Journal of Educational Psychology
1996, Vol. 88, No. 3, 572–585

0022-0663/96/$3.00

Conceptual Maps and Simulated Teaching Episodes as Indicators of Competence in Teaching Elementary Mathematics

Rebecca L. Gomez and O. D. Hadfield
New Mexico State University

Lynn D. Housner
West Virginia University

Network representation was used to assess knowledge obtained during a teaching methodology course in elementary mathematics. Participants were the course instructor, 4 teacher educators, and 53 prospective teachers. Relatedness ratings on key terms were used to construct associative networks. Teacher educator networks shared significant similarities to the course instructor's network, and similarities between the teacher educator and prospective teacher networks were as predictive of course grades as similarity to the course instructor. Fourteen of the prospective teachers participated in a simulated teaching task. Network similarity predicted teaching for conceptual understanding, as did final course grade, and was more predictive than were either exams or lab scores. The advantage of associative networks may be in representing patterns of concept relations underlying mental models of teaching.

Teaching for conceptual understanding has become one of the fundamental components of reform movements in mathematics education. Educational researchers and professional associations alike are advocating a greater emphasis on the development of conceptual knowledge to complement the traditional emphasis on procedural knowledge (Eisenhart et al., 1993; Even, 1993; Mathematical Sciences Education Board, 1991; Mathematical Association of America, 1991; National Council of Teachers of Mathematics, 1991).

Procedural knowledge is defined as knowledge of the symbols, rules, and algorithms that are used for solving mathematical problems (Eisenhart et al., 1993), whereas *conceptual knowledge* is defined as "the relationships and interconnections of ideas that explain and give meaning to mathematics" (Eisenhart et al., 1993, p. 9). In short, conceptual knowledge is an understanding of the concepts that underlie procedures. A procedural approach to teaching division of fractions would emphasize a series of steps such as (a) writing down the problem, (b) inverting the divisor, and (c) multiplying the resulting fractions. A conceptual approach, in contrast, would use models and manipulatives (e.g., cuisenaire rods, fraction bars, drawings) to convey the relationships of concepts such as whole number division, multiplication, and fractions, to the concept division of fractions. A conceptual approach might also emphasize the relation of fractions to proportions.

Educational researchers have also shown an increased interest in the relationship between subject matter knowledge and the instructional processes that teachers use in conveying such knowledge to learners. This idea has become known as *pedagogical content knowledge.* According to Shulman (1987), pedagogical content knowledge is "the capacity of a teacher to transform the content knowledge he or she possesses into forms that are pedagogically powerful yet adaptive to the variations in ability and background presented by the students" (p. 15). Conveying subject matter in understandable forms is fundamental to the notion of pedagogical content knowledge and involves constructing alternate ways of representing ideas to students. Shulman advocates the use of analogies, metaphors, examples, demonstrations, and simulations for this purpose.

Since the introduction of pedagogical content knowledge, educational researchers have embarked on research programs for defining such knowledge in specific subject matter areas. Research in mathematics education has focused on the relationship between knowledge of the subject matter and the application of pedagogical content knowledge in prospective and experienced teachers (Ball, 1990; Ball & McDiarmid, 1988; Borko & Livingston, 1989; Cobb, Wood, & Yackel, 1990; Eisenhart et al., 1993; Even, 1993; Fennema, Frank, Carpenter, & Carey, 1993; Leinhardt & Smith, 1985; Livingston & Borko, 1989, 1990; Tirosh & Graeber, 1990). The research generally demonstrates that experienced teachers have richer, more well-instantiated cognitive representations about subject matter, pedagogical content knowledge, classrooms, and the nature of children than do inexperienced teachers. The breadth and depth of experienced teachers' knowledge enable them to provide instruction that is at once comprehensive and responsive to student needs. Furthermore, experienced teachers are able to re-

Rebecca L. Gomez, Department of Psychology, New Mexico State University; O. D. Hadfield, Department of Curriculum and Instruction, New Mexico State University; Lynn D. Housner, School of Physical Education, West Virginia University.

The alphabetical ordering of authorship reflects equal contributions to this research. This research was supported in part by a grant from the Center for Teaching Excellence, Eastern New Mexico University. We would like to thank Roger Schvaneveldt for his helpful comments regarding this article.

Correspondence concerning this article should be addressed to Rebecca L. Gomez, Department of Psychology 3452, Box 30001, New Mexico State University, Las Cruces, New Mexico 88003. Electronic mail may be sent via Internet to rebecca@crl.nmsu.edu.

spond effectively to a diverse set of student learning problems through flexible and improvisational application of robust, field-tested instructional routines and heuristics (i.e., pedagogical content knowledge built up in memory as a function of extensive teaching experience). In contrast, prospective and beginning teachers may have adequate subject matter knowledge, but when confronted with student comprehension problems, they often have difficulty generating alternate methods for conveying content knowledge. Furthermore, prospective and beginning teachers may believe in the value of teaching mathematics conceptually but may lack a firm enough foundation in the subject matter or the pedagogical content knowledge to teach for conceptual understanding. Finally, student teachers often encounter a conflict between the views espoused by university teacher educators and school administrators and the actual implementation of mathematics curricula in the schools. For example, teacher educators, central administrators, and principals may encourage teaching for conceptual understanding, but a number of factors may work against such efforts. Required standardized testing, management systems based on procedural knowledge, and the commitment of some cooperating teachers to teach "only" procedural knowledge reduces the opportunities for student teachers to acquire the pedagogical content knowledge necessary for teaching conceptual knowledge (Eisenhart et al., 1993).

Statement of the Problem

Livingston and Borko (1989) suggest that educational research should focus more on determining the types of teacher education experiences that aid novices in developing expert-like knowledge structures. Additionally, little is known about the actual content and organization of teacher educator knowledge and how that knowledge is imparted to prospective teachers (Floden & Klinzing, 1990; Howey & Zimpher, 1990). Finally, research on prospective teachers has not assessed the substance and organization (or structure) of knowledge obtained during teaching methodology classes in elementary mathematics. Instead, the research conducted thus far has focused on observations of student teachers and beginning teachers as they teach mathematics to children, and then infers the nature of conceptual and pedagogical content knowledge from these observations. Direct assessments of conceptual mathematical knowledge acquired during teacher education coursework and the application of this knowledge to student teaching has not been forthcoming. An attempt to assess and compare the conceptual and pedagogical content knowledge of teacher educators and prospective teachers has potential for contributing to an understanding of novice–expert differences in elementary math education, as well as methods for enhancing and developing such knowledge in prospective teachers.

Structural-Based Assessment

Conceptual and pedagogical content knowledge can be assessed structurally using the Pathfinder network scaling algorithm (Schvaneveldt, 1990; Schvaneveldt, Durso, & Dearholt, 1981, 1989) and an associated measure of network similarity. The Pathfinder network scaling algorithm generates empirically derived network representations of the associative structure among a set of concepts by taking relatedness ratings as input to use in generating networks where each concept is represented by a node, and the relations between concepts are represented by links between nodes. The algorithm operates by computing all paths between two nodes and includes a link between those nodes only if the link represents the most direct (or minimum length) path between the two concepts. In Pathfinder networks, highly related concepts are directly linked and less related concepts are separated by two or more links. The reader will recognize that Pathfinder networks are essentially semantic networks without labeled relations. Additionally, the primary method for constructing semantic networks is theoretical in nature (Collins & Loftus, 1975; Meyer & Schvaneveldt, 1976; Quillian, 1969), whereas Pathfinder generates networks empirically from estimates of psychological distance.

The first step in the general procedure for producing Pathfinder networks requires generating a list of concepts representing the domain of inquiry (in the present domain the concepts reflect the course instructor's view of conceptual and pedagogical content knowledge critical to elementary math education). Then relatedness data are obtained by having participants rate every possible pair of concepts on a numerical scale. Once produced, the networks can be used to compare and contrast differences in knowledge structure using numerical indices of similarity and by means of visual inspection.

Pathfinder has been used for representing knowledge in a number of domains, including basic research in memory (Cooke, 1992; Cooke, Durso, & Schvaneveldt, 1986), learning (Gomez & Schvaneveldt, 1994), problem solving (Durso, Rea, & Dayton, 1994), representation of belief in psychosomatic illness (Gomez, Schvaneveldt, & Staudenmayer, 1996), representation of the cognitive structures underlying expertise (Cooke & Schvaneveldt, 1988), and of particular relevance to the present study, structural assessment of knowledge growth for academic subject matter (Acton, Johnson, & Goldsmith, 1994; Goldsmith, Johnson, & Acton, 1991; Gonzalvo, Canas, & Bajo, 1994; Housner, Gomez, & Griffey, 1993a, 1993b; Johnson, Goldsmith, & Teague, 1994).

For example, Goldsmith et al. (1991) used a set-theoretic measure, C, to assess the similarity between instructor and student Pathfinder networks in the context of a university course on statistics and experimental design. C capitalizes on the degree to which a node in two graphs is connected to similar sets of nodes (see Goldsmith & Davenport, 1990, for a detailed discussion of the properties of this measure). The correlation between C and exam scores ($r = .74$) was a better predictor of performance than were correlations between the untransformed relatedness ratings and exam scores, or measures based on the use of multidimensional scaling (another structural-based assessment approach). Given previous research suggesting that multidimensional scaling captures global properties of the relations among concepts whereas Pathfinder captures the local properties (Cooke et al., 1986), these findings suggested that reducing the untransformed relatedness ratings to only the most salient relations among concepts, as is done in Pathfinder analysis, might be capitalizing on the "configural character" of domain knowledge. Goldsmith et al. (1991) argued that if knowledge is indeed based on sensitivity to the relationships among concepts, then a method that captures and represents the configural character of these relationships should be particularly useful.

Housner et al. (1993b) found similar results in a study involving prospective teachers enrolled in pedagogy courses. Knowledge of key pedagogical concepts was organized more consistently and was more similar to the instructor's organization after completion of the course than it was in the beginning of the course. Furthermore, Housner et al. (1993a) demonstrated that the prospective teachers' knowledge showed the same growth trend in relationship to the referent structures of other teacher experts, and to a composite network based on these teachers, thus adding significantly to the generalizability and content validity of this research (see Acton et al., 1994, for similar results in the domain of computer programming, as well as a more thorough treatment of the issue of appropriate referent structures).

Finally, Gonzalvo et al. (1994) used the configural properties of Pathfinder networks to predict the acquisition of conceptual knowledge in the domain of history of psychology. Gonzalvo et al. conducted a fine-grained analysis of the relationship of the concepts in Pathfinder networks to students' abilities to define these concepts. Each concept in the instructor and student networks was assessed structurally in terms of the concepts to which it was directly linked. Well-structured concepts in the student networks were defined as those sharing greater degrees of similarity (in terms of direct links to the same set of concepts) with those links found in instructor networks. Ill-structured concepts were defined as those sharing less student–instructor link similarity. Gonzalvo et al. found positive correlations between the goodness of students' definitions of concepts and structural similarity of their concepts to the concepts in the instructor networks, as well as an increase in the number of well-structured concepts at the end as compared with the beginning of the course.

In short, these studies suggest that a structural-based approach can provide a useful means for representing and assessing the growth of domain knowledge. Relatedness ratings represent a more indirect, but possibly objective, approach to discovering the structure of knowledge than do standard exams. Students are required to make a judgment about every pair of concepts; thus, they are tested over a wider range of comparisons than would be possible on an essay or multiple-choice exam. Because no context is provided other than that which arises from the relationship between the two concepts, relatedness ratings are not restricted to the context implied by the test question. Relatedness judgments also rely more on recognition than recall and thus are less susceptible to the problems associated with recall (such as forgetting and selective remembering). Another problem with standard exams has to do with devising simple and objective grading systems. This is especially true in the case of percentage-based exam scores, where student knowledge is assessed relative to other students rather than directly in relation to the course instructor. Thus, traditional exams may test the factual knowledge of one student relative to another but give little insight into how students organize knowledge in relation to the instructor or some other referent (Goldsmith et al., 1991). On this view, the use of relatedness ratings in combination with techniques for determining the most salient relationships among the concepts may provide a means for better reflecting a system of knowledge than would the more traditional method of eliciting answers to specific questions on an exam.

Ironically, although the motivation for using structural-based assessment has been largely justified on the grounds that traditional course exams are inadequate as indicators of the configural nature of knowledge (Goldsmith et al., 1991), validation of this approach has relied almost exclusively on comparisons with traditional course performance measures. The use of course exams as the validation criterion is a serious limitation in research on structural assessment. Therefore, showing that network similarity is a valid indicator of teaching competence would be an important extension of structural-based assessment approaches. It is also important to assess how exam scores fit into this relationship. Surprisingly little is known about how performance on course exams relates to the ability to generalize knowledge in an applied setting.

Participants in the present study were prospective math teachers enrolled in an elementary mathematics methods course, the course instructor, and four other experienced math teacher experts. Given the results of previous research using the Pathfinder methodology, we expected to find a positive relationship between course performance and network similarity to the course instructor (Goldsmith et al., 1991; Housner et al., 1993b). We also expected to find a high degree of between-expert similarity and a positive relationship between course performance and network similarity to a composite of the experts' network (Acton et al., 1994; Housner et al., 1993a). However, our primary research objective was to examine students' ability to apply knowledge obtained in a teaching methodology course in the context of a simulated teaching task. We were particularly interested in determining the relationship between the level of subject matter understanding as reflected in network similarity and the ability to transform and convey this knowledge in a teaching situation. In other words, we wanted to know how network similarity relates to procedural knowledge, conceptual understanding, and the ability to apply pedagogical content knowledge. We took the next reasonable step in a line of research (Acton et al., 1994; Goldsmith et al., 1991; Gonzalvo et al., 1994; Housner et al., 1993a, 1993b; Johnson et al., 1994) by moving beyond *recall* of knowledge (as reflected in traditional course assessment measures) to the *application* of knowledge (as reflected in a simulated teaching task).

Method

Participants

The course instructor. The course instructor taught an elementary mathematics methods course at New Mexico State University. The instructor was selected because of his commitment to conceptually based approaches to mathematics instruction and his extensive experience as a teacher educator (7 years) and as a teacher of elementary children (11 years).

The teacher educators. Four mathematics teacher educators from different universities were recruited to participate. Each of the teacher educators had a doctorate in mathematics education, 5 years experience as a teacher educator, and a commitment to conceptually based approaches to teaching mathematics.

The undergraduate students. Fifty-three students who were enrolled in the elementary mathematics teaching methodology course participated.

The Teaching Methodology Course

The focus of the course was on conceptually based approaches to teaching elementary mathematics. The course consisted of assigned readings, lectures, and class discussions. A critical component of the course was a laboratory section in which prospective teachers were provided with modeled teaching demonstrations pertaining to conceptually based approaches and were then required to apply these ideas in microteaching experiences with fellow students. The prospective teachers also developed lesson plans and teacher-made activities in the laboratory and were given appropriate feedback on these assignments relative to conceptually based teaching. The final course grade was based on performance in the laboratory sessions and objective exams. Performance in the laboratory sessions accounted for 50% of the course grade; objective exams accounted for the remaining 50%.

Materials

The initial step in the study was to delineate the knowledge structure of the course instructor. The researchers conducted interviews with the course instructor, focusing on his perceptions of effective mathematics teaching. The critical knowledge identified by the course instructor consisted of 27 terms, including teaching topics, concepts relevant to each topic, and manipulatives used for teaching concepts (see Table 1). For instance, concepts of importance for teaching whole number operations are algorithms, regrouping, partitive/measurement, and basic facts. Appropriate manipulatives for demonstrating these concepts are cuisenaire rods and arrays. Of the 27 terms identified, 7 were teaching topics, 13 were concepts, and 7 were manipulatives. Manipulatives and materials that can be used to convey more than one concept are repeated in Table 1.

Procedure

Representing the course instructor's knowledge. The instructor used a computerized program to rate every possible pairwise combination of the terms on a 9-point relatedness scale (1 = *unrelated* and 9 = *highly related*). The pairs were presented in random order for rating. Once the ratings were obtained, the data were converted to distance measures by subtracting each rating from 10. The Pathfinder network algorithm was then used to generate a network representing the course instructor's organization of knowledge. The parameters used to compute the network were set at r = infinity and $q = n - 1$, where n refers to the number of terms in the data (in the present study, $n = 27$). The r parameter was chosen to match the ordinal properties of the data, and the q parameter was chosen for the purpose of generating the sparsest network possible from the given data. The first parameter, the Minkowski r metric, determines how distance between two nodes not directly linked is computed. When r = infinity, the

Table 1
Key Math Concepts Used in the Experiment

Topics	Concepts	Manipulatives/materials
Prenumeration	Classification/sorting comparison	Attribute/blocks Discrete/continuous materials
Place value	One-to-many correspondence Proportional/nonproportional	Base 10 blocks
Whole number operations	Algorithms Regrouping Basic facts Partitive/measurement	Cuisenaire rods Arrays
Fractions	LCM/GCF Rename	Shaded bars/squares/ circles
Decimals		Shaded bars/squares/ circles
Percents	Ratios and proportions	Shaded bars/squares/ circles
Measurement	Area Volume	Color cubes

Note. LCM/GCF = lowest common multiple/greatest common factor.

length of a path is equal to the magnitude of the maximum link on the path. This means that only ordinal assumptions are required of the data because when r = infinity, the same links will be included for any monotonically increasing transformation of the data. The second parameter, q, limits the number of links allowed in searching for shorter alternative paths. When $q = n - 1$, where n equals the number of concepts being compared, there is essentially no limit on the number of links allowed in paths because the longest possible path has $n - 1$ links (see Schvaneveldt et al., 1989, for more information on choosing r and q parameters).

Eliciting knowledge from the teaching educators. The four experienced math teacher educators rated the set of concepts using the same procedure as the course instructor. Pathfinder networks were generated for each teacher educator in the same manner as for the course instructor. A composite expert network was generated by converting each teacher educator's ratings to z scores and then averaging the converted ratings. The individual and composite networks were then compared with the network structure generated by the course instructor.

Eliciting knowledge from the prospective teachers. During the last week of class, all 53 students in the math education course rated the same set of concepts as the course instructor and the math teacher educators. The relatedness ratings were used to generate networks for all students in the class. A measure of network similarity (described below) was used to compare student networks with the networks for the course instructor and the teaching educators. Network similarity to the course instructor was used to identify the top 10 and bottom 10 students in the class. Of these, 7 in each group were recruited to participate in a simulated teaching task. These students returned 3 months later to participate in the simulated teaching task and were each paid $15 for their efforts.

In the simulated teaching task, the prospective teachers were asked to explain and demonstrate how they would teach three specific areas of elementary school mathematics to elementary school children. The three areas of math concepts chosen for this task were whole number operations, fractions, and measurement problems. Three problems were generated for each math concept. The tasks and protocol for the simulated teaching are presented in Table 2. Students were told to use appropriate manipulatives to teach each concept and also to work each problem out by hand. The purpose of the latter instruction was to determine whether the students could produce the correct solution to the problem. The participants were also asked to describe real-world situations where they would apply multiplication and division of fractions and where they would have to find perimeter and volume. Participants were given their choice of eight types of manipulatives to use in teaching the problems in the simulated teaching task. The manipulatives were Base 10 blocks, shaded fraction bars, color tiles, color cubes, cuisenaire rods, decimal squares, pattern blocks, and color counters. The latter three manipulatives were included as distracter items. The experimenter read the problem to the student and the students were videotaped as they simulated teaching the concept to the experimenter. Students were randomly assigned to each of the three experimenters. The experimenters were "blind" as to whether the student was high or low on network similarity to the course instructor.

Scoring

Simulated teaching task. After the simulated teaching tasks were videotaped, each tape was rated by four independent raters who were paid for their efforts. Two of the raters were instructors for the elementary mathematics methods course at New Mexico State University. The third rater taught math for education majors in the mathematics department of the university, and the fourth rater was a public school teacher who had received awards for mathematics teaching in the public school system. Each problem was scored according to four levels of performance difficulty, as shown in Table 3. The first level of mastery reflects surface knowledge of the procedures in this particular domain of math problems. The second level reflects the ability to remember which manipulatives are used for teaching particular classes of problems, but provides no indication of whether the prospective teacher can use the manipulative correctly. The third and fourth levels of mastery reflect a deeper conceptual understanding in the ability to use manipulatives to explain math problems for conceptual understanding (Level 3) and in the ability to generate real-world analogies (Level 4). That is, Levels 3 and 4 reflect degree of conceptual and pedagogical content knowledge, whereas Levels 1 and 2 reflect surface and procedural knowledge. Interrater reliability between all pairs of observers ranged from .62 to .92, with a mean of .83. These levels of agreement were considered acceptable. The observers' ratings were then aggregated for each student by computing the average of the raters' scores for each problem, at each level of performance.

Network similarity. Pathfinder networks were compared using a network similarity index called NETSIM. NETSIM is based on the expected similarity between networks and is computed in the following manner. First, the observed similarity is computed by dividing the number of links shared by both networks (those in the intersection) by the number of links in either network (those in the union). Next, because the probability that two networks (Net 1 and Net 2) will share k links can be computed from the hypergeometric probability distribution, this information can be used to compute the expected similarity of two random networks.[1] The expected similarity is subtracted from the observed similarity between two networks to get the NETSIM index, which is relative to the chance level of similarity (or NETSIM = 0). For instance, with $n = 27$ concepts, if two networks (Net 1 and Net 2) contain $L_1 = 66$ and $L_2 = 55$ links, respectively, and if $k = 21$ of the links are shared, then there are 100 links in the union (i.e., $66 + 55 - 21 = 100$). The probability of sharing 21 links is .00013, observed similarity is 0.210 (i.e., 21/100), expected similarity is .094, and NETSIM is

[1] Applying the hypergeometric probability distribution to graph similarity yields the equation,

$$p(I = k) = \frac{\binom{L_1}{k}\binom{N - L_1}{L_2 - k}}{\binom{N}{L_2}},$$

where maximum $(0, L_1 + L_2 - N) \leq k \leq \text{minimum}(L_1, L_2)$. In this equation, $p(I = k)$ refers to the probability that Net 1 and Net 2 will share k links, N denotes the total number of links possible in a network (given n concepts, $N = n(n - 1)/2$), and L_1 and L_2 refer to the number of links in Net 1 and Net 2, respectively. Expected similarity, E[Sim], is then obtained by the usual method for computing expected values:

$$\mathrm{E}[Sim] = \sum_{k=\min(I)}^{\max(I)} \mathrm{p}(I = k) \cdot \frac{k}{(L_1 + L_2 - k)},$$

where E[Sim] = expected similarity, min (I) = maximum $(0, L_1 + L_2 - N)$ and max(I) = minimum (L_1, L_2).

Table 2
Problems Used in the Simulated Teaching Task

Task	Problem
Whole number operations	
Subtraction	Show how the subtraction algorithm works using the problem: 53 − 28 = ?
Multiplication	Show how the multiplication algorithm works using the problem: 3 × 47 = ?
Division	Show how the division algorithm works using the problem: 736 ÷ 3 = ?
Fractions	
Subtraction	Show how to subtract the fractions 1¼ − ⅓.
Multiplication	Show how to multiply the fraction ⅔ × ¾. Describe a real world situation where you would multiply two fractions like the ones here.
Division	Show what it means to divide the fractions 1⅔ ÷ ⅓. Describe a real world situation where you would divide two fractions like the ones here.
Measurement	
Perimeter	Demonstrate how you would teach the concept of perimeter using the given shape. (Students were given the outline of a T-shaped area.) Describe a real world situation where you would find the perimeter of an irregular shape like the one here.
Area	Demonstrate how you would teach the concept of area using the given shape. (Students were given the outline of an L-shaped area.)
Volume	Demonstrate how you would teach the concept of volume using the given shape. (Students were given the outline of an L-shaped area.) Describe a real world situation where you would find the volume of a shape like the one here.

0.116. A positive NETSIM value indicates a greater degree of similarity between two networks than that expected by chance. Alternately, a negative NETSIM value means that the observed network similarity is less than that expected by chance. Statistical significance can be used as a criterion for determining whether the NETSIM values statistically exceed the degree of network similarity expected by chance. Using this index, one would expect to find a higher degree of network similarity between high-knowledge students and the course instructor or the teacher experts than between low-knowledge students and the referent networks.

Although NETSIM is conceptually similar to the *C* measure used in previous research (see Acton et al., 1994; Goldsmith & Davenport, 1990; Goldsmith et al., 1991; Gonzalvo et al., 1994; Housner et al., 1993b; Johnson et al., 1994), it should be noted that the measures are not identical. The *C* measure is based on the similarity of the "neighborhoods" of the corresponding nodes in two networks, whereas NETSIM is derived from the similarity of sets of links found in two networks. In the computation of *C*, the neighborhood of each node in a network is taken to be the set of nodes connected to the node in the network. For the corresponding nodes in two networks, the similarity of their neighborhoods is computed by the ratio of the number of nodes that are in both neighborhoods over the number of unique nodes in either neighborhood (the number of nodes in the intersection of the neighborhoods over the number in the union). These values are computed for every node in the network, and *C* is the mean of these node-based values. NETSIM is similar in spirit, but the computation is based on the networks as wholes rather than node by node. The primary advantage of NETSIM is that a chance value can be computed, whereas attempts to find the distribution of *C* have not

Table 3
Criteria Used by Expert Raters to Score Performance on the Simulated Teaching Task

Level of performance	Criteria
Level 1: Performed algorithm correctly	0 = *No* 1 = *Yes*
Level 2: Selected appropriate manipulative	0 = *No* 1 = *Fair* 2 = *Best*
Level 3: Used manipulative to explain concept effectively	0 = *Ineffective*
	1 = *Moderately effective* (some confusion, lacked confidence, did not relate explanation to algorithm)
	2 = *Effective* (clear, confident, and related explanation to algorithm)
Level 4: Explained real world application of concept	0 = *None*
	1 = *Application suggested but lacked accurate explanation*
	2 = *Application accompanied by accurate and detailed explanation*

been successful. NETSIM values can be interpreted relative to a statistical distribution as well as relative to each other. C can be interpreted only relative to other values. NETSIM values are quite similar to C, with mean correlations of .93 in this study (the correlations ranged from .80 to .99, depending on which referent networks were used in the comparison).

Results

The Pathfinder network generated from the course instructor's ratings is shown in Figure 1. The network reflects many of the same relationships shown in Table 1. This is particularly interesting given that the course instructor constructed the table analytically, whereas the network was constructed from the course instructor's judgments of relatedness. For example, the concepts "LCM/GCF" (lowest common multiple/greatest common factor), "Rename," and "Shaded bars/squares/circles" are all related to the topic "Fractions" and these relationships also show up as links in the course instructor's Pathfinder network. One expects substantial overlap between the table and the network, given that both representations were generated from the same source (i.e., the course instructor's knowledge); however, the network shown in Figure 1 is an empirical demonstration of this point. It should be noted that the primary source of information in Pathfinder networks is in the presence or absence of a link. Spatial layout and link lengths are determined arbitrarily and therefore have no meaningful interpretation.

Next, the generalizability of the course instructor's knowledge was assessed by comparing his knowledge structure with those of the four other experienced teacher educators. NETSIM values comparing the Pathfinder networks of the course instructor, the four experienced teacher educators, and the composite networks are shown in Table 4. The corresponding C values are also shown in Table 4. The averaged ratings used to compute the five different composite networks changed as a function of which teacher educator was the focus of the comparison. For example, the composite network used to compute network similarity in comparison with the course instructor was generated from the z-transformed ratings of Teacher Educators 1–4. In contrast, the composite used to investigate network similarity for the first teacher educator was based on the ratings obtained from the course instructor and Teacher Educators 2–4. The probability of obtaining NETSIM values as large or larger than those shown in Table 4 indicates levels of agreement considerably above chance (all $ps < .001$).

Next, NETSIM values were computed for all 53 students in the math pedagogy course. Students' networks were compared with the six referent networks: the course instructor's network, those from the four teacher educators, and the composite network based on the averaged ratings of the four teacher educators. The resulting NETSIM values are shown in Table 5. The analyses reported were all performed on NETSIM values. Analyses were also performed using C values that, although not reported here, were consistent with the analyses using NETSIM. Results for the seven students with the highest NETSIM scores and the seven students with the lowest NETSIM scores are shown in Table 5. The NETSIM values shown in Table 5 were all statistically ($p <$ .003) greater than baseline (NETSIM = 0). A repeated measures ANOVA with six levels of referent structure was computed for all students to determine whether NETSIM values varied as a function of referent. The ANOVA revealed a significant main effect of referent, $F(5, 260) = 4.71$, $MSE = 0.0008$, $p < .001$. Post hoc t tests were

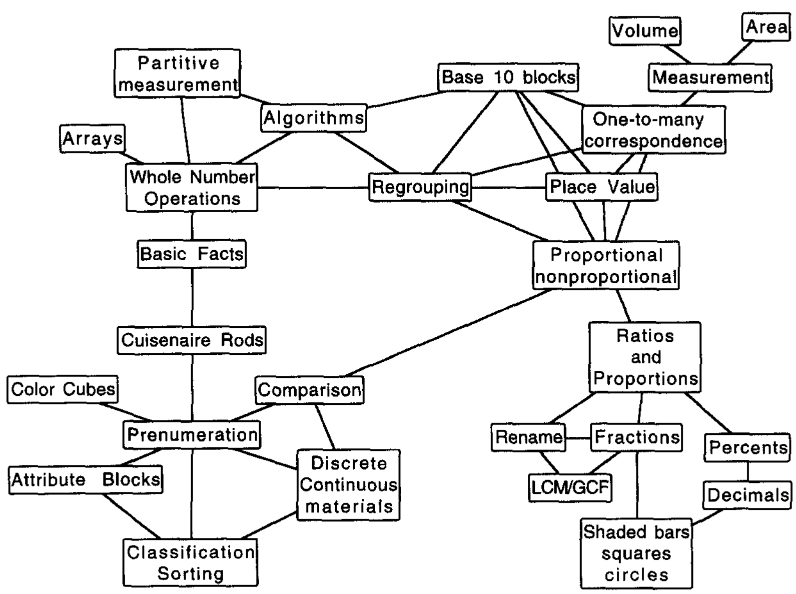

Figure 1. Pathfinder network for the course instructor ($q = n - 1$, r = infinity). LCM = least common multiple; GCF = greatest common factor.

Table 4
NETSIM and C Values for Comparisons Between the Pathfinder Networks of the Course Instructor, the Four Experienced Teacher Educators (TE 1–4), and the Respective Composite Networks

Teacher educators	NETSIM (Observed minus expected similarity) 1	2	3	4	5	6
1. Instructor (42 links)	—					
2. TE 1 (55 links)	—					
NETSIM	.154	—				
C	.200	—				
3. TE 2 (66 links)						
NETSIM	.288	.116	—			
C	.384	.215	—			
4. TE 3 (66 links)						
NETSIM	.176	.104	.190	—		
C	.302	.190	.319	—		
5. TE 4 (91 links)						
NETSIM	.165	.128	.219	.208	—	
C	.272	.242	.351	.335	—	
6. Composites (27–31 links)						
NETSIM	.249	.133	.222	.216	.166	—
C	.381	.174	.296	.312	.251	—

Note. The number of links in a given network are shown in parentheses. NETSIM and *C* are both network similarity indexes.

conducted for the purpose of determining whether NETSIM to any one referent was better than NETSIM to the other referent structures. Because of the number of tests involved in these comparisons, the significance criterion was adjusted to control for familywise error rate ($EF/k = .05/15 = .003$). Post hoc *t* tests, comparing NETSIM to each referent with NETSIM to every other referent, failed to result in statistically significant differences (all $ps > .03$). A similar analysis was conducted using only the 14 students who participated in the simulated teaching task. A 2 (knowledge: high vs. low) × 6 (referent structure) ANOVA, where knowledge was treated as a between-subjects variable and referent structure was a within-subject variable, resulted in a main effect of knowledge, $F(1, 12) = 99.27$, $MSE = 0.0022$, $p < .001$; a main effect of referent, $F(5, 60) = 2.74$, $MSE = 0.0006$, $p = .027$; and a Knowledge × Referent Structure interaction, $F(5, 60) = 3.80$, $MSE = 0.0006$, $p = .005$. As with the analysis using all 53 students, post hoc *t* tests were conducted to determine whether NETSIM was statistically higher to any one referent than to the other referents. With the significance criterion adjusted to control for familywise error rate ($EF/k = .05/15 = .003$), post hoc *t* tests, comparing NETSIM to each referent with NETSIM to every other referent, resulted in only one difference. For

Table 5
Means and Standard Deviations for NETSIM and C Comparisons to the Six Referent Networks, Presented Separately for All Students and for High- and Low-Knowledge Students

Group and value	Course instructor	TE 1	TE 2	TE 3	TE 4	TEs 1–4
All students ($n = 53$)						
NETSIM	.12 (.06)	.10 (.05)	.11 (.05)	.11 (.05)	.12 (.04)	.12 (.06)
C	.22 (.07)	.20 (.05)	.23 (.06)	.23 (.05)	.26 (.05)	.20 (.06)
High-knowledge students ($n = 7$)						
NETSIM	.19 (.03)	.13 (.03)	.16 (.02)	.16 (.05)	.14 (.04)	.16 (.03)
C	.28 (.02)	.23 (.03)	.28 (.03)	.27 (.05)	.29 (.05)	.23 (.02)
Low-knowledge students ($n = 7$)						
NETSIM	.05 (.02)	.04 (.02)	.05 (.03)	.06 (.01)	.08 (.03)	.05 (.01)
C	.14 (.01)	.15 (.03)	.18 (.03)	.20 (.02)	.24 (.05)	.13 (.01)

Note. TE = teacher educator. NETSIM and *C* are both network similarity indexes. Numbers in parentheses are standard deviations.

the high-knowledge students, NETSIM was significantly higher to the course instructor than to Teacher Educator 1, $t(12) = 3.80$, $p = .003$ (all other $ps > .04$). Post hoc t tests for low-knowledge students showed no differences in NETSIM values as a function of referent structure (all $ps > .03$). Most important, there were no statistical differences in NETSIM values to the composite network compared with NETSIM values to the course instructor (for all or for high- or low-knowledge groups), suggesting that both networks were equally predictive of course performance.

Pearson product–moment correlations were computed between course performance measures and the students' NETSIM values. The correlations for the entire class and the correlations for the 14 students participating in the simulated teaching task are shown in Table 6 (note that the latter correlations include two extreme groups of $n = 7$). Mean test average for all 53 students was 86.14 percentage points ($SD = 8.68$; range = 61.00–98.75), mean lab score was 96.80 ($SD = 3.10$; range = 87.10–100.00), and mean final grade was 91.47 ($SD = 5.01$; range = 75.98–99.22). Mean test average for the 14 students who participated in the simulated teaching task was 87.39 percentage points ($SD = 6.82$; range = 70.75–95.00), mean lab score was 95.46 ($SD = 2.87$; range = 88.39–99.35), and mean final grade was 91.40 ($SD = 3.87$; range = 82.79–96.28). Although correlations between NETSIM and lab scores for the entire class were numerically lower than correlations between NETSIM and the other two course performance measures, all but one correlation reached statistical significance. The correlations between NETSIM and lab scores were somewhat higher when only the 14 students participating in the teaching demonstration were included in the analysis; three of the six correlations reached statistical significance (NETSIM to Teacher Educators 1 and 3 and to the composite network). When the entire class was included in the analysis, all of the correlations of NETSIM with test average and final grade were statistically significant. However, the correlations between test average and NETSIM were numerically lower than the correlations between final grade and NETSIM. NETSIM to the course instructor, and also to the composite network, accounted for approximately 30% of the variance in final course grade. Interestingly, the correlation of final grade with NETSIM to Teacher Educator 2 was slightly higher than the correlations for final grade with NETSIM to the course instructor and with the composite. When only the 14 students participating in the simulated teaching task were included in the analyses, the correlations of test average with NETSIM were also numerically lower than were the correlations of final grade with NETSIM. In this case, NETSIM to the course instructor accounted for 55% of the variance in final course grade and NETSIM to the composite network accounted for 50% of the variance in final course grade.

These results (for all students and for the 14 participating in the teaching demonstration) replicate the findings reported by Acton et al. (1994) and Housner et al. (1993a) by demonstrating how network similarity to a composite of experts can be as predictive of a course performance variable (such as final grade) as similarity to the course instructor ($rs = .74$ and $.71$, respectively). Such a result is important because it rules out the possibility that the student teachers are merely organizing their knowledge in terms of the particular views espoused by the course instructor. Instead, this result suggests that student teachers are learning something more general with regard to the way teacher educators organize their knowledge.

Given that the present results replicate those found earlier in the literature, both in terms of predicting course performance (Goldsmith et al., 1991; Gonzalvo et al., 1994; Housner et al., 1993b) and in terms of generalizing to a composite network based on experts other than the course instructor (Acton et al., 1994; Housner et al., 1993a), the next objective was to assess performance on the simulated teaching task and its relationship to network similarity.

Table 7 shows the percentage of points obtained for the high- and low-knowledge groups as a function of problem type (whole number operations, measurement problems, and fractions) and performance on the simulated teaching task (Levels 1–4). Note that the cell for Level 4 ("Example") is empty for whole number operations. This is because students were not asked to provide real-world analogies to these problems. Table 7 shows that students were generally

Table 6
Correlations Between Course Performance Measures and NETSIM to the Six Referent Networks, for All Students and Those Participating in the Simulated Teaching Task

Course performance variables	Course instructor	TE 1	TE 2	TE 3	TE 4	TEs 1–4
			All students ($N = 53$)			
Final grade	.55***	.31*	.63***	.54***	.56***	.54***
Test average	.52***	.28*	.60***	.50***	.54***	.51***
Lab scores	.32*	.20	.36**	.35*	.30*	.31*
		Students participating in the simulated teaching task ($n = 14$)				
Final grade	.74**	.54*	.72**	.61*	.60*	.71**
Test average	.64*	.38	.66*	.46	.54*	.56*
Lab scores	.48	.57*	.37	.55*	.34	.58*

Note. TE = teacher educator. NETSIM is a network similarity index.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 7
Mean Percent Correct Scores and Standard Deviations for High- and Low-Knowledge Groups at Four Different Levels of Mastery, Across Three Problem Types

Level of mastery	Whole number operations High ($n = 7$)	Whole number operations Low ($n = 7$)	Measurement problems High ($n = 7$)	Measurement problems Low ($n = 7$)	Fractions High ($n = 7$)	Fractions Low ($n = 7$)
Algorithm	100 (0)	100 (0)	90 (25)	76 (32)	81 (26)	71 (30)
Manipulative	95 (13)	90 (13)	57 (9)	64 (22)	93 (19)	60 (36)
Explanation	52 (11)	38 (8)	62 (18)	40 (13)	21 (23)	5 (8)
Example	—	—	50 (14)	46 (17)	11 (28)	4 (9)

Note. Numbers in parentheses are standard deviations.

able to perform at the first level of mastery. That is, most students were able to solve the problems procedurally. In cases of whole number operations, both high- and low-knowledge students were able to choose an appropriate manipulative for teaching the problem (Level 2). High-knowledge, but not low-knowledge, students were able to choose an appropriate manipulative for teaching fractions. Both groups performed poorly when selecting an appropriate manipulative for teaching measurement problems. Performance was generally poor when teaching for conceptually based understanding (Level 3) and when attempting to generate real-world analogies or examples (Level 4).

The overall analysis was broken into two separate ANOVAs because of the empty cells for whole number operations at the fourth level of mastery. A 2 (high vs. low knowledge) × 3 (problem type) × 3 (Mastery Levels 1–3) ANOVA, where knowledge was a between-subjects variable and problem type and mastery level were within-subject measures, resulted in main effects of knowledge, $F(1, 12) = 7.81$, $MSE = 571.52$, $p = .016$; problem type, $F(2, 24) = 17.32$, $MSE = 359.49$, $p < .001$; and level of mastery, $F(2, 24) = 60.92$, $MSE = 483.44$, $p < .001$. The main effects of problem type and level of mastery were mediated by a Problem Type × Mastery Level interaction, $F(4, 48) = 8.33$, $MSE = 322.34$, $p < .001$. The interaction is due primarily to the finding that the prospective teachers had far more difficulty explaining fractions for conceptual understanding than explaining whole number operations and measurement problems. Students also had relatively more difficulty selecting the appropriate manipulative for measurement problems. A 2 (high vs. low knowledge) × 2 (problem type) ANOVA for performance at Level 4, where knowledge was a between-subjects variable and problem type was a within-subject variable, resulted in a main effect of problem type, $F(1, 12) = 29.94$, $MSE = 394.34$, $p < .001$, but no effect for knowledge or interaction of these variables (both Fs < 1). Prospective teachers were far better at generating real-world examples for measurement problems than for fractions. This finding is consistent with that reported in the previous analysis, where prospective teachers had more difficulty teaching fractions for conceptual understanding as compared with teaching measurement problems. Although prospective teachers were far more successful in selecting the correct manipulative for teaching fractions as compared with measurement problems, this apparent success did not factor into their ability to teach fractions, as indicated by performance on Levels 3 and 4 of the simulated teaching task.

Finally, correlations were calculated to clarify the relationship between network similarity, course performance, and performance on the simulated teaching task. We were particularly interested in determining how course performance and network similarity mapped onto level of teaching skill. Table 8 shows the correlations of simulated teaching performance (collapsed across problem type) to final grade, test average, lab score, network similarity to the course instructor, and network similarity to the composite network. Levels 1, 2, and 4 performance failed to correlate

Table 8
Correlations Between Performance at Each Level of Mastery on the Simulated Teaching Task and Network Similarity to the Course Instructor, the Composite Network of the Four Teacher Educators, Final Course Grade, Test Average, and Lab Score (n = 14)

Level of mastery	Course instructor NETSIM	Composite NETSIM	Final grade	Test average	Lab score
Level 1: Algorithm	.42	.50	.35	.28	.28
Level 2: Manipulative	.31	.26	.37	.22	.49
Level 3: Explanation	.69**	.68**	.62*	.52	.45
Level 4: Real-world example	.11	.17	−.05	−.11	.11

Note. NETSIM is a network similarity index.
* $p < .05$. ** $p < .01$.

significantly with either NETSIM or course performance scores. However, the ability to teach for conceptually based understanding (Level 3) was significantly correlated with the two NETSIM measures and final course grade. NETSIM to the course instructor and the composite network accounted for 48% and 46% of the variance in Level 3 performance, and final course grade accounted for 38% of the Level 3 performance variance.

To obtain a better understanding of the relationship between network similarity and course grades as predictors of simulated teaching performance, we computed the partial correlations of Level 3 teaching performance with NETSIM to the course instructor, NETSIM to the composite, final grade, test average, and lab scores, with the variance contributed by the latter five indices individually held constant (see Table 9). If NETSIM is accounting for variance not accounted for in course performance measures, then NETSIM should continue to predict Level 3 teaching performance, even when the variance due to course measures is partialled out. Alternately, if NETSIM shares a significant degree of variance with a particular performance measure, then partialling the measure out should result in a loss of predictiveness. The second row in Table 9 shows the Level 3 predictiveness of NETSIM to the course instructor, with NETSIM to the composite network, final course grade, test average, and lab score individually held constant. It can be seen that NETSIM to the course instructor correlates significantly with Level 3 performance when test average and lab scores are held constant, but not when NETSIM to the composite or final course grade are partialled out. The third row in Table 9 shows a similar pattern of results. That is, NETSIM to the composite network is a significant predictor of performance on Level 3 of the simulated teaching task when either test average or lab score is partialled out, but not when NETSIM to the course instructor or final grade is held constant. In contrast, the fourth, fifth, and sixth rows in Table 9 show how when either of the two NETSIM measures is individually partialled out, course performance measures show no correlation with the ability to teach math concepts for understanding. This pattern of findings suggests that NETSIM measures based on comparisons of Pathfinder networks capture unique variance not captured in exams or lab scores.

Finally, NETSIM to the course instructor and final course grade were included in a multiple regression equation to determine whether these factors together would be a better predictor of Level 3 teaching performance than either factor taken alone. Combining these two factors resulted in $R = .71$ ($R^2 = .51$), $F(2, 11) = 5.74$, $MSE = 65.06$, $p = .019$. Although using either NETSIM or final course grade in the regression equation significantly predicted Level 3 teaching performance, $R^2 = .48$, $F(1, 11) = 11.20$, $MSE = 65.06$, $p = .006$ for NETSIM, and $R^2 = .38$, $F(1, 11) = 7.68$, $MSE = 65.06$, $p = .017$ for final course grade, the additional improvement due to adding the other independent variable was not large enough to be statistically significant, $F(1, 11) = 0.29$, $MSE = 65.06$, when adding final course grade to NETSIM, and $F(1, 11) = 3.81$, $MSE = 65.06$, when adding NETSIM to final course grade. Thus, these two variables appear to convey overlapping information with regard to simulated teaching performance.

Discussion

The objectives of the research reported here were twofold. The first objective was to replicate previous research using the Pathfinder network scaling methodology (Schvaneveldt, 1990; Schvaneveldt et al., 1981), but in the domain of math education. The results showed that the course instructor shared a high degree of structural knowledge with the other experienced teacher educators. Furthermore, network similarities to the composite network, generated from the experienced teacher educators' concept ratings, were virtually as predictive of the prospective teachers' course grades as were network similarities to the course instructor, replicating the results of Acton et al. (1994) and Housner et al. (1993a). Network similarity to the four teacher educators was also predictive of course grades.

The second objective of the study was to take the next logical step in research on structural-based assessment by examining the relationship between knowledge structure and the ability to apply this knowledge. Previous studies have used course grades as validation criteria for structural-based assessment. However, as the research in this particular study demonstrates, the relationship between course

Table 9
Partial Correlations Between Knowledge Indexes as a Predictor of Performance on the Explanation Component (Level 3) of the Simulated Teaching Task With Every Other Knowledge Index Held Constant (n = 14)

Knowledge index used as predictor (Level 3: Explanation)	1	2	3	4	5	6
1. None						
2. NETSIM to instructor	.69**	—	.22	.44	.55*	.61*
3. NETSIM to composite	.68**	.03	—	.42	.54*	.57*
4. Final grade	.62*	.23	.28	—	.44	.52
5. Test average	.52	.15	.24	−.21	—	.52
6. Lab score	.45	.18	.09	.20	.44	—

Note. NETSIM is a network similarity index.
$*p < .05$. $**p < .01$.

performance and teaching skill is somewhat tenuous. We assessed teaching performance on four different levels of mastery and discovered that although most participants were able to solve the problems procedurally and many were able to select the correct manipulative for demonstrating the problems, students generally had greater difficulty translating their knowledge of procedures into conceptually based explanations and real-world analogies to problems. Interestingly, network similarity was as sensitive an index of the ability to teach for conceptual understanding as was final course grade, and more sensitive than were exams and lab scores. When test average and lab scores were individually held constant in partial correlations, NETSIM to the course instructor and the composite network were predictive of performance on the simulated teaching task, whereas the opposite was not true. That is, partialling out the variance due to NETSIM to either referent network essentially eradicated the relationship between test average, lab performance, and the ability to convey math problems for conceptual understanding. When final course grade was held constant, NETSIM to the course instructor and the composite were no longer predictive of performance on the simulated teaching task, suggesting that NETSIM and final course grade were tapping shared variance. These results demonstrate two important points. The first is that network similarity can be useful for assessing a specific aspect of knowledge, in this case conceptual and pedagogical content knowledge. Secondly, the information captured using relatedness ratings and the Pathfinder methodology was just as effective at measuring key knowledge as final course grade. That Pathfinder was a more sensitive predictor of simulated teaching performance than test average is consistent with the view that standard exams may fail to represent the configural nature of knowledge captured in structural approaches (Goldsmith et al., 1991).

A potential drawback of this study is in the use of two very small groups on extreme ends of a continuum, with each group consisting of only seven students. Given the resources we had available, it was possible to test only a small number of participants. Twenty students (10 with low network similarity to the course instructor and 10 with high network similarity to the instructor) were originally targeted for the simulated teaching task. It was impossible to obtain all 20 students due to the constraints that students either be in the upper or lower quadrant of NETSIM scores and not be concurrently fulfilling their student teaching requirements. The latter constraint served the purpose of avoiding a confound due to differential degrees of teaching experience. Under any circumstances the use of such small groups raises questions regarding the stability and generalizability of the results, but the fact that we show statistical differences with such a small sample suggests that the effects are sizable. More important, however, the correlations of teaching performance with network similarity to the course instructor accounted for nearly 50% of the variance observed in the data. Under such circumstances it is hard to imagine how the use of additional participants could change the conclusions drawn from this study.

There may be another problem resulting from the selection procedure, as it relates to predicting simulated teaching performance. Because high- and low-knowledge students were selected on the basis of NETSIM and not on the basis of course grades, there could be an advantage for the former compared with the latter. If scores were more variable on NETSIM than on the traditional course measures, one might expect higher correlations between NETSIM and simulated teaching than between course performance and simulated teaching. However, when variance in NETSIM to the course instructor for students with the seven highest and seven lowest final grades ($M = .131$, $S^2 = .00422$) was compared to that of the seven high- and seven low-knowledge students ($M = .118$, $S^2 = .00593$), the difference was statistically indistinguishable, $F(13, 13) = 0.71$, $p > .05$.

Why would network similarity be a more sensitive index of conceptually based understanding than exams or lab scores? One explanation might be that microteaching performance was assessed in the lab portion of the course immediately after new concepts were presented by the course instructor. This meant that students were relying on recent memory to perform the lab exercises. Thus, the lab score was reflecting short-term retention whereas network similarity was reflecting longer term retention. However, this does not explain the failure of exams to contribute to measurement of teaching skill. A second factor could be a lack of variance in exams or lab scores. However, examination of the distributions associated with course performance suggests otherwise. Another explanation is that the exams in this class were clearly designed to tap declarative knowledge, whereas the lab scores were based on procedures relevant to teaching skill. Some combination of declarative and procedural knowledge, reflected in final course grade, is clearly more predictive of teaching skill than either source of knowledge alone. Such a combination may also be more reliable because it is based on two "subtests," (i.e., labs and exams), rather than either one by itself. With regard to network similarity, associative networks might provide a better means for representing a mental model of a domain than does assessment based on short-term retention and the declarative knowledge typically measured in exams. That is, standard assessment measures may reflect factual knowledge in a domain, but may fail to tap into the system of relations that hold among concepts. According to Johnson-Laird (1983), mental models are functional representations or mental replicas (Craik, 1943) of phenomena in the world. Mental models are used in vision, control of movement, reasoning, and in communication. In short, mental models are used not only to understand the world, but also to operate on the world. Mental models are far more complex than associative networks, involving recursive procedures, propositions, inferences, analogies, images, and symbols, but at the core is the idea that mental models capture the functional relationships among ideas. Thus, the sensitivity of network representations for the particular problem explored here, teaching elementary mathematics problems for conceptual understanding, may lie in the power to represent the systems (or patterns) of structural relations that characterize mental models in this domain.

The value of such an explanation may depend, however, on distinguishing between the strength of an association and its propositional meaning (or link label).[2] One argument against the methodology used here is that it represents knowledge by means of numerical relatedness when ultimately teachers must convey content knowledge in the form of propositions. That is, the measures of structural similarity used here may provide insight into the system of relations among concepts, but they are silent on the particular labels participants use when making numerical ratings. Thus, a student might see two concepts as being highly related but for very different reasons than the course instructor. However, the knowledge required for teaching, or any skill for that matter, is complex and multifaceted, consisting of both implicitly and explicitly known relations (Greeno, 1983). It is possible that strength of association better reflects implicit knowledge than knowledge that is readily verbalized. Strength of association may also serve to disambiguate link labels in a way that propositions cannot. For example, the concept pairs "dog–pet" and "mouse–pet" could share the relation "is a type of," but the former would likely receive a higher numerical rating than the latter. Given the importance of distinguishing correct from incorrect propositional knowledge in the domain of teaching skill, an important extension of this work would be to investigate the nature of the conceptual relations that hold among concepts. One method, which has been used in the past (Housner et al., 1993b), is to have prospective teachers label the links in an expert's network and then compare the novices' link labels with those assigned by an expert. Such an approach has potential for providing a richer understanding of the progression of knowledge underlying teaching skill, as well as a means for diagnosing the accuracy of pedagogical content knowledge.

In summary, the findings reported here provide initial support for the use of associative network methodologies as tools for capturing changes in the pedagogical knowledge structures of prospective teachers. Associative networks appear to represent knowledge at a deeper conceptual level than is reflected in certain traditional assessment measures. In addition, the simulated teaching sessions used in this study show promise as an in-depth method of assessing pedagogical knowledge in the area of elementary school mathematics. Both of these methods appear to be superior to traditional course examinations. Perhaps both methods could be used in teacher education programs prior to the student teaching assignment as a means for determining which students may need closer observation or assistance during their practicum. These methods could also be used to assess the progression of teaching skill as experience in student teaching is obtained. This study extends past research on the structural assessment of classroom learning, not only by using teaching competence instead of course grades as the criterion variable, but also by investigating the relationship between knowledge structure and the application of knowledge in a simulated teaching task. It is important to note, however, that simulations may only approximate the authentic nature of teaching. On this view, future attempts to examine the application of mathematical knowledge by beginning teachers must be extended to schools and classrooms where the exigencies of real-world teaching are in operation.

[2] We would like to thank an anonymous reviewer for raising this issue.

References

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

Ball, D. L. (1990). Prospective elementary and secondary teachers' understanding of division. *Journal for Research in Mathematics Education, 21,* 132–144.

Ball, D., & McDiarmid, G. (1988). *"Many moons": Understanding teacher learning from a teacher education perspective* (Issue Paper 88-5). East Lansing, MI: National Center for Research on Teacher Education.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26,* 473–498.

Cobb, P., Wood, T., & Yackel, E. (1990). Classrooms as learning environments for teachers and researchers. In R. Davis (Ed.), Constructivist views on the teaching and learning of mathematics. *Journal for Research in Mathematics Education: Monograph No. 4.* Reston, VA: National Council of Teachers of Mathematics.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Cooke, N. M. (1992). Predicting judgment time from measures of psychological proximity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 640–653.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 538–549.

Cooke, N. M., & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man–Machine Studies, 29,* 407–427.

Craik, K. (1943). *The nature of explanation.* Cambridge, England: Cambridge University Press.

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science, 5,* 94–98.

Eisenhart, M., Borko, H., Underhill, R., Brown, C., Jones, D., & Agard, P. (1993). Conceptual knowledge falls through the cracks: Complexities of learning to teach mathematics for understanding. *Journal of Research in Mathematics Education, 24,* 8–40.

Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal of Research in Mathematics Education, 24,* 94–116.

Fennema, E., Frank, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using children's knowledge in instruction. *American Educational Research Journal, 30,* 555–583.

Floden, R. E., & Klinzing, H. G. (1990). What can research on teacher thinking contribute to teacher preparation? A second opinion. *Educational Researcher, 19,* 15–20.

Goldsmith, T. E., & Davenport, D. M. (1990). Assessing structural similarity of graphs. In R. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 75–88). Norwood, NJ: Ablex.

Goldsmith, T. E., Johnson, P. J., & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83,* 88–96.

Gomez, R. L., & Schvaneveldt, R. W. (1994). What is learned from artificial grammars? A transfer test of simple associative knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 396–410.

Gomez, R. L., Schvaneveldt, R. W., & Staudenmayer, H. (1996). Assessing beliefs about environmental illness/multiple chemical sensitivity. *Journal of Health Psychology, 1,* 107–123.

Gonzalvo, P., Canas, J., & Bajo, M. (1994). Structural representations in knowledge acquisition. *Journal of Educational Psychology, 86,* 601–616.

Greeno, J. G. (1983). Response to Phillips. *Educational Psychologist, 18,* 75–80.

Housner, L. D., Gomez, R. L., & Griffey, D. (1993a). A Pathfinder analysis of pedagogical knowledge structures: A follow-up investigation. *Research Quarterly for Exercise & Sport, 64,* 291–299.

Housner, L. D., Gomez, R. L., & Griffey, D. (1993b). Pedagogical knowledge in prospective teachers: Relationships to performance in a teaching methodology course. *Research Quarterly for Exercise & Sport, 64,* 167–177.

Howey, K. R., & Zimpher, N. L. (1990). Professors and deans of education. In W. R. Houston (Ed.), *Research on teacher education* (pp. 349–370). New York: Macmillan.

Johnson, P. J., Goldsmith, T. E., & Teague, K. W. (1994). Locus of the predictive advantage in Pathfinder-based representations of classroom knowledge. *Journal of Educational Psychology, 86,* 617–626.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Leinhardt, G., & Smith, D. (1985). Expertise in mathematics instruction: Subject matter knowledge. *Journal of Educational Psychology, 77,* 247–271.

Livingston, C., & Borko, H. (1989). Expert–novice differences in teaching: A cognitive analysis and implications for teacher education. *Journal of Teacher Education, 40,* 36–42.

Livingston, C., & Borko, H. (1990). High school mathematics review lessons: Expert–novice distinctions. *Journal for Research in Mathematics Education, 21,* 372–387.

Mathematical Association of America. (1991). *A call for change: Recommendations for the mathematical preparation of teachers of mathematics.* Washington, DC: Author.

Mathematical Sciences Education Board. (1991). *Counting on you: Actions supporting mathematics teaching standards.* Washington, DC: National Academy Press.

Meyer, D. E., & Schvaneveldt, R. W. (1976, April). Meaning, memory structure, and mental processes. *Science, 192,* 27–33.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

Quillian, M. R. (1969). The teachable language comprehender: A stimulation program and theory of language. *Communications of the ACM, 12,* 459–476.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1981, November). *General semantic networks.* Paper presented at the annual meeting of the Psychonomic Society, Philadelphia, PA.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249–284). New York: Academic Press.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1–22.

Tirosh, D., & Graeber, A. O. (1990). Evoking cognitive conflict to explore preservice teachers' thinking about division. *Journal for Research in Mathematics Education, 21,* 98–108.

Received June 14, 1995
Revision received December 8, 1995
Accepted December 28, 1995 ■

Acknowledgment of Ad Hoc Reviewers

The *Journal* wishes to thank persons who have reviewed manuscripts for this issue. (Numbers in parentheses indicate assistance with more than one manuscript.)

Philip Abrami
Marilyn Jager Adams
Lawrence Aleamoni
Elisha Y. Babad
Jay Belsky
James R. Booth (2)
Judith Bowey
Robert H. Bradley
Earl Butterfield
Thomas P. Carpenter
Allan S. Cohen
Lyn Corno
Peggy DeCorte
Sharon Derry
Patrick Dickson
Beverly Dretzke (2)
John Duda
John Fantuzzo
David Feldman
Martin E. Ford (2)
Mark Grabe
Ronald Johnson (2)
Daniel Keating
Timothy Z. Keith (3)
Kenneth A. Kiewra
Clifford Konold
James Kulik
Barbara G. Licht (2)
Franklin R. Manis (2)
Irene Miura
James Moser
William E. Nagy
Michelle Perry
Arthur J. Reynolds
Thomas Romberg
Hollis Scarborough
Linda Siegel
Robert E. Slavin
M. J. Snowling
Steven A. Stahl
Robert Sternberg
Michael Subkoviak
David H. Uttal
Philip Vernon
Irwin Waldman (2)
Alvin Y. Wang
Noreen M. Webb