying the history of psychology textbook, student networks and multidimensional solutions became more like those of their instructors. We performed correlated-sample t tests and compared the C, GTDs, and AD measures obtained in the pretest phase of the pretest–posttest group with those obtained in the posttest phase of the same group. All comparisons were significant, t(32) = 5.60, p < 0.001, for C; t(32) = 4.20, p < 0.001, for GTDs; and t(32) = 4.60, p < 0.001, for AD. We calculated independent-sample t tests to compare the pretest phase of the pretest–posttest group with the posttest phase of the posttest-only group. These comparisons were also significant, with t(67) = 5.24, p < 0.001, for C; t(67) = 3.90, p < 0.001, for GTD; and t(67) = 4.18, p < 0.001, for AD. Therefore, all of the pretest versus posttest comparisons were significantly different from each other. In contrast, the C values, GTDs, and ADs were not different when the posttest phases of the pretest–posttest group and of the posttest-only group were entered into the comparison, all ts < 1 in absolute value.

Comparisons Between Similarity Measures

The obtained C values, GTDs, and ADs were able to capture structural changes as a function of learning. Thus, students organized the history of psychology concepts more like their instructors did once the students had studied the history of psychology book. This better organization was shown by the higher C values and GTDs and by the lower ADs in the posttest phase relative to the pretest phase.
This change in structural knowledge should be related to the way in which students perform the definition task. Therefore, similarity measures should be able to predict students’ performance on the definition task. Traditional tests are closer to our definition task than to the relatedness rating task. Hence, before claiming that the relatedness task is a good way to assess knowledge, one should show that the similarity measures extracted from the direct rating task are able to predict students’ scores on the definition task. Therefore, we computed correlations between each similarity measure and the students’ scores on the definition task. Definitions scores correlated 0.54 with C, 0.44 with GTD; –0.47 with AD, and 0.56 with DR. The DRs were also correlated because it is important to show that our similarity measures are able to explain part of the variance that is not explained by the DRs.

Both the MDS and Pathfinder procedures claim to capture organizational properties that are not present in the DRs. Therefore, part of the explained variance should be due to the organizational properties captured by those procedures. To assess whether Pathfinder and MDS provided prediction of students’ definition scores beyond that provided by direct judgments per se, we computed partial correlations between the four measures extracted from the rating task and the students’ grades in the definition task. Table 2 shows the partial correlations of each individual measure (DR, C, GTD, and AD) as predictors of students’ definition scores with every other measure held constant.

The second and fourth rows of Table 2 show that both the C and AD measures correlated significantly despite the inclusion of each control variable. Therefore, both of them captured unique predictive variance from the concept ratings. This result is interesting because it shows that when weighted multidimensional analyses are used and are based on experts’ and novices’ ratings entering into the multidimensional solution, both Pathfinder and MDS similarity indices (C and AD, respectively) are predictive of students’ scores. In contrast, GTD did not correlate significantly when the variance associated with C was held constant, thus indicating that the information extracted by GTD is present to some extent in the obtained C values.

Interestingly, DR was still predictive when the other similarity measures were partialed out, thus indicating that the raw ratings still contained unique information that was not captured by ADs and Cs. A stepwise multiple regression analysis (that was based on standardized beta weights) indicated that the best predictive model was as follows:

\[
\text{Scores} = 0.378 (\text{DR}) - 0.253 (\text{AD}) + 0.273 (\text{C}).
\]

The three variables together (DR, AD, C) were significantly predictive of scores, \(F(3, 98) = 30.23, p < 0.001\), and accounted for 48% of the total variance. The \(r_5\) associated with the standardized beta weights were 4.64, \(p < 0.001\), for DR; –3.18, \(p = 0.002\), for AD; and –3.14, \(p = 0.002\), for C.

**Conceptual Analysis**

In the analyses performed in the two previous sections we intended to assess overall learning and to predict students’ definition scores from the network and dimensional structures provided by the Pathfinder and multidimensional procedures. Obviously, an important question is do these procedures make it possible to identify concepts and dimensions that were understood and misunderstood by students? Concept understanding could be defined in terms of the similarity of students versus experts’ conceptual representations of the same concept.

We performed different analyses to characterize network and multidimensional representations, because the information provided by them is different. The conceptual analysis performed with Pathfinder involved the assessment of changes in concept-to-concept structural representation as a function of learning, whereas the analysis performed with the multidimensional procedure involved the assessment of changes in the dimensional–global representation.

**Characterizing the network.** Concept understanding was defined in terms of the degree of similarity between students’ and experts’ representations in the network. A concept representation in a student network could be linked to the same concepts as in the experts’ networks (understanding of the concept) or could be linked to different concepts (misunderstanding of the concept). Therefore, it is possible to identify well- or ill-structured concepts by looking at the C values associated with each of them. Large C values would be associated with a good understanding of the concept, whereas low C values would be associated with a poor understanding of it. An obvious prediction is that concepts’ C values should be highly correlated with concepts’ scores in the definition task. Furthermore, learning should reduce the number of misunderstood and misdefined concepts. Thus, we expected a reduction in the number of misunderstood concepts from the pretest phase to the poststudy phase.

Therefore, our first analysis was directed at finding correlations between concepts’ C values and their associated definition scores. C values for the concepts were found for each of the 32 concepts. Although we explored the relation between the grades and the C values at the individual level, there was too much noise in the data to obtain a clear correlational pattern. Therefore, we averaged the C values
for each of the concepts over the students to calculate the mean C value for each concept, group, and phase. Table 3 shows the mean C values and scores for each of the 32 concepts as a function of the learning phase. The correlations between C values and scores were $r = 0.59$, $p < 0.01$, for the pretest phase and $r = 0.47$, $p < 0.01$, for the posttest phase. As predicted, there was a general tendency for students to give better definitions to concepts for which their relational structure was more similar to that of the experts. Obviously the correlations are far from perfect possibly because of differences in the type of knowledge required to perform the definition and relational tasks.

A counterintuitive finding of these correlational analyses was that the correlations between the concepts’ C values and scores was higher (0.59) for the pretest phase than for the posttest phase (0.47). Before studying the textbook, participants shared popular knowledge of some of the more familiar psychology concepts (e.g., sexuality, dreams) and lacked knowledge of the more technical ones (catharsis, eugenics). Lack of knowledge about a concept would result in a completely disorganized representational structure and in a very poor definition. That is, before studying the textbook, a student might not have known anything about a concept such as catharsis and hence would not have been able to provide a good definition for the concept or to relate it to psychoanalysis. Therefore, poor knowledge about a concept would be reflected in both its representational structure and definition. However, after studying the textbook, students would have learned more technical knowledge about the concepts, although this knowledge was still far from perfect. This partial knowledge might produce some inconsistencies between the definitions and the relational structure. For example, after studying the textbook a student might have learned that catharsis was an important concept for the psychoanalytic school but might not have remembered what its meaning was. Therefore, its representational structure would be good, but its definition would be poor. These inconsistencies would result in relatively lower correlations between C values and scores following study of the textbook.

Another interesting feature of the data in Table 3 is that although most of the concepts improve from one phase to the next, there is a tendency for those concepts with low

<table>
<thead>
<tr>
<th>Concept</th>
<th>C</th>
<th>Score</th>
<th>M</th>
<th>SD</th>
<th>C</th>
<th>Score</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instinct</td>
<td>0.07</td>
<td>0.10</td>
<td>1.62</td>
<td>0.76</td>
<td>0.18</td>
<td>0.16</td>
<td>2.45</td>
<td>0.83</td>
</tr>
<tr>
<td>Functionalism</td>
<td>0.04</td>
<td>0.09</td>
<td>1.50</td>
<td>0.82</td>
<td>0.06</td>
<td>0.14</td>
<td>2.40</td>
<td>0.56</td>
</tr>
<tr>
<td>Consciousness</td>
<td>0.13</td>
<td>0.11</td>
<td>1.75</td>
<td>0.60</td>
<td>0.17</td>
<td>0.13</td>
<td>2.45</td>
<td>0.60</td>
</tr>
<tr>
<td>Anxiety</td>
<td>0.07</td>
<td>0.10</td>
<td>1.59</td>
<td>0.60</td>
<td>0.14</td>
<td>0.20</td>
<td>2.59</td>
<td>0.58</td>
</tr>
<tr>
<td>Computer science</td>
<td>0.03</td>
<td>0.08</td>
<td>2.43</td>
<td>0.72</td>
<td>0.10</td>
<td>0.15</td>
<td>2.50</td>
<td>0.59</td>
</tr>
<tr>
<td>Association</td>
<td>0.10</td>
<td>0.11</td>
<td>1.63</td>
<td>0.64</td>
<td>0.12</td>
<td>0.10</td>
<td>2.28</td>
<td>0.58</td>
</tr>
<tr>
<td>Gestalt</td>
<td>0.02</td>
<td>0.06</td>
<td>1.45</td>
<td>0.77</td>
<td>0.02</td>
<td>0.07</td>
<td>2.91</td>
<td>0.62</td>
</tr>
<tr>
<td>Determinism</td>
<td>0.08</td>
<td>0.11</td>
<td>1.55</td>
<td>0.74</td>
<td>0.08</td>
<td>0.17</td>
<td>1.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Inheritance</td>
<td>0.08</td>
<td>0.11</td>
<td>3.16</td>
<td>0.94</td>
<td>0.18</td>
<td>0.18</td>
<td>3.38</td>
<td>0.59</td>
</tr>
<tr>
<td>Psychophysics</td>
<td>0.02</td>
<td>0.05</td>
<td>1.18</td>
<td>0.51</td>
<td>0.04</td>
<td>0.09</td>
<td>1.38</td>
<td>0.61</td>
</tr>
<tr>
<td>Positivism</td>
<td>0.12</td>
<td>0.17</td>
<td>1.68</td>
<td>0.77</td>
<td>0.16</td>
<td>0.15</td>
<td>2.43</td>
<td>0.83</td>
</tr>
<tr>
<td>Apperception</td>
<td>0.01</td>
<td>0.04</td>
<td>1.07</td>
<td>0.18</td>
<td>0.09</td>
<td>0.20</td>
<td>2.13</td>
<td>0.87</td>
</tr>
<tr>
<td>Sensationism</td>
<td>0.01</td>
<td>0.04</td>
<td>1.59</td>
<td>0.60</td>
<td>0.02</td>
<td>0.07</td>
<td>2.27</td>
<td>0.76</td>
</tr>
<tr>
<td>Structuralism</td>
<td>0.14</td>
<td>0.14</td>
<td>1.70</td>
<td>0.76</td>
<td>0.18</td>
<td>0.14</td>
<td>2.50</td>
<td>0.59</td>
</tr>
<tr>
<td>Empiricism</td>
<td>0.08</td>
<td>0.11</td>
<td>1.33</td>
<td>0.54</td>
<td>0.19</td>
<td>0.18</td>
<td>2.44</td>
<td>0.62</td>
</tr>
<tr>
<td>Empiricism</td>
<td>0.08</td>
<td>0.11</td>
<td>1.33</td>
<td>0.54</td>
<td>0.19</td>
<td>0.18</td>
<td>2.44</td>
<td>0.62</td>
</tr>
<tr>
<td>Cynicism</td>
<td>0.18</td>
<td>0.16</td>
<td>1.30</td>
<td>0.48</td>
<td>0.23</td>
<td>0.18</td>
<td>2.12</td>
<td>0.82</td>
</tr>
<tr>
<td>Eugenics</td>
<td>0.07</td>
<td>0.13</td>
<td>1.06</td>
<td>0.27</td>
<td>0.22</td>
<td>0.31</td>
<td>2.51</td>
<td>1.04</td>
</tr>
<tr>
<td>Sexuality</td>
<td>0.14</td>
<td>0.25</td>
<td>1.89</td>
<td>0.68</td>
<td>0.29</td>
<td>0.24</td>
<td>2.26</td>
<td>0.53</td>
</tr>
<tr>
<td>Atomism</td>
<td>0.13</td>
<td>0.13</td>
<td>1.33</td>
<td>0.54</td>
<td>0.19</td>
<td>0.18</td>
<td>2.44</td>
<td>0.62</td>
</tr>
<tr>
<td>Habit</td>
<td>0.10</td>
<td>0.17</td>
<td>1.27</td>
<td>0.57</td>
<td>0.21</td>
<td>0.19</td>
<td>1.88</td>
<td>0.85</td>
</tr>
<tr>
<td>Stimulus</td>
<td>0.20</td>
<td>0.14</td>
<td>2.07</td>
<td>0.75</td>
<td>0.25</td>
<td>0.15</td>
<td>2.43</td>
<td>0.59</td>
</tr>
<tr>
<td>Thinking</td>
<td>0.20</td>
<td>0.08</td>
<td>2.54</td>
<td>1.01</td>
<td>0.26</td>
<td>0.14</td>
<td>2.74</td>
<td>0.59</td>
</tr>
<tr>
<td>Response</td>
<td>0.21</td>
<td>0.12</td>
<td>2.86</td>
<td>0.76</td>
<td>0.21</td>
<td>0.13</td>
<td>3.34</td>
<td>0.58</td>
</tr>
<tr>
<td>Intelligence</td>
<td>0.22</td>
<td>0.15</td>
<td>1.98</td>
<td>0.83</td>
<td>0.23</td>
<td>0.15</td>
<td>2.62</td>
<td>0.76</td>
</tr>
<tr>
<td>Psychoanalysis</td>
<td>0.23</td>
<td>0.13</td>
<td>2.12</td>
<td>1.00</td>
<td>0.44</td>
<td>0.15</td>
<td>3.06</td>
<td>0.64</td>
</tr>
<tr>
<td>Mind</td>
<td>0.26</td>
<td>0.12</td>
<td>1.74</td>
<td>0.56</td>
<td>0.29</td>
<td>0.15</td>
<td>2.49</td>
<td>0.64</td>
</tr>
<tr>
<td>Environment</td>
<td>0.19</td>
<td>0.11</td>
<td>2.74</td>
<td>0.86</td>
<td>0.22</td>
<td>0.14</td>
<td>2.91</td>
<td>0.48</td>
</tr>
<tr>
<td>Behavior</td>
<td>0.27</td>
<td>0.13</td>
<td>1.83</td>
<td>0.52</td>
<td>0.31</td>
<td>0.16</td>
<td>2.79</td>
<td>0.52</td>
</tr>
<tr>
<td>Unconsciousness</td>
<td>0.25</td>
<td>0.18</td>
<td>1.95</td>
<td>0.85</td>
<td>0.33</td>
<td>0.14</td>
<td>2.60</td>
<td>0.65</td>
</tr>
<tr>
<td>Behaviorism</td>
<td>0.27</td>
<td>0.13</td>
<td>2.27</td>
<td>1.00</td>
<td>0.36</td>
<td>0.18</td>
<td>3.12</td>
<td>0.65</td>
</tr>
<tr>
<td>Dreams</td>
<td>0.42</td>
<td>0.31</td>
<td>2.21</td>
<td>0.63</td>
<td>0.52</td>
<td>0.27</td>
<td>2.66</td>
<td>0.46</td>
</tr>
</tbody>
</table>

* C values above the median of 0.184.
C values in the pretest phase also have low C values in the posttest phase. In fact, the correlation between the concepts’ C values in the pretest and posttest phases was $r = 0.91$, $p < 0.01$. This indicates that all concepts improved to similar degrees, with more difficult concepts remaining more difficult after studying. Therefore, learning was evenly distributed across more and less difficult concepts.

To assess understanding of the concepts, we calculated the median C value across both phases and found it to be 0.184. We considered concepts above the median to be better understood than concepts below the median. It is possible to identify well or poorly understood concepts by looking at their position in relation to the median. Concepts above the median are marked with a superscript letter a in Table 3. As shown, only 12 of the 32 concepts were above the median before studying, whereas 19 of them were above this value after studying. Before studying, students had some relational knowledge about concepts such as habit, stimulus, thought, response, and so forth. After studying the students understood relational aspects of concepts such as structuralism, empiricism, cognitivism, eugenics, atomism, sexuality, and catharsis. However, there was still a large number of poorly understood concepts.

Inspection of the students’ posttest scores in the definition task showed considerable variability. The posttest definition task scores ranged from 3.5 to 1.4 (out of 5), with a mean of 2.53. Thirty-eight of the students were above the mean (better performing group), and 31 were below it (poorer performing group). The large number of misdefined concepts in the posttest phase was caused by the 31 poorer performing students in that phase. The number of misdefined concepts in the better performing group was only 9. This contrasts with the 19 concepts below the median that we found for students in the poorer performing posttest group.

Therefore, as a final qualitative analysis, the Pathfinder networks for the pretest; poorer performing, posttest; better performing, posttest; and expert groups were obtained. We generated these networks by averaging the ratings of the students within each group and obtaining a single Pathfinder network for each of the averaged proximity data. Hence, we generated the pretest network by averaging the pretest ratings of the 33 students that composed the pre-

![Figure 1. Pathfinder network for students in the pretest phase.](image-url)
post study group; we obtained the better performing, poststudy network by averaging the ratings of the 38 students with definition scores above the mean; we obtained the poorer performing, poststudy network by averaging the ratings of the 31 students with scores below the mean; finally, we obtained the expert network by averaging the ratings of the 4 experts. Figures 1, 2, 3, and 4 show those networks. An interesting feature of these networks is the gradual organization of concepts as a function of the learning group. Thus, the networks of the pretest and the poorer performing poststudy groups look very dense and disorganized, and it is difficult to find organizational principles. In contrast, the network from the better performing poststudy group starts to show some organization, with an important organizational focus on such concepts as mind, determinism, behaviorism, association, and psychoanalysis. Finally, the experts' networks show a clear organization, with concepts organized around the schools of thought to which they are more central. Concepts such as consciousness, thinking, and mind are directly linked to cognitivism; behaviorism is directly associated with concepts such as stimulus, determinism, and empiricism; psychoanalysis is directly linked to catharsis, instinct, dreams, and unconsciousness.

At a local concept-to-concept level, it is also possible to observe some interesting tendencies. For example, a technical concept such as eugenics is erroneously linked to unconsciousness and apperception in the pretest network. This concept also shows some misconceptions in the poorer performing poststudy network because it is directly linked to consciousness. However, the correct link between eugenics and heritage is already present, thus indicating that something has been learned about the concept. Finally, in the better performing poststudy network erroneous links are eliminated and eugenics is directly linked only to heritage.

Characterizing the dimensional space. To assess changes in the dimensional representation of concepts as a function of learning, we constructed four group matrices with the distances obtained from the ratings. One matrix was constructed by pooling conceptual distances from the pretest phase of the pretest–poststudy group (N = 33); a second matrix was constructed for the conceptual distances from the poorer performing poststudy group (pooling students in pretest–poststudy and poststudy only groups, N = 38); a third matrix was constructed for better performing students in the poststudy phase (N = 31); and the fourth matrix contained conceptual distances that were

---

**Figure 2.** Pathfinder network for poorer performing students in the poststudy phase.
obtained from the experts’ ratings ($N = 4$). Therefore, this analysis and the obtained multidimensional representation parallel the group analysis performed by the Pathfinder algorithm on the averaged group matrices. We entered these four matrices into the INDSCAL procedure to find a multidimensional space common to the four groups. A three-dimensional space yielded the greatest degree of stress and interpretability.

Figures 5 and 6 show the conceptual space defined by Dimensions 1 and 2 and Dimensions 2 and 3, respectively. Figures 7 and 8 show each groups’ weights on each dimension. Figure 5 shows that Dimension 1 separated some of the behaviorist and psychoanalytic concepts from the rest of the concepts. Figure 7 showed that the students weighted much more heavily on Dimension 1 than did the experts. Interestingly, 9 of the 12 concepts that had C values above the median in the prestudy phase were among those that weighted positively on this dimension (e.g., stimulus, habit, dreams, unconscious). Hence, we interpreted Dimension 1 as capturing some general familiarity with the concepts. Popular psychology stresses the role of psychoanalysis and behaviorism as the primary psychological schools, and therefore, some psychoanalytic and behaviorist concepts might be more familiar to people in general and hence, more familiar to our students. This first dimension seemed to lack importance to our experts. In contrast, experts weighted very heavily on Dimension 2, with better performing, poorer performing, and pretest students graded on their weights on this second dimension. Therefore, the weights on this dimension appeared to depend on the degree of learning. Figure 5 showed that this second dimension separated mental-rationalist concepts (e.g., mind, consciousness) from external–empiricist ones (e.g., empiricism, stimulus). Thus, the second dimension seems to be associated with the degree of mentalism attributed to the concept. The combinations of Dimension 2 and 3 also had the property of discriminating among different learning groups. As shown in Figure 8, all the points representing the different learning groups are located on the diagonal, with the pretest group weighting very little on both dimensions and with the experts assigning strong weights to both of them. Dimension 3 separated the biological and environmental determinants of behavior (e.g., heritage, sexuality) from the mental ones (e.g., thinking, consciousness). Combination of the second and third dimensions separated the space in four quadrants each containing a set of concepts. The upper right quadrant

![Pathfinder network for better performing students in the poststudy phase.](image)
contained concepts associated with mental processes or states (e.g., mind, intelligence), the upper left quadrant contained concepts linked to biological determinants of behavior (e.g., instinct, heritage), the lower left quadrant contained concepts associated with environmental determinants (e.g., behavior, environment), and the lower right quadrant represented empiricist concepts that were, however, linked to internal states (e.g., association, sensation).

Discussion

Results from the present study provided preliminary support for the indication that the acquisition of knowledge implies modification in the way in which knowledge is represented and structured. The classic expert-knowledge difference was replicated in the domain of the history of psychology. Experts' representation of concepts was more organized than was the students' conceptual representation. Thus, experts' networks were clearly structured around a few central concepts, whereas students' networks showed less clearly the organizational principles and central concepts in their representation. In addition, experts and students showed differences in the global dimensions by which they organized their history of psychology knowledge. Thus, familiarity with the concepts was an important organizational dimension for students, whereas mentalism and determinants of behavior were the important dimensions for experts.

Besides the student–expert comparisons, structural changes in knowledge as a function of learning were assessed by comparisons among students showing different degrees of learning. Thus, students' networks and dimensional weights varied depending on whether they had already studied the textbook (prestudy vs. poststudy comparisons) and on whether they had learned the material according to their scores on the definition task (comparisons between poorer performing and better performing students). These comparisons are important because a clear understanding of how students' performance improves in a knowledge domain cannot be obtained by only comparing them to a group of experts whose knowledge and aptitude in the domain is far beyond that of the novices. Prestudy–poststudy comparisons allowed for direct assessment of changes in performance and in the structure of knowledge as knowledge of the domain increased. Clearly, our students provided better definitions, had more structured networks,

![Pathfinder network for the expert group.](image.png)
and attributed more weight to central dimensions after studying the textbook and attending psychology classes than they had before.

Important aims of our study were exploring how to capture and represent the organizational properties of knowledge and assessing how these representations were modified with the acquisition of new knowledge. Comparisons between student–expert similarity measures indicated that both representational approaches provide valid techniques for representing and assessing structural knowledge. First, C values, GTDs, and ADs were able to capture learning so that students in the poststudy phases showed greater C values and smaller distances than did students in the presudy phase. Therefore, the three measures are valid indices of learning. Second, C values and ADs were able to predict students’ scores in the definition task. ADs and C values showed significant correlations with grades when the variance that was due to the ratings and the other similarity measures was partialed out. This indicated that both metrics contain unique predictive variance and that they possibly convey different information (Schvaneveldt, Durso, & Goldsmith, 1985). In contrast, GTDs were not able to predict new variance once the variance contributed by the C values was partialed out. This indicated that the information extracted by GTD is already present in the C values, and therefore C may be a better index of network similarity.

The demonstration of predictive power for both the MDS and Pathfinder techniques did not support results from some recent studies showing the superiority of the Pathfinder technique at predicting performance (Goldsmith et al., 1991). In these studies researchers have compared the predictive validity of MDS and Pathfinder solutions and their abilities to assess changes in mental representation as a function of learning (e.g., Cooke et al., 1986; Goldsmith et al., 1991), and we conclude that Pathfinder solutions are better predictors of students’ performance. For example, Goldsmith et al. (1991) explored the validity of MDS and Pathfinder to assess students’ cognitive representation of classroom learning. Students enrolled in an introductory statistics class judged the relatedness of 30 statistical and design concepts at the beginning and at the end of the course. Researchers learning by comparing the Pathfinder and MDS

![Conceptual space as defined by Dimensions 1 (general familiarity) and 2 (degree of mentalism).](image)

Figure 5. Conceptual space as defined by Dimensions 1 (general familiarity) and 2 (degree of mentalism).
solutions of students at the beginning and end of the course with those of their instructor. Thus, the similarity between each student and the instructor was assessed using four different indices: correlations of raw relatedness ratings, correlations of MDS distances, correlations of GTDs, and correlations of C values. The correlations of these indices with the students’ grades after the course showed that the best predictors of students’ exam performance were the Pathfinder C values ($r = 0.74$). In fact, when the original ratings were partialled out, the only significant correlation was that of the C values. When the C values were partialled out, none of the other indices correlated significantly with course grades. This finding led Goldsmith et al. to suggest that Pathfinder networks better represent conceptual domains than do MDS spatial models. However, this conclusion might be due to the particular MDS procedure used by Goldsmith et al. and not to the lack of predictive power of MDS techniques. In their study, Goldsmith et al. used unweighted procedures so that distances of concepts were correlated with the original ratings of a single student to obtain the optimal MDS solution for students. Thus, different MDS solutions were obtained for each of the novices and experts in the experiment. Then, the conceptual distances from the novice and expert solutions were correlated to assess similarities in their conceptual representations.

This unweighted procedure poses two problems for comparing novice and expert conceptual representations. First, the dimensions defining the MDS space would not be the same for expert and novices. Mathematically this is irrelevant because expert–novice comparisons are made on Euclidean distances, and these distances remain invariant across the rotation of axes. However, although the approach does not require that comparisons be made on the same dimensional axes, the psychological meaning of distances that are based on different dimensions differs. For example, stimulus and response might be closely located in a student MDS solution because both concepts are familiar to the student, and both are located high in the general familiarity dimension. These two concepts might also be closely located in the expert solution but because they have a low degree of mentalism. The similarity between these distances would result in a high student–expert similarity index.

![Figure 6](image)

Figure 6. Conceptual space as defined by Dimensions 2 (degree of mentalism) and 3 (determinants of behavior).
However, this index very probably would not be predictive of students’ performance on the definition task, because a definition that is based on global familiarity with a concept would necessarily be poor.

Second, with Goldsmith et al.’s (1991) unweighted approach, the correlation between experts’ and novices’ MDS distances and the correlation between experts’ and novices’ original ratings will be similar. This is because unweighted
MDS solutions are obtained by maximizing the correlation between individual ratings and distances. MDS creates a spatial representation so that distances between concepts in the solution correspond to distances between concepts in the raw proximity matrix. The difference between MDS distances and distances in the raw proximity matrix is that MDS represents distances in a low-dimensional space, whereas the original ratings exist in a high-dimensional space \((n-1)\) dimensions when there are \(n\) concepts. The two sets of distances would agree perfectly in an \(n-1\) dimensional space. As the number of dimensions is reduced, the agreement between the two sets of distances would diminish, but it will still remain high because MDS attempts to keep stress low. Although MDS distances and raw ratings distances are not exactly the same, for the levels of stress to remain reasonable, they will necessarily be correlated. Partial correlations between MDS distances and course grades with original ratings held constant would not be significant because ratings and MDS distances would contain overlapping information.

These two problems may be causing the lack of predictive power of MDS distances in Goldsmith et al.'s (1991) study. In the present experiment, a weighted MDS procedure was used so that the novices' and experts' rating matrices were taken simultaneously to obtain a common optimal solution. In the weighted procedure, the points representing the concepts are moved in the space to find a final location that optimizes the correlation between distances and both novice and expert ratings. The obtained solution is common to both, but the procedure assigns weights to the novice and expert contribution to each of the dimensions defining the MDS space. Therefore, it was possible to create a multidimensional space in which the points represent novice and expert weights on the dimensions, and the similarity between experts and novices was estimated by the AD between the points representing them. Conflicting results might be due to the particular MDS procedure used and not to the lack of predictive power of MDS procedure. Results of the present study indicated that when MDS weighted procedures were used, the MDS solution also contained unique predictive variance.

This last result is important because it suggests that both the Pathfinder and the MDS techniques could be used simultaneously to provide different structural aspects of the knowledge domain. The qualitative analyses performed on the Pathfinder and MDS data provided very different information about the way in which history of psychology knowledge is represented and organized by the students and their instructors. The Pathfinder technique provided local concept-to-concept structural information. Thus, it was possible for us to assess participants' understanding of individual concepts by looking at student–expert similarities in the network connections to individual concepts, and it was possible for us to show how these local connections changed as a function of learning. The MDS technique, on the other hand, provided information about the dimensions structuring the knowledge domain and about how these dimensions are also differentially weighted by students and instructors as well as by students with different degrees of knowledge. Hence, MDS solutions capture global information, whereas networks represent more local relationships. This suggestion should lead researchers to the joint use of both techniques when assessing the structural properties of knowledge.

References


1995 APA Convention Call for Programs

The *Call for Programs* for the 1995 APA annual convention appears in the September issue of the *APA Monitor.* The 1995 convention will be held in New York, New York, from August 11 through August 15. The deadline for submission of program and presentation proposals is December 2, 1994. Additional copies of the *Call* are available from the APA Convention Office, effective in September. As a reminder, agreement to participate in the APA convention is now presumed to convey permission for the presentation to be audiotaped if selected for taping. Any speaker or participant who does not wish his or her presentation to be audiotaped must notify the person submitting the program either at the time the invitation is extended or before the December 2 deadline for proposal submission.