ASSESSMENT TECHNOLOGIES FOR ADAPTIVE INSTRUCTION: DIAGNOSIS OF STAGE-SEQUENTIAL LEARNING PROGRESS IN PROBLEM-SOLVING CONTEXTS

by

MIN KYU KIM

(Under the Direction of J. Michael Spector)

ABSTRACT

Students often experience a lack of educational care suited to their particular needs and conditions. To address that issue, teachers need appropriate methods to understand what students know. This dissertation claims that dynamic formative assessments of how learners are thinking about and responding to problem situations provide a viable approach. The studies of this dissertation thus focus on theories and methodologies applicable to formative assessment. The sequence of studies involves the following: (a) defining a cognitive model of learning progress in complex problem-solving contexts; (b) devising a robust concept map technology to elicit an individual's understanding to a problem situation; and (c) developing diagnostic methodologies to assess and respond appropriately to individual cognitive changes.

The theory of mental models explains that students understand a complex problem based on their own knowledge base that is likely a structure. Drawing on the theory of mental models this dissertation suggested two theoretical frameworks involving (a) the features of knowledge structure (3S: Surface, Structure, and Semantic) and (b) the five-stage model of learning progress. Learning progress was considered as a process of transitioning from one stage to another within a student's knowledge base.

It is necessary that assessment methods take into account the complex, dynamic structure of mental models so that diagnostic, formative information becomes precise. The concept map technique was assumed to represent descriptive and complex knowledge structures in instances in which semantic relations elicited from students' natural language responses were used. A comparison study in this dissertation proved that the semantic relation approach could construct more meaningful concept maps.

The results from Confirmatory Factor Analyses (CFAs) supported that knowledge structure is likely to consist of the three features (3S: Surface, Structure, and Semantic). Latent class modeling methods were used to validate the stages of learning progress. The results did not confirm that all the stages assumed in the model exist in the data. It was argued that missing stages can be theoretically and statistically justified. In short, the proposed stage-sequential model of learning progress is able to serve as a diagnostic model of learning progress in problem-solving situations.

INDEX WORDS: Formative assessment, Learning progress, Mental models, Cognitive change, Problem solving, Concept map, Latent class model

ASSESSMENT TECHNOLOGIES FOR ADAPTIVE INSTRUCTION: DIAGNOSIS OF STAGE-SEQUENTIAL LEARNING PROGRESS IN PROBLEM-SOLVING CONTEXTS

by

MIN KYU KIM

B.S., Yonsei University, Seoul, Korea, 1997

M.A., Younsei University, Seoul, Korea, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

ASSESSMENT TECHNOLOGIES FOR ADAPTIVE INSTRUCTION: DIAGNOSIS OF STAGE-SEQUENTIAL LEARNING PROGRESS IN PROBLEM-SOLVING CONTEXTS

by

MIN KYU KIM

Major Professor: J. Michael Spector

Committee: Michael J. Hannafin
Allan Cohen

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2012

DEDICATION

To the Lord my God.

You answered my prayer and allowed me to build a new life through this five year journey.

To my wife, So Mi, and my daughter, Claire.

The only reason I could complete this work is because of your consistent love and support.

ACKNOWLEDGEMENT

I am eternally grateful to my major advisor, Dr. Spector. You accepted me as your student when I struggled to find my research agenda, and you let me follow you to the University of Georgia. You have provided wisdom and guidance ever since and inspired me to become a professional educator. My co-chair, Dr. Hannafin, has nurtured and mentored me in developing my studies. I give you my special thanks for your support for and help with my transfer to UGA and my becoming a Bulldog. Dr. Cohen, another member of my committee, has guided me in building a methodological foundation for my research. I have really enjoyed hearing your expertise in educational measurement that has pushed my own thinking.

I am also indebted to Dr. Reiser and Ms. Thomas: Dr. Reiser who was willing for me to include a case study in my dissertation that he used in his Trends and Issues in ID&T course at Florida State University; Ms. Thomas who placed my sessions in her classrooms so that I could gather data from her students. Finally, I am thankful to the students who participated in my study. Their time and willingness made my work with them valuable and meaningful. Without the support of all of you, this dissertation would not have been possible.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Background of the Studies

This dissertation is in concert with the claim that a critical and persistent issue in instructional design and technology is to provide students with learning environments suited to their individually different needs and characteristics (Cronbach, 1957; Park, 1996). However, it remains a challenge to implement personalized support on a large scale, primarily due to a lack of appropriate methods and the heavy burden placed on teachers.

To address the issue, my professional research agenda centers on designing and developing *adaptive learning environments.* Adaptive learning environments can be defined as instructional conditions that involve diagnosing a student's existing knowledge and providing tailored instructional supports to each individual (or a group having the same status) so that an individual student can progress toward the learning goals (Azevedo & Jacobson, 2008; Lee & Park, 2007; Shute & Zapata-Rivera, 2007).

To plan and implement an *adaptive* environment, the emphasis should be on the individual and individual differences, especially in terms of learner preconceptions and cognitive progression over time. Examples of adaptive learning environments are the intelligent tutoring systems (ITS) that aim at diagnosing a student's learning needs and accommodating instruction to the student's ongoing changes in learning (Lee & Park, 2007). There have been studies to advance ITS for the last two decades. Some ITS studies developed instructional tools and simulators to model human tutors and human cognition (Aleven, Popescu, & Koedinger, 2001;

Koedinger & Anderson, 1998; Seidel & Park, 1994). For example, knowledge representation methods such as semantic networks were included in some ITS to present knowledge based on students' responses to the tasks (Akhras & Self, 2002; Shute & Psotka, 1996).

Ohlsson (1987, 1993) criticized ITSs, asserting that they have limited adaptability for teaching so that ITSs are not intelligent when compared to human expert teachers. Although focusing on technologies such as AI (Artificial Intelligence), ITSs failed to employ rigorous learning principles and instructional strategies used for formative assessment and personalized instruction (Park, Perez, & Seidel, 1987). It is largely admitted that the challenge is to build a theoretical framework that accurately assesses an individual's or a group of learners' characteristics such as levels of knowledge and skills (Conati, 2002; Park & Lee, 2003; Snow, 1994).

This dissertation claims that dynamic formative assessments of how learners are thinking about and responding to problem situations provide a more viable approach. The studies of this dissertation focus on theories and methodologies applicable to formative assessment in terms of the recognition of the relevance of mental models in the development of knowledge and expertise, and technologies that now make it possible to elicit meaningful representations of mental models and assess those against productive mental models of experienced problem solvers.

More specifically, Spector (2004) framed a central question pursued in the field of instructional technology as "how to assess progress of learning in a complex domain." (p. 276). Learning progress can be defined as the changes in a learner's understanding, which are gradually modified through instruction in the direction of expert-like knowledge and performance. It is assumed in this study that, in response to a problem situation, learners experience qualitatively distinct cognitive stages representing diverse mental models. This dissertation work as an initial effort focuses on (a) defining a cognitive model of learning progress in complex problem-solving contexts, (b) devising a robust concept map technology to elicit an individual's understanding of a problem situation, and (c) developing diagnostic methodologies to assess and respond appropriately to individual cognitive changes.

Theoretical Orientation for the Studies

Theory of Mental Models

Drawing on the theory of mental models, developmental psychology, and studies of expertise development, this dissertation develops testable propositions that are applicable to the effective design of formative assessment. Mental models are defined as iconic cognitive artifacts resulting from perception and linguistic comprehension, representing certain aspects of external situations in specific domains (Johnson-Laird, 2005a, 2005b). Presumably, these cognitive artifacts are constructed by an individual based on his/her preconceptions, cognitive skills (e.g., critical thinking and meta-cognitive skills), and perceptions of the problem. The cognitive artifacts evolve and are gradually modified through experience and instruction (Carley & Palmquist, 1992; Collins & Gentner, 1987; Seel, 2001, 2003, 2004; Seel & Dinter, 1995; Shute & Zapata-Rivera, 2008; Smith, diSessa, & Roschelle, 1993).

The theory of mental models describes both knowing and teaching in terms of how knowledge is represented in the human mind, how learning evolves, and how learning progress is conceptualized in the context of instruction. When a learner compares new information with his/her existing model, the new situation can be perceived in terms of an existing structure. Otherwise, a learner can modify an entire model to fit a new experience when a learner fails to adjust (Johnson-Laird, 1983; Ifenthaler, 2010; Norman, 1983; Seel, 2003; Seel & Dinter, 1995).

The latter cognitive change results in qualitatively different mental models (Seel, 1983, 2006). Instruction for the purposes of the studies included in this dissertation can be characterized as an effort to facilitate productive cognitive changes in critical reasoning and the ability to solve ill-structured problems. Spector (2004) argued that one could view learning progress in complex problem solving as cognitive changes in the direction of expert-like mental models.

Learning progress is likely to be well-illustrated by the notion of mental models when learning progress is characterized as a set of directional changes in a learner's mental representations (Schlomske & Pirnay-Dummer, 2008; Schvaneveldt, Durso, Goldsmith, Breen, & Cooke, 1985; Spector & Koszalka, 2004). Mental models can be hypothesized as progressing through different levels of structural knowledge. Jonassen, Beissner, and Yacci (1993) proposed structural knowledge as a distinctive type of awareness separated from and intermediating between declarative and procedural knowledge. Problem solving often relies on a structural knowledge base that requires the integration of ideas and concepts (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Segars, 1997). In that sense, assessment of problem solving necessarily takes into account the organization of the knowledge base, which requires a theoretical framework of knowledge structures involved in problem solving so that different levels of problem solving can be illustrated (Gijbel, Dochy, Van den Bossche, & Segers, 2005).

One cannot see a mental model directly. Mental models are inferred entities. However, it is possible to elicit representations of these internal constructions and use them as the basis for judging how an individual's ability to reason about complex problems is developing. Mental model representations are believed to consist of propositional representations as structured symbols and images as visualized icons (Johnson-Laird, 2005b; Newell, 1990). Concept maps−structural knowledge representations consisting of concepts and relations (Clariana, 2010;

Narayanan, 2005; Novak & Cañas, 2006; Spector & Koszalka, 2004)‒are generally accepted as re-representations of a student's mental models.

In order to elicit structural knowledge, a number of technologies have been developed, including the following: DEEP (Dynamic Evaluation of Enhanced Problem-solving; see Spector & Koszalka, 2004); SMD (Surface, Matching, and Deep Structure; Ifenthaler, 2007); T-MITOCAR (Text Model Inspection Trace of Concepts and Relations; Pirnay-Dummer, & Ifenthaler, 2010); CmapTools (Novak & Cañas, 2006); jMap (Jeong, 2008); ACSMM (Analysis Constructed Shared Mental Model; O'Connor & Johnson, 2004); KU-Mapper (Clarian & Wallace, 2009); ALA-Mapper (Analysis of Lexical Aggregates-Mapper; Taricani & Clariana, 2006); ALA-Reader (Analysis of Lexical Aggregates-Reader; Clariana & Wallace, 2007, Clariana et al., 2009); and KNOT (Knowledge Network Orientation Tool; Schvaneveldt, 1990).

Developmental Psychology and Expertise Development

Developmental psychologists believe that both learning and development proceed through similar qualitative changes (Chen & Siegler, 2000; Opfer & Siegler, 2004; Siegler, Thompson, & Opfer, 2009; Werner, 1957; Vygotsky, 1934/1978). Learners may experience qualitatively different levels of knowledge structures when engaged in problem solving. Just as Piaget (1964) argued that children experience qualitatively-distinct sequential knowledge states, developmental psychologists have supported the idea that learning and development evolve as the learner constructs a qualitatively-distinct knowledge structure (Alexander, 2003, 2004; Flavell & Miller, 1998; Siegler, 2005; Siegler, Thompson, & Opfer, 2009; Werner, 1957; Vygotsky, 1934/1978). According to Siegler (2005), although some findings still seem to be related to the age-dependent changes asserted by Piaget, more findings indicate that children, even in a short-term period of learning, experience qualitative changes in mental model states (e.g., Chen & Siegler, 2000; Opfer & Siegler, 2004; Siegler et al., 2009).

It is argued in this dissertation that plausible stages applicable to the learning progress can be adopted from the studies of the development of expertise. In this view, moving toward higher levels of expertise involves "changes to different knowledge structures and complex acquired mechanisms" (Ericsson, 2003, p. 67). Considering the earlier discussion that people experience qualitatively distinct cognitive changes in the short- as well as long-term, changes in expertise development is likely to be applicable to describing the stages of learning progress. For example, current expertise studies tend to be interested in gaining expertise in domain specific learning and instruction (e.g., Alexander, 2003, 2004; Chi, 2006).

Design of the Studies Comprising this Dissertation

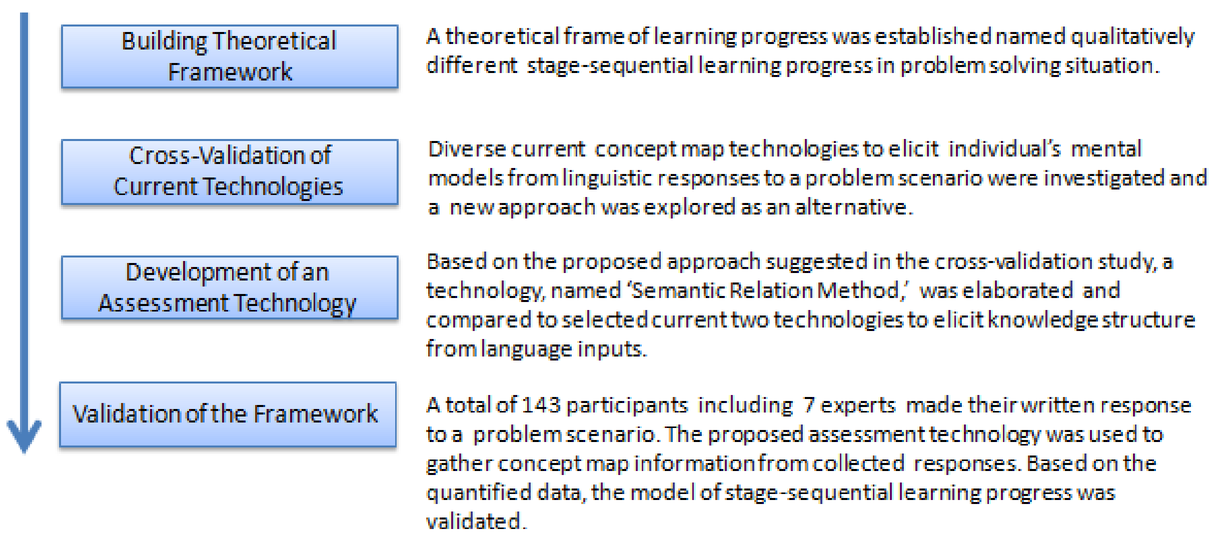


Figure 1.1. Research process.

As Figure 1.1 illustrates, although this dissertation is composed of four independent manuscripts, the four individual studies were associated with one another and conducted in a sequential order. The first study as a conceptual paper aimed at building a theoretical framework for diagnosing learning progress in problem-solving contexts. In order to evaluate the suggested framework of qualitatively distinct stages of learning progress, it was necessary to have a concept map technology able to yield descriptive and precise data. Thus, the next step was to explore and validate current concept map technologies by comparing them to an alternative approach serving as a benchmark model. The third study, based on the findings of the second, proposed a new concept map technology and elaborated specific concept mapping algorithms. The new technology provided ways to gather the concept map information used in the last study. In this final study, the theoretical framework suggested in the first was tested based on quantitative validation methods using 143 written responses to a problem scenario gathered from 136 students and seven experts.

Manuscripts

The following four papers included in this dissertation encapsulate the research conducted since August 2010. All papers were submitted to a professional journal. The first paper (chapter 2), *Theoretically grounded guidelines for assessing learning progress: Cognitive changes in problem-solving contexts*, provides a theoretical framework for learning progress in complex problem solving by drawing on the theory of mental models, the development of expertise, and the transfer of learning. The framework of learning progress consists of qualitatively distinctive sequential stages, each of which illustrates particular features of knowledge structure representing a problem situation. This paper was published in *Educational Technology Research and Development* DOI: 10.1007/s11423-012-9247-4.

The second paper (chapter 3), *Cross-validation study of methods and technologies to assess mental models in a complex problem-solving situation*, investigates current methods and technologies that yield concept maps. Concept maps consisting of concepts (nodes) and relations (links) are often used as representations of students' mental models. The focus of this paper is to identify more reliable, valid methods and technologies, or their alignments when natural language (i.e., written text) is used as raw data. This paper was published in *Computers in Human Behavior*, *28* (2): 703-717.

The third paper (chapter 4), *Development of an assessment technology for measuring knowledge structures using natural language response to a complex problem scenario*, aims at devising and validating a new concept map technology for drawing concept maps from students' written responses to a complex problem-solving situation. The proposed concept map technology is labeled as the *Semantic Relation Model*. Semantic relations of paired concepts distilled from text play a key role in creating a concept map analogous to a student's internal semantic structure. The journal to which this paper was submitted in April 2012 is *Contemporary Educational Psychology*.

The fourth paper (chapter 5), *Investigation of a model of stage-sequential learning progress in problem-solving*, validates the theoretical framework suggested in the first paper, 'a model of stage-sequential learning progress,' based on empirical data. The concept map technology proposed in the third paper is used for analyzing students' written responses to a problem scenario and eliciting their concept maps. In addition, in this study, diverse parameters describing the features of concepts maps are defined and explored; then, a set of parameters are determined as the best indicators of concept maps. Lastly, drawing on latent class modeling methods, the hypothesized model of learning progress is tested. This paper was submitted to *Contemporary Educational Psychology* in December 2011.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273-298). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Akhras, F. N., & Self, J. A. (2002). Beyond intelligent tutoring systems: situations, interactions, processes and affordances. *Instructional Science*, *30*(1), 1–30.

Aleven, V., Popescu, O., & Koedinger, K. R. (2001). Towards tutorial dialog to support self-explanation: adding natural language understanding to a cognitive tutor. In J. Moore, L. Redfield, and W. Johnson (Eds.), *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future* (pp. 246–255). Amsterdam: IOS Press.

Azevedo, R., & Jacobson, M.(2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. Educational Technology Research and Development, 56(1), 93-100.

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces, 70*, 215-225.

Chen, Z., & Siegler, R. S. (2000). Intellectual development in childhood. In R.Sternberg (Ed.) *Handbook of intelligence* (pp. 92-116). New York: Cambridge.

Chi, M.T.H. (2006). Two approaches to the study of experts' characteristics. In K.A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 121-130), Cambridge University Press.

Clariana, R. B. (2010). Multi-decision Approaches for Eliciting Knowledge Structure. In *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 41-59). Retrieved from http://dx.doi.org/10.1007/978-1-4419-5662-0_4

Clariana, R., Wallace, P., & Godshalk, V. (2009). Deriving and measuring group knowledge structure from essays: The effects of anaphoric reference. *Educational Technology Research and Development*, *57*(6), 725-737. doi:10.1007/s11423-009-9115-z

Collins, A. & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought*, (pp. 243-265). Cambridge, MA: Cambridge University Press.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence, 16*(7–8), 555–575.

Cronbach, J. L. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, *13*(5), 533-568. DOI:10.1016/S0959-4752(02)00025-7

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31-83). New York: Cambridge University Press.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon & D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 851–898). New York: Wiley.

Gijbels, D., Dochy, F., den Bossche, P. V., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27-61.

Ifenthaler, D. (2007, October). Relational, structural, and semantic analysis of graphical *representations and concept maps. Introducing the SMD-Technology*. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, Anaheim, CA.

Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research & Development, 58*(1), 81-97. doi:10.1007/s11423-008-9087-4.

Jeong, A. (2008). *jMap*. Retrieved May 5, 2010, from http://garnet.fsu.edu/~ajeong

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185-208). Cambridge University Press.

Johnson-Laird, P.N. (2005b). The history of mental models. In K. I. Manktelow, & M. C. Chung (Eds), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179-212). Psychology Press.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Koedinger, K. R., &Anderson, J. R. (1998). Illustrating principled design: the early evolution of a cognitive tutor for algebra symbolization. *Interactive Learning Environment*, 5, 161–180.

Lee, J., & Park, O. (2007). Adaptive instructional systems. In J. M. Spector, M. D. Merill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 469-484). Routledge: Taylor & Francis Group.

Narayanan, V. K. (2005). Causal mapping: An historical overview. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for research in information technology* (pp.1-19). Hershey, PA: Idea Group Publishing.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Norman, D. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7-14). Hillsdale, NJ: Erlbaum.

Novak, J. D., & Cañas, A. J. (2006). *The origins of the concept mapping tool and the continuing evolution of the tool*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.3382

O'Connor, D. L., & Johnson, T. E. (2004). *Measuring team cognition: Concept mapping elicitation as a means of constructing team shared mental models in an applied setting*. First International Conference on Concept Mapping, September 14-17, 2004, Pamplona, Spain.

Ohlsson, S. (1987). Some principles of intelligent tutoring. In R. W. Lawler & M. Yazdani (Eds.), *Artificial Intelligence and education* (pp. 203–237). Norwood, NJ: Ablex.

Ohlsson, S. (1993). Learning to do and learning to understand: a lesson and a challenge for cognitive modeling. In P. Reimann & H. Spada (Eds.), *Learning in humans and machines* (pp. 37–62*)*. Oxford: Pergamon Press.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49*, 301-332.

Park, O. and Lee, J. (2003). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 651–685). Mahwah, NJ: Lawrence Erlbaum Associates.

Park, O., Perez, R. S., & Seidel, J. R. (1987). Intelligent CAI: old wine in new bottles or a new vintage? In G. Kearsley (Ed.), *Artificial Intelligence and Instruction: Applications and Methods* (pp. 11–45). Boston, MA: Addison-Wesley.0

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.

Schlomske, N., & Pirnay-Dummer, P. (2008). Model based assessment of learning dependent change during a two semester class. In Kinshuk, Sampson, D., & Spector, M. (Eds.), *Proceedings of IADIS International Conference Cognition and Exploratory Learning in Digital Age 2008*, Freiburg, Germany, pp. 45-53.

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

Schvaneveldt, R.W., Durso, F.T., Goldsmith, T.E., Breen, T.J., & Cooke, N.M. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23,* 699-728.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation, 23*(4), 373-398.

Seel, N. M. (1983). Asking Questions and Cognitive Structuring. *Psychologie in Erziehung Und Unterricht, 30*(4), 241-252.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science, 29* (4-5), 403-427.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition, and Learning, 1* (1), 59-85.

Seel, N. M. (2004). Model-centered learning environments: Theory, Instructional design, and ethics. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instruction design: International perspectives* (pp. 49-74). NJ: Lawrence Erlbaum Associates, Inc.

Seel, Norbert M. (2006). Mental models in learning situations. In C. Held, G. M. Vosgerau, & M. Knauff, *Mental Models & the Mind, Volume 138: Current developments in Cognitive Psychology, Neuroscience and Philosophy of Mind* (1st Ed.) (pp. 85-107). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/B8G4W-4NJ2V40-7/2/88ab15869a290abbebc878cf06020bbc

Seel, N.M., & Dinter, F.R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. Educational Research and Evaluation, 1 (1), 4-35.

Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: past, present and future. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 570–600). New York: Macmillan.

Shute, V. J., & Zapata-Rivera, D. (2007). Adaptive technologies. In J. M. Spector, M. D. Merill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 227-294). Routledge: Taylor & Francis Group.

Seidel, R. J. and Park, O. (1994). An historical perspective and a model for evaluation of intelligent tutoring systems. *Journal of Educational Computing Research*, 10, 103–128.

Siegler, R. S. (2005). Children's Learning. *American Psychologist, 60*, 769-778.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*, 143-150.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg and R. K. Wagner (Eds.), *Mind in context: interactionist perspectives on human intelligence* (pp. 3–37). Cambridge, U.K.: Cambridge University Press.

Spector, J. M. (2004). Multiple uses of information and communication technology in education. In N. Seel & S. Dijkstra (Eds.), *Curriculum plans and processes of instructional design*. Mahwah, NJ: Erlbaum.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Taricani, E. M., & Clariana, R. B. (2006). A Technique for Automatically Scoring Open-Ended Concept Maps. *Educational Technology Research & Development, 54*(1), 65-82. doi:Article

Vygotsky, L. (1934/1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis, MN: University of Minnesota.

CHAPTER 2

THEORETICALLY GROUNDED GUIDELINES FOR ASSESSING LEARNING PROPGRESS: COGNITIVE CHANGES IN PROBLEM-SOLVING CONTEXTS[1]

[1] Kim, M. 2012. *Educational Technology Research and Development.* DOI: 10.1007/s11423-012-9247-4 Reprinted here with permission of publisher.

Abstract

It is generally accepted that the cognitive development for a wide range of students can be improved through adaptive instruction - learning environments optimized to suit individual needs (e.g., Cronbach, 1957; Lee & Park, 2007; Shute & Zapata-Rivera, 2007). Designers of adaptive instruction, however, have been slow in responding to the needs of individual learners in complex problem solving situations. It is vital to diagnose an individual’s learning progress in terms of cognitive changes in complex problem-solving contexts. This paper reports the development of theoretically grounded guidelines to advance the research and implementation of an assessment of learning progress in complex problem solving domains. Drawing on the theory of mental models, the development of expertise, and the transfer of learning, I present a framework of learning progress applicable to the process of developing and implementing adaptive instruction that involves detecting and monitoring changes in mental models and then generating formative support, providing feedback, and assessing progress.

*Keywords***:** learning progress, mental models, cognitive change, problem solving

The purpose of this paper is to develop a theoretically grounded framework to guide and advance the assessment of learning progress. The discussion centers on instructional challenges related to cognitive changes in problem solving. The sequence of the challenges is compatible with the process of adaptive instruction that includes (a) assessing learner states (e.g., preconceptions), (b) monitoring learning progress, and (c) promoting transfer of learning to a larger group of problem situations. The focus is on promoting the development of expertise in solving complex and ill-structured problems (such as those that occur in many domains including engineering design, environmental planning, etc.) as it is already well established how to design instruction to support learning simple problem solving procedures. Drawing on the theory of mental models, studies of expertise development, and developmental psychology, testable propositions are developed that are applicable to the effective design of adaptive instruction.

Learning necessarily involves change (Siegler, 2005; Spector, 2009). Instruction is an intentional effort to facilitate, verify, and elaborate desirable changes in the short-term and long-term. Instruction can be characterized as an effort that includes assessing a person's existing knowledge, determining appropriate learning methods and activities, and facilitating an individual learner's progression toward an explicit learning goal, which is the desired change stated in general terms (Lee & Park, 2007; Shute & Zapata-Rivera, 2007).

However, effecting cognitive change is not a simple task because learning is a complex process affected by many factors. For example, science competency goes beyond simply having specific factual knowledge and includes conceptual understanding of complex scientific phenomena, the capacity to engage in critical reasoning and the ability to solve ill-structured problems that typically have incomplete information at the outset and that lend themselves to multiple solutions and solution approaches (NRC, 2005). It is now widely acknowledged that students should have more opportunities to be educated with authentic and complex problems in order to better develop their problem solving abilities (Herrington & Oliver, 2000; Jonassen, 2000; Jonassen & Rohrer-Murphy, 1999; Spector, 2008b).

Studies in the development of expertise focus on identifying, explaining, and utilizing expertise, including investigating the ways that people develop expertise in specific domains (Ericsson, 2006; Spector, 2008a, Sternberg, 1999). Spector (2004) argued that one could view learning progress in complex problem solving as changes in the direction of expert-like knowledge and performance. In an instructional situation, one may experience a couple of challenges with regard to assessing and monitoring learning progress and promoting improvements in problem solving.

First, instruction that aims to develop a student’s expertise in problem-solving situations probably begins with the assessment of the student’s level of understanding and the identification of areas for improvement (van Gog, Ericsson, Rikers, & Paas, 2005). Accordingly, a method to elicit a learner’s problem conceptualization is required and determination of a learner’s progress necessitates having a reference model (Carver, 2006; Ifenthaler & Seel, 2005; Schlomske & Pirnay-Dummer, 2008; Spector, 2008a). However, in a complex problem-solving situation, it is difficult to establish standards or reference models.

Second, a theoretically sound and systematic assessment model is required to determine levels of expertise, to explain learning progress, and to provide adaptive instruction that meets individual requirements in terms of differences in developmental stages of expertise, established problem space and meta-cognition (Spector, 2008a).

Third, based on an assessment model likely composed of qualitatively distinct stages, the methodology to monitor transitions among different stages of learning progress is vital when deciding a learner's developmental position. In addition, it is anticipated that there is often a reversion to an earlier or less sophisticated stage of learning progress when confronted with a particularly challenging problem. These cognitive reversions to earlier stages of development need to be theoretically established (e.g., when they are likely to occur, to which prior stages a person is likely to revert, why reversion occurs, etc.) and then identified in particular cases so that appropriate interventions and support can be deployed.

Fourth, the transfer of expertise from one problem situation to another, be it similar or dissimilar, is a critical goal of problem-solving instruction. Specifically, it is important to determine whether learners achieve particular levels of problem-solving knowledge and skills and are able to adapt their knowledge and problem-solving skills to new situations.

In this paper, the theoretical foundations involved in learning progress and the development of problem solving knowledge and skills are explored so as to provide a framework for the assessment of learning progress in complex problem solving. Emphasis is placed on mental model theory as a promising basis for such an instructional framework.

Assessment of Learner Understanding

Re-represented Mental Models as Learner Models

The theory of mental models describes both knowing and teaching in terms of how knowledge is represented in the human mind, how learning evolves and how learning progress is conceptualized in the context of instruction. Mental models are defined as iconic cognitive artifacts resulting from perception and linguistic comprehension, representing certain aspects of external situations in specific domains (Johnson-Laird, 2005a, 2005b).

These cognitive artifacts are presumably constructed by an individual based on preconceptions, his/her cognitive skills (e.g., critical thinking, meta-cognitive skills), and a perception of the problem itself. The cognitive artifacts evolve and are gradually modified through experience and instruction (Carley & Palmquist, 1992; Collins & Gentner, 1987; Seel, 2001, 2003, 2004; Seel & Dinter, 1995; Shute & Zapata-Rivera, 2008; Smith, diSessa, & Roschelle, 1993). The theory of mental models provides a comprehensive perspective on knowledge building that includes both domain-general processes (i.e., those that transcend and contribute to cognitive development across domains; Kail, 2004; Sternberg, 2008) and domain-specific knowledge (Chi, Glaser, & Farr, 1988; Vosniadou, Vamvakoussi, & Skopeliti, 2008). Fundamentally, mental models are considered to be domain-specific in that they mirror the structure of what they represent (Johnson-Laird, 2005b; Seel, 2001; Spector, 2008a). However, the process of model building is largely influenced by domain-general processes such as meta-cognition, self-regulation, and cognitive flexibility (Collins & Gentner, 1987; Kail, 2004; Sternberg, 2008). According to cognitive psychologists, problem solving involves finding a reasonable course of action, often by making use of mental models (Johnson-Laird, 2005a; Seel, 2001, 2004). Manipulation of mental models is accompanied or led by meta-cognition (e.g., reflection). For example, a law school student reading the testimony of an eyewitness in a criminal case might first construct an internal representation of the situation and then test that representation against other facts and testimony. The internal representations are typically quite specific to the situation. Afterwards, the student may reflect on the situation and reasoning about that testimony, deliberate on the implications and assumptions and possible alternatives, and then modify the representation to explain to him/ herself what probably happened. The interim deliberative process involves meta-cognitive processes that may lead to another domain-specific internal representation.

Second, the theory of mental models explains the process of building expertise. The process of building expertise can be explained as a way of constructing mental models. According to Piaget (1964), there are two different cognitive processes involved in mental model development − assimilation and accommodation. When encountering a new situation, a learner activates existing models analogous to the situation. By comparing the new information with the model, the new situation can be perceived in terms of the existing structure (assimilation). Accommodation involves the modification of an entire model to fit a new experience when a learner fails to adjust (Johnson-Laird, 1983; Ifenthaler, 2010; Norman, 1983; Seel, 2003; Seel & Dinter, 1995).

One way to promote cognitive change is to provide appropriate cognitive conflict. This approach is grounded on a few well-known learning theories such as Piaget's (1976) process of equilibration, Festinger's (1962) notion of cognitive dissonance, and Vygotsky's (1934/1978) zone of proximal development. Vygotsky (1934/1978) said that "the only good learning is that which is in advance of development" (p. 82).

Third, the theory of mental models provides a measurable feature of existing knowledge. At first it might appear that mental models could not be the foundation for any kind of instructional planning or implementation since they are permanently hidden from view. That is to say that one cannot see a mental model directly. Mental models are inferred entities. However, it is possible to elicit representations of these internal constructions and use those as the basis of judging how an individual's ability to reason about complex problems is developing. Mental model representations are believed to consist of propositional representations as structured symbols and images as visualized icons (Johnson-Laird, 2005b; Newell, 1990). A semantic network as the externally represented structural components of mental models implies that a latent structure exists in the human brain. According to this theoretical perspective, detecting those latent structures and changes as they evolve is a way to assess the development of complex problem solving skills.

Measurable Features of Mental Models as Structural Knowledge

Learning progress is likely to be well-illustrated by the notion of mental models when learning progress is characterized as a set of directional changes in a learner's mental representations. Many empirical studies utilizing concept map techniques have shown that students' structural comprehension in a domain becomes more coherent and expert-like as they gain competence in a discipline (e.g., Schlomske & Pirnay-Dummer, 2008; Schvaneveldt, Durso, Goldsmith, Breen, & Cooke, 1985; Spector & Koszalka, 2004).

Mental models can be hypothesized as progressing through different levels of structural knowledge. From this standpoint, knowledge appears to be a configuration of mental representations as a whole that contain symbols and their relationships corresponding to the properties of that which they represent (Johnson-Laird, 1983).

Problem solving often relies on a structural knowledge base that requires the integration of ideas and concepts (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Segars, 1997). In that sense, assessment of problem solving necessarily takes into account the organization of the knowledge base, which requires a theoretical framework of knowledge structures involved in problem solving so that different levels of problem solving can be illustrated (Gijbel, Dochy, Van den Bossche, & Segers, 2005).

Table 2.1

Constructs Related to Knowledge Structure

Features of Knowledge Structure (3S) (Ifenthaler, 2006, 2010; Pirnay-Dummer, 2006; Spector & Koszalka, 2004)	Elements of Knowledge Structure (Sugrue, 1995)	Analogy Study (Gentner & Medina, 1998; Holyoak & Koh, 1987; Judd, 1908; Simon & Hayes, 1976)	Linguistic Comprehension (Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1972; Katz & Postal, 1964; Kintsch & van Dijk, 1978)
Surface feature	Concepts	Surface	Surface structure
Structural feature	Links from concepts and principles to conditions and procedures for application	Deep	Deep structure
Semantic feature	Principles		

Spector and Koszalka (2004) first introduced three features (3S) of knowledge structure that likely describe mental models: (a) surface, (b) structure, and (c) semantic features (see Table 2.1). Those features have been used as a framework for developing assessment measures for mental models (Ifenthaler, 2006; Pirnary-Dummer, 2006). The 3S features of knowledge structure were confirmed in this study, supported by studies in related areas: elements of knowledge structure, analogy study, and linguistic comprehension.

First, the surface feature indicates the descriptive information about the components of a knowledge structure. According to Sugrue's (1995) elements of a knowledge structure targeted by the assessment of problem solving, the surface feature relates to an understanding of concepts defined as "a category of objects, events, people, symbols or ideas that share common defining attributes or properties and are identified by the same name" (Sugrue, 1995, p. 9). Likewise, cognitive scientists account for the surface level of mental models as salient objects and aspects of the context (Holyoak & Koh, 1987; Simon & Hayes, 1976), and according to linguists studying linguistic comprehension, the surface structure of linguistic representations characterize the shape of sentences in terms of concepts and their relations in a text (Katz & Postal, 1964).

The second indicator of a knowledge structure is a structural feature that describes the levels of size, complexity, and cohesiveness of a mental model. In this feature, concepts and principles are situated in particular conditions and procedures for application (Sugrue, 1995). The focus in assessing problem solving ability is on the “extent to which the student’s knowledge structure is organized around key concepts and principles that are linked to conditions and procedures for application” (Gijbels et al., p. 35). That is, the structural feature indicates a deep level in terms of a well-organized knowledge structure within a particular context in which underlying causal principles, including key variables and their relations, are subsumed (Gentner & Medina, 1998; Judd, 1908).

As the third indicator, the semantic feature shows the levels of understanding concepts and their relations in a knowledge structure, including principles that can be “a rule, law, formula, or if-then statement that characterizes the relationship between two or more concepts” (p. 9). Katz and Postal (1964) claimed that a substantial part of the meaning emerges from the semantic information of the deep structure. This idea is supported by studies of linguistic comprehension that argue that meaning stems from information integrated in the whole corpus (Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1972; Kintsch & van Dijk, 1978).

Problems in Building Reference Models of Complex Problems

Problem solving is a goal-directed cognitive effort and requires appropriate and adequate understanding of the problem (Anderson, 1980). Diagnosis of the levels of problem-solving knowledge and skills involves comparison between mental models representing a problem-solver's existing understanding in the form of learner models and a targeted expert's model as a reference model. Likewise, learning progress involves a series of changes toward goals. For example, Snow (1990) claimed that the progress of mental models involves learning-dependent and developmental transitions between preconceptions and causal explanations. Both problem-solving and learning require reference models to identify, monitor and promote levels of expertise in problem-solving situations. Particularly for instructional contexts, reference models may denote instructional goals, which have the critical functions of assessing individual learning status and providing adaptive feedback in problem-solving tasks.

A possible argument against the need for a reference model is that people create their own understanding; it is not plausible or reasonable to build reference models in complex and ill-structured problem situations because even experts' models will vary depending on their experience. This study concedes that an understanding of a problem situation is ultimately achieved when an individual internally constructs his/ her own representations of a problem. It is also argued here that even personal representations are social artifacts partially or substantially constructed through the interactions of a group; this position is called social constructivism (Glasersfeld & von Ernst, 1995). For example, as Wittgenstein (1922) claimed, language as a social enterprise plays a critical role in building and mediating an individual's internal representations of the external world. As a consequence, interacting with others via externalized representations, such as verbalized language expressions, allows experienced people to come to a shared understanding, at least in some problem cases.

Theoretical discussions of problem-solving can provide a few plausible or pragmatic ways to reconcile the need for reference models with the impossibility of building references in complex and ill-structured problem situations. Pretz and colleagues (2003) defined problem solving as a set of mental activities composed of: (a) recognizing the problem, (b) defining and representing the problem, (c) developing a solution strategy, (d) organizing one's knowledge about the problem, (e) allocating mental resources for solving the problem, (f) monitoring one's progress toward the goal, and (h) evaluating the solution. These problem-solving processes can be dichotomized into two phases: planning (which includes (a) and (b)) and development (which consists of (c) through (h)).

Based on this theoretical background, it does seem possible to build reference models at least in the planning phase of problem solving. While solutions to ill-structured complex problems may be multiple and constructed through diverse paths, the planning phase can be somewhat invariant. The planning phase of problem solving is believed to end with the problem representation, which refers to "the manner in which the information known about a problem is mentally organized" (Pretz et al., 2003, p. 6). Newell and Simon (1972) claimed that a problem solver conceptualizes the problem space in which all of the possible conditions of a problem exist. Studies of expertise demonstrate a clear distinction among mentally represented problem spaces between experts and novices (e.g., Chi, Glaser, & Farr, 1988; Spector & Koszalka, 2004).

Spector (2008a) argued that "experts would exhibit clearly recognizable patterns in their problem conceptualization, although experts did develop a variety of problem responses" (p. 31). That is, although levels of problem representations vary depending on an individual's level of expertise, experts in the same discipline typically have a relatively similar understanding of the problem space.

The problem space to a problem situation can be represented as a single canonical knowledge structure. Even more often, there are admittedly multiple acceptable representations depending on the problem situation. In an instructional setting, a single or multiple reference models can be established to facilitate learners' problem-solving development. For example, drawing on measurement methodologies to compare knowledge representations (Ifenthaler, 2006; Pirnay-Dummer, 2006; Spector & Koszalka, 2004), a variety of knowledge representations elicited from multiple experts can be compared so that a set of common entities to a problem are generated. In another way, a student representation can be analyzed against a variety of reference models so that his/ her understanding of a target knowledge structure can be assessed by multiple standards, providing diverse aspects of a problem situation.

Stage-Sequential Learning Progress

Insights from Developmental Psychology

Behavioral learning theorists believed that human understanding grows continuously, without qualitatively distinct cognitive stages (e.g., Kendler & Kendler, 1962). In contrast, Piaget (1964) believed that children experience qualitatively distinct, sequential knowledge states (e.g., sensorimotor, preoperational, concrete operational and formal operational) while growing up. Although his notion of qualitative changes in a child's thinking has continued in contemporary developmental theories (Flavell & Miller, 1998), there are several departures from Piaget's conventional notion of developmental stages.

First, cognitive development of a child is probably more learning-dependent and less age-dependent than Piaget's early studies suggest. Piaget's (1964) main claim was that a child's development is an age-dependent progress. He suggested that a child of a young age is not likely to perform tasks requiring higher level cognitive structure because his/her cognitive ability is still immature (Siegler & Klarh, 1982). In contrast to Piaget, Bruner (1961) agreed with the qualitative stages of cognitive development, but not with the age-dependent progress. Findings in developmental studies support Bruner's points. For example, many observations have discovered that some infants show precocious abilities (Flavell, 1992). In addition, unless adults have adequate and appropriate education or training, they may not be able to resolve Piaget's formal operational tasks (Shaklee, 1979).

Second, cognitive development is promoted by interactions between domain-general and domain-specific processes. There has been a longstanding debate whether development proceeds through a fixed sequence of stages across learning domains or through a specific fractionated manner within a particular subject area (Carlson, 2002; Case, 1992; Fischer & Silvern, 1985; Flavell, 1985). Piaget's (1964) theory postulates static, invariant developmental stages, which are domain-independent. Contrarily, contemporary studies suggest that the development of the human mind is neither explained by totally general, domain-independent stages nor by only domain-specific, fractionated knowledge (Flavell, 1992; Sternberg, 2008). Accordingly, it is important to note that a general stage-like development can still account for changes of understanding. In contrast, domain-specific knowledge appears to be more critical to building expertise.

Third, both learning and development proceed through similar qualitative changes. Piaget (1964) believed that learning and development are fundamentally dissimilar. Stage-sequential qualitative stages only occur in development (long-term change). Learning is viewed as short-term changes aimed at obtaining and accumulating domain-specific content knowledge (Siegler, 2005; Siegler, Thompson, & Opfer, 2009). In contrast, Werner (1957) and Vygotsky (1934/ 1978) believed that both learning and development basically evolve as the learner constructs qualitatively distinct knowledge structures. According to Siegler (2005), although some findings still seem to be related to the age-dependent changes asserted by Piaget, more findings indicate that children, even in a short-term period of learning, experience qualitative changes in mental model states (e.g., Chen & Siegler, 2000; Opfer & Siegler, 2004; Siegler et al., 2009). It is notable that children proceed through a similar progression of qualitatively distinct stages in both the short term and the long term, which may occur based on quantitative changes in the frequency of existing approaches (Siegler et al., 2009; Vosniadou et al., 2008).

Stages When Expertise Develops

This study argues that plausible stages applicable to the learning progress can be adopted from the studies of the development of expertise. Ericsson (2003, 2005, 2006) suggested that expertise is developed by a deliberate practice in which learners engage in appropriate and challenging tasks carefully picked by masters, devote years of practice to improving their performance, and refine their cognitive mechanisms by self-monitoring efforts such as planning, reflection, and evaluation. In this view, moving toward higher levels of expertise involves “changes to different knowledge structures and complex acquired mechanisms” (Ericsson, 2003, p. 67). Ericsson and Simon (1980, 1993) used the think-aloud protocol method to understand how a highly superior performer is thinking in a given problem situation and explained that there might be qualitative changes in mental models when expertise develops. Nonetheless, Ericsson and colleagues have not yielded a model in which the developmental stages of expertise explicitly exist.

Considering the earlier discussion that people experience qualitatively distinct cognitive changes in the short- as well as long-term, changes of expertise development is likely to be applicable to describing the stages of learning progress. For example, current expertise studies tend to be interested in gaining expertise in domain specific learning and instruction (e.g., Alexander, 2003, 2004; Chi, 2006).

A seminal work providing acceptable stages of the development of expertise is the scale Dreyfus and colleagues (1986) proposed (see Table 2-2). They suggested that expertise develops through five stages as a novice experiences a variety of situations. The novice is a beginner in a domain and learns detached rules and facts with little benefit of experience. When experiencing real situations or observing a number of examples, the learner becomes able to recognize situational knowledge as well as context-free features (the advanced learner). At this point, he or she still has difficulty in discriminating what is more important in any particular situation. The next stage is the competent learner who determines critical elements of the situation and then restricts their attention to the selected relevant features. With a sufficient number of experiences and competent perspective, a proficient learner has holistic understanding on a discipline. He or she immediately recognizes the important aspects of the current situation but still needs to deliberate about what to do. The most distinctive trait of the expert is her or his immediate response to a situation because she or he intuitively knows how to achieve the goal due to the vast amount of experienced situations classified in sub-classes.

Table 2.2

Stages in the Development of Expertise

Level	Dreyfus, Dreyfus, & Athanasiou (1986)	Chi (2006), Hoffman (1998)	Alexander (2003, 2004)
1	Novice	Novice	Acclimation
2	Advanced Beginner	Initiate	Competence
3	Competence	Apprentice	Proficiency-Expertise
4	Proficiency	Journeyman	
5	Expert	Expert	
6		Master	

Many other models have been introduced since Dreyfus and colleagues’ work (see Table 2.2). For example, Chi (2006) developed a proficiency scale of expertise development, which was adapted from Hoffman (1998). She believed that the ultimate goal of studying for relative expertise is to enable less-skilled novices to become more knowledgeable by providing adaptive instruction based on their levels of understanding. The proficiency scale of expertise development is composed of six stages from novice to master as follows: (a) Novice (someone who is new), (b) Initiate (a novice who has begun introductory instruction), (c) Apprentice (one who is learning), (d) Journeyman (a person who can perform a task under guided orders), (e) Expert (one who is distinguished or a brilliant journeyman), and (f) Master (any journeyman or expert who is qualified to teach someone at a lower-level) (p. 22). These scales denote different levels of personal expertise in terms of roles and abilities that a learner can experience from learning and instruction.

Alexander (2003, 2004) introduced an account of multiple stages of expertise development focusing on the nature of developing expertise in academic domains. The developmental phases are hypothesized as having three stages: acclimation, competence, and proficiency-expertise. The acclimation stage refers to the initial level in which learners become familiar with the given unfamiliar domain. Accordingly, it is typical in this stage for learners to only have limited and fragmented knowledge that is not cohesive or interconnected. In contrast, learners experience quantitative and qualitative changes in the competence stage. They construct sufficient, well-organized, domain-specific knowledge. An increased familiarity with the given domain leads learners to further delve into the domain. In the proficiency-expert stage, learners feed on new knowledge of the domain and markedly improve in higher levels of domain knowledge. Alexander's model of domain learning provides an insight into the features of knowledge structure that a learner constructs at each stage.

A Framework of Stage-Sequential Learning Progress

Based on the aforementioned theoretical review of mental models, expertise development, and developmental psychology, a model of stage-sequential learning progress as a potential framework for learning progress is presented below (see Table 2.3). This model draws on the five stages of expertise development (Dreyfus et al., 1986) and three types of knowledge features (i.e., surface, structure, and semantic) (see Table 2.1), associating the two approaches according to the notion that knowledge appears to be a configuration of mental.

Table 2.3

A Framework of Qualitatively Distinct Stages of Learning Progress

Stage	Description
Novice (Irrelevant structure)	A beginner starts to learn context-free abstract knowledge and faces situations in a new domain; concepts and relations are low in quantity compared to the reference model.
Advanced beginner (Surface structure)	A learner recognizes situational knowledge as well as non-situational knowledge but shows a lack of sense of what is important in a particular situation.
Competent Learner (Deep structure)	With increasing experience, a learner chooses a perspective and then determines which elements of a situation are critical; most key concepts are posed in a learner's mental model, but the propositional relations among concepts are somewhat different from the reference model.
Proficient Learner (Semantically deep structure)	A learner approaches a problem holistically and immediately recognizes problem space; the expected concepts and relations among them are represented in a learner’s mental model including situational and non-situational concepts.
Intuitive Expert (Advanced semantic structure)	An expert intuitively makes a decision about what the problem is and how it is resolved. An expert has tacit knowledge based on a vast number of relevant experiences.

Although the scale proposed by Dreyfus and Dreyfus (1986) anticipates many other models, the scale appears to be more informative in accounting for the qualitative features of mental models at each stage and the progress of problem-solving skills. The five levels of expertise are used to denote a learner’s stages of learning progress in a problem-solving situation. Associating the features of knowledge structure and the development of expertise characterizes each stage of learning progress that a learner's mental model falls into on a given occasion. A series of changes in the stages show a learner's learning trajectory, or learning progress.

Novice learner (Irrelevant structure). Novices are new to the domain, so they have difficulty linking their prior-knowledge to the new domain, thus rendering them as yet unable to represent new knowledge. This level represents a stage at which learners lack knowledge on both the contextual and principle levels, resulting in a poor representation of their thoughts and ideas. Two types of mental models are assumed at this stage: (a) all features (surface, structure, and semantic) of novices' knowledge structures are quite dissimilar to those of experts, or (b) the structure feature could be understood as mastered because mental models consisting of a small number of concepts and relations are likely to look cohesive and connected. This idea is in accord with the claim that a structural graphical approach is insufficient for interpreting mental models (Forbus, Gentner, Markman, & Ferguson, 1998; Kubose, Holyoak, & Hummel, 2002).

Advanced beginner (Surface structure). The advanced beginner stage represents a mental structure in which an adequate amount of contextual knowledge is recognized, but his or her knowledge lacks the building of a proper knowledge structure associated with principles. Otherwise, some instructed principles (key concepts/abstract knowledge) might exist but are not properly connected with one another. Accordingly, two types of mental structure are assumed: (a) mental modes in this stage have similar surface features with those of a reference model but not with structure and semantic features, or (b) there is a high similarity of semantic features but dissimilarity in surface and structure features between a student model and a reference model.

Competent learner (Deep structure). A competent learner comes to identify key concepts underpinning a situation and thus to organize the surface features. However, his or her mental structure still has missing relations among key variables. In other words, learners create either complex knowledge structures or an appropriate set of principles accompanied by a significant number of concepts and relations. There are likely two types of deep structure: (a) one has an adequate structural complexity along with a proper surface fit, which is not necessary to guarantee a semantic fit, however, and (b) the other consists of an appropriate number of contextual and principle concepts that are not yet well-structured.

Proficient learner (Semantically deep structure). The next stage is the proficient learner stage, in which learners conceptualize a sufficient problem space. This study regards experts as persons at the level of proficient performance. Mental models in this stage are assumed to be well-featured at all levels (surface, structure, and semantic). In addition, there is possibly another type of knowledge structure in which a significant number of principles creates a cohesive structure but with a small total number of concepts (surface). This model is supported by the claim that experts sometimes create mental models having an 'optimal' rather than 'maximum' number of concepts and relations that are very efficient (Glaser, Abelson, & Garrison, 1983; Glaser, 1992).

Intuitive Expert (Advanced semantic structure). Dreyfus and Dreyfus (1986) regard the level of intuitive expert as somewhat mysterious ─neither well understood nor clearly supported. At this level, intuitive decision-making takes place based on an advanced semantic structure unlikely to theorize its measurable structure at this point (Dreyfus & Dreyfus, 2005). The difference between logical thinking (the proficient learner) and intuitive decision-making (the expert learner) toward problem-solving is not easily discerned by investigating a single set, or a few sets, of mental observations, such as concept models, because the measurable features of intuitive experts' mental models are still not well understood. Admittedly, experts might not be fostered solely by instruction in a domain, for experiences both before and after instruction could be more influential in the formation of an expert. For example, Ericsson (2003, 2005, 2006) claimed that becoming an expert in a domain on the whole takes 10 years of devotion to highly disciplined, focused, and reflective practice.

Monitoring Learning Progress

Detecting Stage Transitions

Changes in the stages of learning progress can provide information about the effects of an instructional intervention as well as about an individual's state of learning. Any instructional intervention in problem-solving learning may, to some extent, direct learners to improve their knowledge and skills. That is, interventions affect mental model changes. A learning environment in which teachers have diagnostic information about their students and provide formative feedback catered to each individual's needs might be an ideal setting. No matter what intervention is designed to improve problem-solving knowledge and skills, it is important to see the learner's progress in a longitudinal manner so that one can evaluate the effects of an instructional intervention. The proposed framework hypothesizing stage-sequential learning progress provides a diagnostic model; further validation and research using this model is of course required.

Possible Regression of Learning Progress

We can anticipate that most students proceed through positive learning progress from lower stages towards higher stages when they have appropriate instructional support, such as individualized feedback. In contrast, a couple of different patterns can appear as follows:

Expertise reversal effect. The expertise reversal effect actually denotes that integrated information (e.g., text with diagram), which is designed to be beneficial for learning, may provide redundant information to expert learners hindering their ability to learn new knowledge (Kalyuga, Chandler, & Sweller, 1998; van Gog, Ericsson, Rikers, & Paas, 2005). In this paper, the expertise reversal effect is used as a general term, indicating unexpected effect of increasing expertise in learning and instruction. For example, in the context of measurement of mental models, the expertise reversal effect may cause an assessment of mental models to be dysfunctional. Suppose that a reference model for a physics problem is developed only considering the classical physics theory (i.e., Newtonian mechanics). This is done based on the assumption that the target students are too immature to deal with more advanced theories, and the assessment will generate incorrect diagnostic information in the case of an advanced learner's response based on the theory of relativity or quantum theory.

Reversion to an earlier stage. Reversion to an earlier stage may be possible. For example, according to Vosniadou and Skopeliti (2005), children who believe the earth is a physical object being rectangular take slow and gradual process to obtain scientific understanding, that is, the earth as a solar object being sphere. It is possible that some students fail to accomplish a conceptual change toward a scientific model of earth and then revert to their naïve belief established from their everyday experiences reinforcing their model that the ground is flat and below and, in contrast, the sky and solar objects are above.

Reversion by expertise discordance. We can assume another reversion affected by conflicting expertise even without degraded performance. For example, when a student in a South Korean high school whose family has defected from North Korea is solving an economic problem, the student may progress to some degree in conceptualizing a problem space based on capitalistic economics. However, at a certain point, the student may return to his/ her former expertise about Marxist economics to understand economic problems. Therefore, the student's response may appear backward. In that case, reverting to an earlier stage is not necessarily the same as degraded performance.

Stay at a stage. In some cases we may see no movement among stages. Although learners are given instructions and feedback, they may stagnate without more progress in problem solving. A couple of situations can be envisioned. Problem solving in a learning context is a goal-oriented activity that progresses toward learner acquisition of a reference model. These goals are external. A student may decide to ignore goals or reject the given feedback on his or her prior performance (e.g., prior problem conceptualization) (Sadler, 1989). In addition, low performing, self-efficacious students may experience a negative affect such as decreased motivation when they continuously receive negative feedback due to a lack of proper concepts in their understanding (Kernis, Broker, & Frankel, 1989). In these cases, progressing forward through stages may not take place.

Promoting the Transfer of Problem-solving Knowledge and Skills

Learning Transfer as a Learning Progress

The framework of stage-sequential learning progress ends with the expert stage. The expert stage is characterized as not only obtaining expertise in a particular domain, but also including the ability to transfer what an expert knows in one domain to solve problems in other contexts (Gentner, Loewenstein, & Thompson, 2003).

Although it is a common belief that people adapt their prior knowledge to solve new problems, the studies of transfer have taken diverse perspectives which seem to locate in a continuum with two polarized ends (e.g., Barnett & Ceci, 2002; Bransford & Schwartz, 1999;

Gentner, Holyoak, & Kokinov, 2001; Gentner et al., 2003; Lave, 1988; Lobato, 2006; Reeves & Weissberg, 1994; Salomon & Perkins, 1989; Singley & Anderson, 1989). On one end, there is the notion of *identical elements* that are shared components between original learning and transfer situations. The extent of shared features determines the occurrence of transfer (That is the foundation of classical perspective; Thorndike, 1906; Cox, 1997). On the other end, according to situated cognition standpoint, Lave (1988) asserted that knowledge cannot be detached from a specific situation in which it is acquired, so that situation-specific knowledge cannot transfer. The framework of the learning progress associated with the theory of mental models possibly reconciles this classical view with alternative approaches.

Mental models for new situations may be created through an analogical thinking process which consists of selecting analogous models, comparing the original models with new situations, and modifying and simulating current mental models (Johnson-Laird, 1983; Norman, 1983; Seel, 2003). In that sense, the theory of mental models admits the notion of analogical transfer (Bransford, Brown, & Cocking, 2000; Gentner et al., 2001; Sternberg & Frensch, 1993).

In addition, mental models are domain-specific and situation-sensitive (Garnham, 1987, 2001). That is, key principles in mental models can be best interpreted in the relation with surface features. In that sense, when transfer occurs, de-contextualized abstract knowledge might not be merely transferred. Rather, the original mental models as a whole might affect a new model in a new context. This view cannot only admit the role of key principles in transfer, but can also be reconciled with critique such as how the classical transfer approach separates knowledge from the situations (Lobato, 2006).

Supposing that assessment technologies, which can detect learner progress by embedding a tool visualizing mental models, are invented based on the suggested framework, it is possible to investigate transfer with more accurate and rapid measures frequently used in a longitudinal study of cognitive change (e.g., Lemaire & Siegler, 1995). For example, a technique visualizing mental models shows the structural accuracy of learners' understanding (accuracy) and diagnosing learning stages in multiple occasions enable researchers to determine when learners reach the expected stage (speed).

A Scenario

A set of problem situations is carefully selected and developed to teach ecological relationships in a biology class. These complex problems have differing contexts but encompass the same key variables (i.e., population, nutrients, and predators). It is assumed that the problem scenarios have their own reference models and are the same in their levels of difficulty and complexity. For each problem situation, learners encounter multiple tests to measure their problem conceptualization. The problems are entitled as follows:

- *Crown-of-thorns starfish*
- *Habu (a viper) and mongoose in Okinawa*
- *Sharp drop of deer population in Georgia*
- *Disappearing iguana in the Galapagos Islands*
- *The attack of bass at the lakes in South Korea*

In problem-solving situations, acquired knowledge and problem-solving skills are expected to be transferred from one situation to the next. The expectation is that learners build their mental models representing the three key variables (i.e., population, nutrients, and predators) and their relationships across problem contexts.

The generalizability of specific problem-solving abilities can be anticipated based upon the continuous success in obtaining reference models of a set of similar problem situations and the speed of reaching the highest stage of learning progress. For example, in the given experimental scenario, most learners will likely have more difficulty in, and spend more time, solving the first problem. Their stage transition may proceed systematically. In the second problem, although they progress without skipping stages, they may reach the target stage faster. In the third and fourth scenarios, some students may present a first response that begins at a higher point and skips the lower stages.

Finally, proficient students may solve the last problem very quickly. The series of examinations provide meaningful evidence: as students successively build a target-understanding of diverse problem situations, the speed to reach the target stage increases through solving subsequent problems. Based on this evidence, one can conclude that students acquire the expected problem-solving knowledge and skills that are transferred across the given problems. Namely, the application of particular problem-solving knowledge and skills towards a larger class of problems is also likely to be successful.

Empirical Studies of the Theoretical Suggestions

This study discussed ways of assessing learning progress. Theoretical suggestions in specific include (a) ways of building a student and reference model, (b) three features of knowledge structure, and (c) a framework of learning progress. There have been three studies as initial efforts to validate the suggestions (Author, 2011a, 2011b, and 2011c). Some findings of the studies are introduced here, accompanied by related comments.

The Problem-Solving Task

The author (Author, 2011a, 2011b, and 2011c) gathered data from 136 undergraduate students enrolled in a course at a university in the southern United States and seven professors teaching at a major university in the United States. All participants made written responses to a specific complex problem. The task provided a simulated situation in which students were assumed to be participating in an evaluation project, the purpose of which was to investigate an unsuccessful project that had as its goal adapting a technology (i.e., a tablet PC) for classroom teaching. The task the author used provides an example of problem cases that could be used in assessment and instruction.

Reference Modeling

The author (2011a, 2011b) included a reference model building procedure with which an expert model was successfully created in accordance with the suggestion of this study. In particular the author employed the Delphi survey procedure (Goodman, 1987; Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). The panel composed of seven professors agreed with a reference model through three iterations of the Delphi survey to develop a refined model.

All panel members in the first round created their own response to the problem. Admittedly, diverse perspectives were observed in their initial responses that seemed not convergent but contrasting, even though there were some points on which they agreed. That observation was congruent with the claim that even experts elicit multiple representations of an ill-structured problem according to their own experiences (e.g., Jonassen, 1997). However, the Delphi procedure helped the panel to learn from each other and to achieve a consensus on the best understanding to the problem. That result proved that a reference model of an ill-structured problem can be created with proper methodological supports such as the Delphi procedure (Goldsmith & Kraiger, 1997; Spector, 2008a).

Three Features of Knowledge Structure

This study proposes three features of knowledge structure (3S): Surface, structure, and semantic. The author (2011c) validated the three features of knowledge structure using the

Confirmatory Factor Analysis (CFA). The author defined 10 parameters of concept maps obtained from the maps' structure and then associated the parameters with three features of knowledge structure that are assumed to be latent factors in the CFA models. The validation procedure includes sequential evaluation of a single factor model with a correlated group factor model. A single factor model was poor (CFI<0.90, NNFI<0.90). In contrast, the results of the CFA demonstrated that there is a three-dimensional feature of knowledge structure with good-fit indices (e.g., CFI>0.90, NNFI>0.90) (see Table 2.4).

Table 2.4

Summary of Fit Indices

Models	$\chi^2(df)$	$\chi^2/(df)$	CFI	NNFI	RMSEA	SRMR	AIC	ABIC
Single factor	699(35)	19.90	0.49	0.34	0.36	0.14	-2518	-2430
Three factor	96(26)	3.69	0.94	0.90	0.14	0.18	-3157	-3042

Note. Indices (their expected values) are: $\chi 2$(df) =Chi-square statistics and degrees of freedom for test of model fit (equal to 0); $\chi 2/(df)$ =the ratio of Chi-square statistic to the degrees of freedom (equal to 1), CFI=comparative fit index(>0.90), NNFI=non-normal fit index (a.k.a., Tucker-Lewis index)(>0.90), RMSEA=root square error of approximation (<0.05), SRMR=standardized root mean square residual (<0.05), AIC(Akaike Information Criterion (close to 0, smaller the better), ABIC(Adjusted Bayesian Information Criterion) (close to zero, smaller the better)

Stages of Learning Progress

The conceptualized model of learning progress was investigated in the author's (2011c) study. The author defined a set of measures indicating the levels of the features of knowledge structure. Relations between the measures and the features of knowledge structures were determined based on theoretical assumptions as well as empirical evidence obtained from the aforementioned CFAs.

Stages in the learning progress are inferred rather than directly observed. Thus, qualitative stages of learning progress are labeled latent classes because of their psychometric characteristics. Accordingly, latent class model (LCM) methods were employed. In particular, the study employed Log-linear Classification Diagnostic Model (LCDM). LCDMs are restricted latent class models that allow latent class models to place linear restrictions on the log-linear parameters (Rupp, Templin, & Henson, 2010). It is important to note that LDCM requires a substantive theoretical model so that researchers can interpret statistical classifications as meaningful latent classes. The hypothetical relations between the three features of knowledge structure and a set of measures in the study provided a theoretical model for LCDM analysis.

Table 2.5

Posterior Probabilities of Class Membership for the Stages of Learning Progress

Stage	Final Class Counts	Posterior Probability
Novice	99.680	0.696
Advanced beginner	11.224	0.078
Competent Learner	0.001	0.00001
Proficient Learner	32.092	0.224
Expert	-	-

Good-fit indices proved that the students proceed toward an expert-like knowledge structure through the suggested levels of learning progress (e.g., $\chi 2 = 48.435$, $p > .05$).As shown in Table 2.5, all stages were present. For example, the novice stage had an estimate of 0.69. This means that approximately 69% of respondents are classified as novices having not mastered any of the three features of knowledge structure. A large number of respondents were classified as novices or proficient learners, which resulted from positively correlated attributes (i.e., the three features of knowledge structure) similar with most assessment situations (Rupp, Templin, & Henson, 2010).

Implications for Research and Practice

Since Cronbach (1957) claimed that the cognitive development of a wide-range of students requires optimal learning environments suited to their needs. Engineering personalized learning has been studied in reference to adaptive instruction, (e.g., Lee & Park, 2007) which is an attempt to provide an individualized or group (when students share similar characteristics) learning environment. Creating adaptive learning environments necessitates knowing the extent to which students understand the given problem situations and the changes in their levels of understanding. In this paper, theories for learning progress are explored in reference to problem conceptualization, a framework of cognitive changes in expertise development, and in the learning-transfer in problem-solving situations. The study reveals some implications and remaining questions.

The theory of mental models accounts for how people conceptualize problem situations. Considering that mental models are symbolic (Johnson-Laird, 2005a, 2005b), symbol systems such as language and diagrams are deemed to be substantial in representing learners' mental models. Drawing upon symbol systems, current concept mapping tools elicit learners' understanding and can be sorted into two categories: (a) direct drawing tools such as Cmaptools (see http://cmap.ihmc.us/) and DEEP(see http://himatt.ezw.uni-freiburg.de/cgi-bin/hrun/himatt.pl) where learners directly draw their concept models, and (b) tools parsing natural language such as T-MITOCAR (http://himatt.ezw.uni-freiburg.de/cgi-bin/hrun/himatt.pl) and ALA Reader (http://www.personal.psu.edu/rbc4/ala.htm) in which learners' texts written in response to a given problem are parsed and represented as concept maps.

While the former type of tools seems to be more informative since it can describe directional and qualitative relationships (e.g., causal relationship) among concepts including annotation, they require extraneous cognitive load such as learning how to use a tool and possibly losing some contextual features of a problem due to its abstract approach. On the other hand, the latter type of tools use students' responses as written text that may represent students' natural thinking processes, but their representation of a concept map is still limited to describing simple propositional relationships without directional and qualitative information. According to the theory of mental models, mentally represented problem space is a structure including diverse relationships. Thus, assessment tools need to be adapted for the complex, dynamic structure of mental models so that diagnostic, formative information becomes more precise.

Second, the measureable features of the expert's mental structure need to be identified. In this study, for two reasons we restrict the functions of adaptive instruction to supporting students in their progress toward expert level rather than on helping them to be experts. The first reason is our lack of a means to assess and determine the expert level; the second is the limited number of problem cases available in an educational context. Nonetheless, in educational contexts, it is significant to identify expert learners, who may possibly bring many experiences into a class or who become experts during instruction, because they also require educational support pertinent to their states. Thus, theoretical suggestions that include some measurable attributes for determination of expert level mental models are requested.

Third, it is required to study longitudinal stage changes of mental models based on the given framework so that we can apply the model of learning progress to evaluating effectiveness of instruction and determine proper educational supports to an individual. Collins and Wugalter (1992) pointed out that psychological research and theory is increasingly turning to longitudinal studies, where development is monitored by following individuals over time. They introduced a measurement theory named dynamic latent variables, which are continuous quantitative variables that change systematically over time. Although these variables can be seen as continuous, many latent variables are best interpreted as a sequence of qualitative stages. Their assumptions are in congruence with the standpoint of developmental psychology (e.g., Bruner, 1961; Flavell, 1992; Piaget, 1964; Siegler, 2005; Siegler et al., 2009; Werner, 1957; Vygotsky, 1934/1976). A statistical test on the methodology can be illustrated in the experimental scenario in which multiple measurements of mental models are taken in reference to the crown-of-thorns starfish problem. In the investigational setting, we can suppose that researchers find a set of patterns in stage transition, their proportion, and transition probabilities among measurement points from the statistical analyses. These results can inform researchers of students' dynamic changes of learning progress and the effects of a given instructional intervention in a longitudinal manner. In short, statistical tests that detect, analyze a variety of transition patterns are required and then the detected patterns need to be validated by a qualitative review of student models.

Finally, we describe a couple of assessment situations in which conjectural assessment technologies, embedding the proposed framework of learning progress, are used. Although assessment tools are anticipated, the stage-sequential learning progress model provides possible qualitative classification by which learners' states are characterized and the learning stage changes are monitored in a longitudinal assessment setting. Based on these assumptions, we also envision reversion to an earlier or less sophisticated stage of learning progress. Regression in learning progress raises a few implications and further questions. For example, considering the expertise reversal effect, an advanced learner's unpredicted response may have dysfunctional assessment methodologies, designers should take into account advanced responses, which exceed their expectations, when designing problem cases and reference models.

Reverse progress implies that the learning progress is not explained purely by cognitive factors, but influenced by intertwined effects which include non-cognitive factors (e.g., interest and self-efficacy). As mentioned earlier, low interest, self-efficacious students may remain stuck in a stage or move backward to lower stages due to their ignorance of learning goals or given feedback (Kernis et al., 1989; Sadler, 1989). In order to understand the learning progress and provide better formative support, answers need to be provided regarding the relationships among cognitive and non-cognitive factors in learning progress. Following are some possible questions. What non-cognitive factors are related to developing expertise? To what extent is each non-cognitive factor associated with the change of learning stages? Like mental models as indicators of cognitive changes, can non-cognitive factors be classified in sequential stages? Can the five stage model of learning progress be a shared model that classifies non-cognitive factors? Can we establish an integrated model of learning progress including cognitive and non-cognitive factors? How can we determine a learner's stage based on an integrated perspective? What instructional and feedback strategies can be elaborated based on the framework of learning progress considering both cognitive and non-cognitive factors?

Closing Thoughts

If we wish to make an instructional program aimed at developing expertise in particular domains, we need to understand students' levels of understanding, monitor their progress, and provide personalized feedback. Students in different stages may have different needs because of the diverse states in their mental representations. In other words, instruction needs to be adaptive to individual differences.

This focused literature review and conceptual development paper provides theoretically grounded guidelines to address some issues of assessing learning progress based on the theory of mental models, the development of expertise, and learning transfer. For example, the theory of mental models provides practical suggestions on how to elicit learner models and how to build reference models for both structured and ill-structured problem-solving tasks. The research about expertise development presents plausible frameworks for qualitatively distinct developmental stages through which learners proceed. These theoretical findings leave significant practical questions, and theory has more value when it contributes to improving learning. For instance, devising the methodologies to estimate learners' stages and monitor their learning trajectories will be one area and the task to generate instructional strategies that adapt to a learner's developmental stages will be another.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273-298). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Anderson, J. R. (1980). *Cognitive psychology and its implications*. New York: Freeman.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). Learning and transfer. In *How people learn: Brain, mind, experience, and school* (pp. 31–78).Washington, DC: National Academy Press.

Bransford, J. D., & Franks, J. J. (1972). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331-350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717-726. doi:16/S0022-5371(72)80006-9

Bransford, J. D.,&Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61–100). Washington, DC: American Educational Research Association.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31*, 21-32.

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces*, *70*, 215-225.

Carver, S. M. (2006). Assessing for deep understanding. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning science* (pp. 205-221). New York: Cambridge University Press.

Chen, Z., & Siegler, R. S. (2000). Intellectual development in childhood. In R.Sternberg (Ed.) *Handbook of intelligence* (pp. 92-116). New York: Cambridge.

Chi, M.T.H. (2006). Two approaches to the study of experts' characteristics. In K.A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 121-130), Cambridge University Press.

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Collins, A. & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought*, (pp. 243-265). Cambridge, MA: Cambridge University Press.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*(1), 131-157.

Cox, B. D. (1997). The rediscovery of the active learner in adaptive contexts: A developmental-historical analysis of transfer of training. *Educational Psychologist, 32*(1), 41–55.

Cronbach, J. L. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671-684.

Gijbels, D., Dochy, F., den Bossche, P. V., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*(1), 27-61.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, *13*(5), 533-568. DOI:10.1016/S0959-4752(02)00025-7

Dreyfus, H. L., & Dreyfus, S. E. (2005). Peripheral vision: Expertise in real world contexts. *Organization Studies (after Jan 1, 2003)*, *26*(5), 779-792.

Dreyfus, H. L., & Dreyfus, S. E., with Athanasiou, T. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Oxford, UK: B. Blackwell.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 31-83). New York: Cambridge University Press.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233-241.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R. Hoffman, R. R. (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 685-706). Cambridge, UK: Cambridge University Press.

Ericsson, K., & Simon, H. (May 1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. doi:10.1037/0033-295X.87.3.215

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data (2nd ed.)*. Boston: MIT Press. ISBN 0262050293.

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93-102.

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology*, *36*, 613-648.

Flavell, J. H. (1985). *Cognitive development* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology, 28*(6), 998-1005.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon & D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 851–898). New York: Wiley.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205.

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chicheser: Ellis Horwood Ltd.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind*. Cambridge, MA: MIT Press.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393-408. doi:10.1037/0022-0663.95.2.393

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*, 263–297.

Gijbels, D., Dochy, F., den Bossche, P. V., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27-61.

Glaser, E. M., Abelson, H. H., & Garrison, K. N. (1983). *Putting knowledge to use: Facilitating the diffusion of knowledge and the implementation of planned change* (1st ed.). San Francisco, CA: Jossey-Bass.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. London: RoutledgeFalmer.

Goldsmith, T. E., & Kraiger, K. (1997). Applications of structural knowledge assessment to training evaluation. In J. K. Ford (Ed.), *Improving training effectiveness in organizations* (pp. 73–95). Mahwah, NJ: Lawrence Erlbaum Associates.

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing, 12*(6), 729-734.

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23–48.

Hoffman, R. R. (1998). How can expertise be defined?: Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds.), *Exploring expertise* (pp. 81-100). New York: Macmillan.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*, 332–340.

Hsu, C. & Sandford, B. A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation, 12*(10). Available online: http://pareonline.net/getvn.asp?v=12&n=10

Ifenthaler, D. (2006). *Diagnose lernabhängiger Veränderung mentaler Modelle. Entwicklung der SMD-Technologie als methodologische Verfahren zur relationalen, strukturellen und semantischen Analyse individueller Modellkonstruktionen*. Freiburg: Universitäts-Dissertation.

Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps.Educational Technology Research & Development, 58(1), 81-97. doi:10.1007/s11423-008-9087-4.

Ifenthaler, D., & Seel, N. M. (2005). The measurement of change: Learning-dependent progression of mental models. *Technology, Instruction, Cognition, and Learning, 2* (4), 321-340.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185-208). Cambridge University Press.

Johnson-Laird, P.N. (2005b). The history of mental models. In K. I. Manktelow, & M. C. Chung (Eds), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179-212). Psychology Press.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *ETR&D, 45*(1), 65-94.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *ETR&D, 48*(4), 63-85.

Jonassen, D., & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development, 47*(1), 61-79. doi:10.1007/BF02299477

Judd, C. H. (1908). The relation of special training to general intelligence. *Educational Review, 36*, 28–42.

Kail, R. V. (2004). Cognitive Development Includes Global and Domain-Specific Processes. *Merrill Palmer Quarterly Journal of Developmental Psychology, 50*(4), 445-455.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31.

Katz, J. J., & Postal, P. M.(1964). *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press.

Kendler, H. H., & Kendler, T. S. (1962). Vertical and horizontal processes in problem solving. *Psychological Review, 69,* 1-16.

Kernis, M. H., Brockner, J., & Frankel, B. S. (1989). Self-esteem and reactions to failure: The mediating role of overgeneralization. *Journal of Personality and Social Psychology, 57*, 707–714.

Author (2011a). ---------------

Author (2011b). ---------------

Author (2011c). ---------------

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394. doi:10.1037/0033-295X.85.5.363

Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (2002). The role of textual coherence in incremental analogical mapping. *Journal of memory and language, 47*(3), 407–435.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life.* Cambridge, United Kingdom: Cambridge University Press.

Lee, J., & Park, O. (2007). Adaptive instructional systems. In J. M. Spector, M. D. Merill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 469-484). Routledge: Taylor & Francis Group.

Lemaire, P, & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General, 124* (1), 83-97.

Lobato, J. (2006). Alternative perspectives on the transfer of learning: History, issues, and challenges for future research. *Journal of the Learning Sciences*, *15*(4), 431-449.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Norman, D. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7-14). Hillsdale, NJ: Erlbaum.

National Research Council (NRC). (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management, 42*(1), 15-29. doi:10.1016/j.im.2003.11.002

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49,* 301-332.

Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered* (pp. 7-20). Ithaca, NY: Cornell University.

Piaget, J. (1976). *Science of education and the psychology of the child*. New York: Orion Press.

Pirnay-Dummer, P. (2006). *Expertise und Modellbildung – MITOCAR*. Freiburg: Universitäts-Dissertation.

Pretz, J. E., Naples, A. J., & Sternberg, R. J. (2003). Recognizing defining, and representing problems. In J. E. Davidson & R. J. Sternberg, *The psychology of problem solving* (pp. 3-30). New York: Cambridge University Press.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*(3), 381–400.

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: The Guilford Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*, 113–142.

Schlomske, N., & Pirnay-Dummer, P. (2008). Model based assessment of learning dependent change during a two semester class. In Kinshuk, Sampson, D., & Spector, M. (Eds.), *Proceedings of IADIS International Conference Cognition and Exploratory Learning in Digital Age 2008*, Freiburg, Germany, pp. 45-53.

Schvaneveldt, R.W., Durso, F.T., Goldsmith, T.E., Breen, T.J., & Cooke, N.M. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23*, 699-728.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science, 29* (4-5), 403-427.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition, and Learning, 1* (1), 59-85.

Seel, N. M. (2004). Model-centered learning environments: Theory, Instructional design, and ethics. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instruction design: International perspectives* (pp. 49-74). NJ: Lawrence Erlbaum Associates, Inc.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation, 1*(1), 4-35.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation, 23*(4), 373-398.

Shaklee, H. (1979). Bounded rationality and cognitive development: Upper limits on growth? *Cognitive Psychology, 11*, 327-345.

Shute, V. J., & Zapata-Rivera, D. (2007). Adaptive technologies. In J. M. Spector, M. D. Merill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 227-294). Routledge: Taylor & Francis Group.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8,* 165–190.

Siegler, R. S. (2005). Children's Learning. *American Psychologist, 60*, 769-778.

Siegler, R. S., & Klahr, D. (1982). When do children learn: The relationship between existing knowledge and the ability to acquire new knowledge. In R. Glaser (Ed.), *Advances in Instructional Psychology: Vol. 2* (pp. 121-211). Hillsdale, NJ: Erlbaum.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3,* 143-150.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Snow, R.E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14* (5), 455–473.

Spector, J. M. (2004). Problems with problem-based learning: Comments on model-centered learning and instruction in Seel (2003). *Technology, Instruction, Cognition and Learning, 1*(4), 359-374.

Spector, J. M. (2008a). Expertise and dynamic tasks. In H. Qudrat-Allah, J. M. Spector, & P. I. Davidsen (Eds.), *Complex decision making: Theory and practice*. Berlin: Springer-Verlag.

Spector, J. M. (2008b). Cognition and learning in the digital age: Promising research and practice. *Computers in Human Behavior, 24*(2), 249-262.

Spector, J. M. (2009). Adventures and advances in instructional design theory and practice. In L. Moller, D. Harvey & J. Huett (Eds.), *Learning and instructional technologies for the 21st century: Visions of the future* (pp. 1-25). Berlin: Springer.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Sternberg, R. J. (1999). The theory of successful intelligence. *Review of General Psychology*, 3(4), 292-316.

Sternberg, R. J. (2008). Applying psychological theories to educational practice. *American Educational Research Journal, 45*(1), 150-165.

Sternberg, R. J.,&Frensch, P. A. (1993).Mechanisms of transfer. In D. K. Detterman&R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 25–38).Norwood, NJ: Ablex.

Sugrue, B. (1995). A theory-based framework for assessing domainl-specific problem-solving ability. *Educational Measurement: Issues and Practice, 14*(3), 29-35. doi:10.1111/j.1745-3992.1995.tb00865.x

Thorndike, E. L. (1906). *Principles of teaching*. New York: Seiler.

van Gog, T., Ericsson, K. A., Rikers, R. M., & Paas, F. (2005). Instructional design for advanced learners: establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *ETR&D, 53*(3), 73-81.

Vosniadou, S., & Skopeliti, I. (2005). Developmental shifts in children's categorization of the earth. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society* (pp. 2325-2330).

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York: Routledge.

Vygotsky, L. (1934/1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis, MN: University of Minnesota.

CHAPTER 3

CROSS-VALIDATION STUDY OF METHODS AND TECHNOLOGIES TO ASSESS MENTAL MODELS IN A COMPLEX PROBLEM SOLVING SITUATION[2]

[2] Kim, M. 2012, *Computers in Human Behavior*, *28* (2): 703-717.
Reprinted here with permission of publisher.

Abstract

This paper reports a cross-validation study aimed at identifying reliable and valid methods and technologies for natural language (i.e., written text) responses to complex problem-solving scenarios. In order to investigate current assessment technologies for text-based responses to problem-solving scenarios (i.e., ALA-Reader and T-MITOCAR), this study compared the two best developed technologies to an alternative methodology. Comparisons amongst the three models (benchmark, ALA-Reader, and T-MITOCAR) provided two findings: (a) the benchmark model created the most descriptive concept maps; and (b) the ALA-Reader model had higher correlation with the benchmark model than did T-MITOCAR. The results imply that the benchmark model is a viable alternative to two existing technologies and is worth exploring in a larger scale study.

Keywords: assessment technology, concept map, mental models, problem solving, validation study

This study investigated current methods and technologies that yield concept maps ─structural knowledge representations consisting of concepts and relations (Clariana, 2010; Narayanan, 2005; Novak & Canãs, 2006; Spector & Koszalka, 2004)─as re-representations of a student's mental models. This study is a kind of cross-validation[3] aimed at identifying which methods work best in terms as forming the basis for dynamic formative feedback. It is assumed that using natural language (written text) responses as a basis for concept map representations of student thinking is likely to provide a reliable foundation for use in providing formative feedback and assessment (Pirnay-Dummer, Ifenthaler, & Spector, 2010).

There is a common belief that problem solving includes conceptualizing the problem space, which involves creating a knowledge structure that integrates ideas and concepts that a problem solver associates with the problem situation (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Jonassen, Beissner, & Yacci, 1993; Newell & Simon, 1972; Segars, 1997). As a consequence, assessing problem solving should naturally take into account the constructed knowledge structure (Gijbel, Dochy, Van den Bossche, & Segers, 2005); simple knowledge tests are somewhat weak measures of problem-solving ability (Grotzer & Perkins, 2000; Thomas, 2005).

In order to capture structural knowledge, a number of technologies have been developed, including: DEEP (Dynamic Evaluation of Enhanced Problem-solving; see Spector & Koszalka, 2004); SMD (Surface, Matching, and Deep Structure; Ifenthaler, 2009); T-MITOCAR (Text Model Inspection Trace of Concepts and Relations; Pirnay-Dummer, Ifenthaler, & Spector 2010); CmapTools (Novak & Canãs, 2006); jMap (Jeong, 2008); ACSMM (Analysis Constructed Shared Mental Model; O'Connor & Johnson, 2004); KU-Mapper (Clarian & Wallace, 2009);

[3] Cross-validation in this study denotes the practice of identifying the most reliable concept map technology by comparing concept maps elicited from multiple technologies using the same data.

ALA-Mapper (Analysis of Lexical Aggregates-Mapper; Taricani & Clariana, 2006; Clariana, Wallace, & Godshalk, 2009); ALA-Reader (Analysis of Lexical Aggregates-Reader; Clariana & Wallace, 2007, Clariana et al., 2009); and KNOT (Knowledge Network Orientation Tool; Schvaneveldt, 1990).

Current technologies either require learners to create an annotated concept map with rich descriptions of links and nodes (DEEP) or else they use text responses as an interim step in generating a concept map (T-MITOCAR and ALA-Reader) that can then be assessed with tools such as SMD or KNOT. All of these technologies have limitations in terms of their suitability, reliability, and validity (Kalyuga, 2006; Seel, 1999; Spector, Dennen, & Koszalka, 2006).

This paper addresses focuses on methods that use text responses to generate a concept map that can then be assessed and explores an alternative approach that attempts to restore rich descriptions of links between nodes. Prominent methods and technologies are classified and analyzed in terms of their merits and deficiencies. Next, alternative methods and technologies to analyze student responses in the form of written text are selected. Finally, cross-validation among the selected technologies is performed, analyzed and reported. Based on the results, an alternative approach to consider in automatically constructing and assessing concept maps based on open-ended text responses to a problem situation is then described.

Concept Maps as Re-represented Mental Models through Language Inputs

Mental Models as Inferred Entities

Mental models are cognitive artifacts resulting from perception and linguistic comprehension, representing certain aspects of external situations (Johnson-Laird, 2005a, 2005b). In this perspective, knowledge appears to be a configuration of holistic mental representations. Mental model representations consist of propositional representations as structured symbols and images (Johnson-Laird, 2005b; Newell, 1990). A concept map, such as the externally-represented structural component of mental models, implies that a latent structure exists in the human mind. In other words, a concept map is a first or second order representation of a primary representation – a mental model, which is an inferred entity. The primary representations (mental models) drive actions and decisions, which are external indicators of learning. In order to provide formative feedback, however, it is necessary to make inferences about the mental models that are behind decision making and problem-solving activities. Many empirical studies utilizing concept map techniques have shown that as students gain competence in a discipline, their structural comprehension becomes more coherent and expert-like (e.g., Schlomske & Pirnay-Dummer, 2008; Schvaneveldt, Durso, Goldsmith, Breen, & Cooke, 1985; Spector & Koszalka, 2004).

Concept Maps as Analogues of Mental Models

Concept mapping is a method that elicits cognitive representations of an individual's structural knowledge involving interrelated concepts (Axelrod, 1976; Carley & Palmquist, 1992; Narayanan, 2005). In concept maps as representations of semantic networks (Quillian, 1968; Collins & Loftus, 1975; Jonassen et al., 1993), the strength of links may be interpreted as the strength of belief in a given semantic relationship, which is reflected by link weights (Shute, Jeong, Spector, Seel, & Johnson, 2009; Shute & Zapata-Rivera, 2008).

Language plays a key role in creating a concept map. Concept maps can represent pairs of related words, such as a noun (concept)-verb (relation)-noun (concept) relationship. The data used for concept mapping is generally collected from interviews or texts (Carley & Palmquist, 1992; Narayanan, 2005). Text-based data collection is economical in terms of time and effort (Brown, 1992), and is based on techniques that avoid recall bias and potentially leading or misleading questions (Axelrod, 1976).

Mental models are formulated symbolically (Seel, 2001); that is, symbols play a central role in representing ideas and thoughts. Meaning is constructed with cognitive effort (thinking and reasoning) often utilizing symbolic notations which help individuals to re-represent their thoughts (Greeno, 1989). According to Garnham (1987, 2001), the theory of mental models provides a unified account of language processing, thinking, and reasoning. A mental model is constructed based on situational inputs and can be re-represented in text or discourse. In order to visually represent concept maps from text, technological support in terms of natural language processing and network analysis technology are required.

State-of-the-Art Concept Map Technologies

Concept maps are generally visually represented through network analysis using a set of techniques to portray patterns of relations among nodes (Coronges, Stacy, & Valente, 2007; Hutchison, 2003; Wasserman & Faust, 1994). Most of the techniques involve mathematical algorithms derived from graph theory (Rupp, Sweet, & Choi, 2010b; Schvaneveldt, Durso, Goldsmith, Breen, & Cooke, 1989; Wasserman & Faust, 1994). In these techniques, proximity data between and among concepts is defined as “judgments of similarity, relatedness, or association between entities frequently used in the study of human cognition” (Schvaneveldt et al., 1989, p. 249). Drawing on graph theory and proximity data, specific statistical methods have been used. Pathfinder techniques (Jonassen et al., 1993, Schvaneveldt, 1990; Schvaneveldt et al., 1985) to analyze simple association networks using multi-dimensional scaling is an early statistical technique used to assess concept maps. More recently, social network analysis has been used (Rupp, Gushta, Mislevy, & Shaffer, 2010a; Shaffer, Hatfield, Svarovsky, Nash, Nulty, & Bagley et al., 2009; Wasserman & Faust, 1994).

Figure 3.1 illustrates relationships amongst methods and technologies in a network analysis procedure. In general, network analysis employs a three-step procedure (Curtis & Davis, 2003; Taricani & Clariana, 2006): (1) elicit judgments about concept relationships; (2) construct concept maps; and (3) compare the concept maps to the reference model.

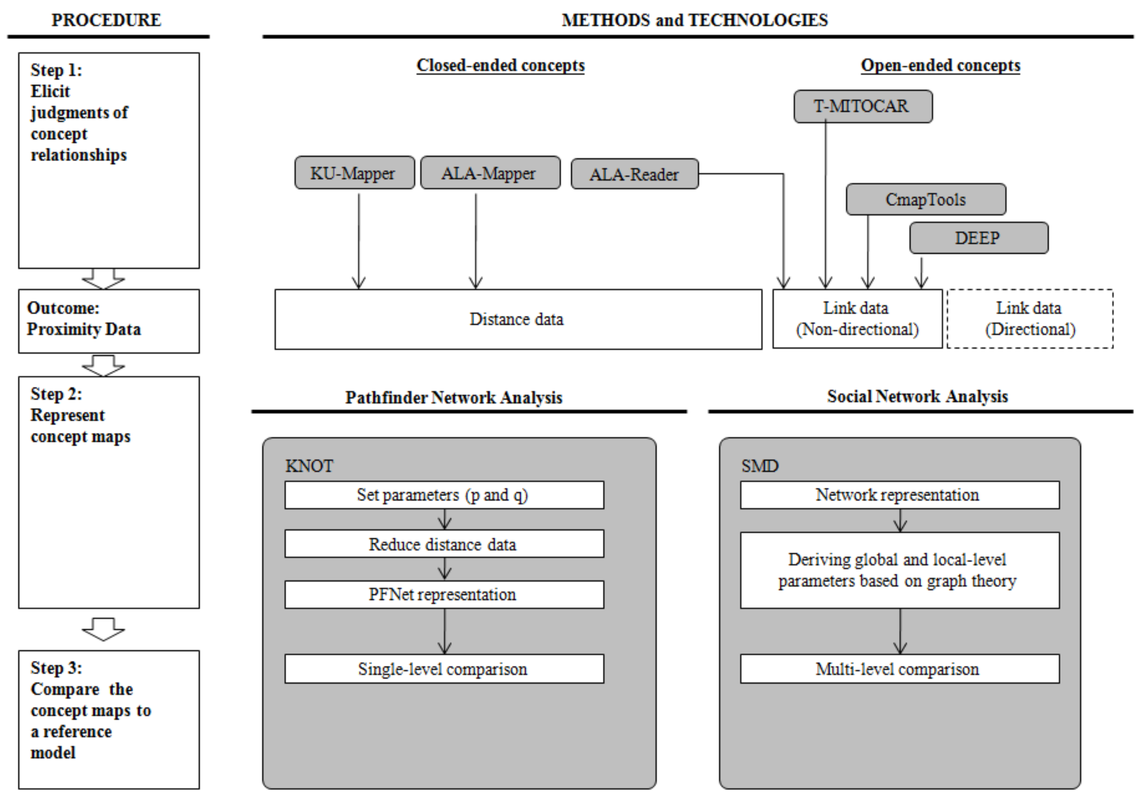


Figure 3.1. Procedure, methods, and technologies applied to network analysis.

Step 1: Elicit Judgments about Concept Relationships

The first step, eliciting judgments about concept relationships, is the essential phase because it yields data that contains all captured concepts and their relationships in a student's response. There are two kinds of concept map approaches involving natural language processing: the 'closed-ended concepts approach' and the 'open-ended concepts approach' (see Figure 3.1). The closed-ended provides the student with a predefined list of concepts and links (Taricani and Clariana, 2006). The open-ended approach allows students to use whatever concepts and linking terms they desire. In Figure 3.1, the class of closed-ended approaches includes KU-Mapper, ALA-Mapper, and ALA-Reader; the class of open-ended approaches includes T-MITOCAR, CmapTools, and DEEP. CmapTools and DEEP allow students to draw concept maps based on either predefined concepts or free use of any concepts. ALA-Reader accepts text input with no limitation (Clariana et al, 2009; Taricani & Clariana, 2006), but that tool is classified as closed-ended concepts because the technology only retrieves a predefined list of words from the open-ended text. Many researchers consider open-ended concept mapping as the gold standard for capturing students' mental models (McClure, Sonak, & Suen, 1999; Ruiz-Primo, Schultz, Li, & Shavelson, 1999; Spector & Koszalka, 2004; Taricani & Clariana, 2006). This study also centers on open-ended concept mapping as a way to elicit a natural and descriptive knowledge structure.

The aforementioned technologies generate two types of proximity data (i.e., an *n* by *n* matrix where *n* is the number of terms−concepts), representing distance or adjacency, depending on the technologies. Distance data generated by such tools (KU-Mapper and ALA-Mapper) include all of the pair-wised distances between terms that are calculated by the location of the terms in a space (e.g., computer screen) or directly judged by students with ordinal scale (e.g., 1 to 9) (Clariana et al., 2009; Taricani & Clariana, 2006). In the case of adjacency data, the relatedness of paired terms is represented in a matrix, the *n* by *n* matrix with '1' (that is entered when two terms are connected) or '0' (that is entered when two terms are not associated) (Clariana et al., 2009; Rupp et al., 2010b; Schvaneveldt, 1990; Schvaneveldt et al., 1989; Wasserman & Faust, 1994). Distance data have mathematical foundations in geometry whereas adjacency data are directly derived from graph theory (Schvaneveldt, 1990; Schvaneveldt et al., 1989). Adjacency data is an alternative to distance data due to their benefit in describing a directional relation where the connection between a pair of concepts has an origin and a destination (Wasserman & Faust, 1994). In the sense that directional adjacency data can describe detailed structural information such as a causal relation between concepts, this study takes account of adjacency data that is more descriptive and complex.

In addition, Clariana and colleagues (2007, 2009) proposed that editing pronouns to nouns the pronouns point out in text is likely to help better capture relevant information about the knowledge structure than does simply ignoring all pronouns. Although their experiment concluded that there was little difference between editing and ignoring pronouns, this study included a comparison between two types of data when determining a better benchmark model.

In regard to the classifications discussed here, we can make a set of combinations of concept mapping approaches that are listed from more complex and descriptive to simple and economic. Again, all listed combinations below can be divided into two (i.e., noun-only and pronoun-edited) according to the data handling subroutine:

- Natural language + open-ended concepts + directional adjacency data
- Natural language + open-ended concepts + non-directional adjacency data
- Natural language + closed-ended concepts + directional adjacency data
- Natural language + closed-ended concepts + non-directional adjacency data

Step 2: Construct Concept Maps

The technologies process proximity data (either distance or adjacency data) and then construct concept maps. Constructing concept maps is based on two network analysis methods: Pathfinder network analysis and social network analysis. Pathfinder network analysis (PFA) is "a data reduction approach that emphasizes the main pair-wise associations in proximity data" (Taricani & Clariana, 2006, p. 730) by constraining the data with *r* and *q* parameters. *r* parameter works as a function of the weights of links in the path and *q* parameter places an upper limit on the number of links in paths (See Schvaneveldt, 1990; Schvaneveldt et al., 1989). In short, pathfinder algorithms with particular *r* and *q* parameters yield the minimum number of links in an *n* by *n* array. Considering the function of PFA, it seems that PFA is a heuristic method to find meaningful relations among many of the concepts (Goldsmith, Johnson, & Acton, 1991). However, it is often observed that many studies with a small set of concepts utilize PFA. Likewise, in the case that key terms are pre-defined from the referent network, the use of PFA would be less effective. Otherwise, the approach to identifying key terms and relations in a large number of concepts appears to be more pertinent to PFA. For example, Coronges and colleagues (2007) conducted more descriptive analysis (i.e., social network analysis) and then ran PFA.

Social network analysis (SNA) is useful in the study of knowledge structure although it was developed to study social relationships (Coronges et al., 2007; Knoke & Kuklinski, 1982; Hage & Harary, 1983; Wasserman & Faust, 1994). Coronges and colleagues (2007) used SNA to analyze a cognitive network labeled Cognitive Associative Network (CAN). Another example is the Epistemic Network Analysis (ENA) by Rupp and colleagues (2010a, 2010b) in which they created knowledge networks (concept maps) and compared the networks using measures resulting from SNA. Network measures computed by SNA include global-level measures that describe the overall entire network (e.g., centralization, size, density, clustering, and path length) and local-level measures, such as centrality of each concept, which can be used to determine the salience of concepts (Coronges et al., 2007; Wasserman & Faust, 1994).

As a result, when we use adjacency data (either directional or non-directional), employing SNA provides more benefits for research in that it generates diverse network parameters that are compared to a reference network or other networks (e.g., a previously elicited network). SMD, a set of analysis functions using adjacency data (Ifenthaler, 2007; Ifenthaler et al., 2009), is consistent with SNA. Table 3.1 shows that the measures used in SMD are directly adapted to the analysis techniques specified in SNA.

Table 3.1

Measures for Analyzing the Organization of Cognitive Structure

SMD Measure	Operationalization	Social Network Measure[a]
Surface structure	The overall number of propositions	The number of edges (links)
Graphical structure	The complexity of a cognitive structure indicates how broad the understanding of the underlying subject matter is	Geodesics distance and diameter of a network
Connectedness	A connected cognitive structure indicates a deeper understanding of the underlying subject matter	Connectedness
Ruggedness	Non-linked vertices of a cognitive structure point to a lesser understanding of the phenomenon in question	Non-linked nodes
Average degree of vertices	As the number of incoming and outgoing edges grows, the complexity of the cognitive structure is taken as more complex	Average degree
Cyclic	A non-cyclic cognitive structure is considered less sophisticated	Cycle (closed walk) or acycle
Number of cycles	A cognitive structure with many cycles is an indicator for a close association of the vertices and edges used	Cohesive subgroups
Vertices	A simple indicator for the size of the underlying cognitive structure	The number of nodes
Vertex matching	The use of semantically correct concepts (vertices) is a general indicator of an accurate understanding of the given subject domain	Shared number of nodes
Propositional matching	The use of semantically correct propositions (vertex-edge-vertex) indicates a correct and deeper understanding of the given subject domain	Configural similarity (KNOT; Taricani & Clariana, 2006)

Note. It was modified from Ifenthaler et al., (2009). a. It referred to Wasserman and Faust (1994).

Step 3: Compare the Concept Maps to the Reference Model

Evaluation of the derived concept maps is often done by comparison with a reference model, which is often elicited from an expert (Curtis & Davis, 2003; Goldsmith & Kraiger, 1997; Coronges et al., 2007; Taricani & Clariana, 2006). Comparison between concept maps is indicated by similarity measures assessed by overlaying network patterns with the concept map information (Coronges et al., 2007; Monge & Contractor, 2003). KNOT (i.e., a PFA tool mostly employed in studies) software includes the function to gauge two similarity parameters: common and configural similarity. The common similarity is simply the total number of links shared by two concept maps. The configural similarity is calculated by dividing the total number of shared propositions by the total number of unique links in the student's and the expert's concept maps ranging from 0 (no similarity) to 1 (perfect similarity). These similarity values are used as concept map scores (Clariana et al., 2009; Taricani & Clariana, 2006). Similarity measures can be extended including diverse network parameters.

Table 3.2

Seven Similarity Measures of T-MITOCAR

Surface	Compares the number of concepts between two models (graphs).
Graphical Matching	Compares the diameters of the spanning trees of the graphical mental models as an indicator for the range of conceptual knowledge.
Structural Matching	Compares the complete structures of two graphs regardless of their contents.
Gamma	Compares the two graphs' gammas that indicate each graph's average percentage of the links that are actually present for a node.
Concept Matching	Compares the sets of concepts within a graph to determine the use of terms.
Propositional Matching	Compares only fully identical propositions between two graphs, which are used for quantifying semantic similarity between two graphs.
Balanced Semantic Matching	Uses both concepts and propositions to match the semantic potential between two model representations.

Note: It refers to Pirnay-Dummer and Ifenthaler (2010).

Spector and Koszalka (2004) first introduced the surface, structure, and semantic similarity features in conceptual form. In accordance with those three features, many cognitive scientists have proposed similarity measures for investigating the achievement of complex problem-solving knowledge and skills (Clariana et al., 2009; Ifenthaler, 2006; Pirnary-Dummer, 2006; Taricani & Clariana, 2006). For example, T-MITOCAR (Pirnay-Dummer and Ifenthaler, 2010) provides six similarity measures as comparison results based on multiple network measures from SMD (see Table 3.2).

Method

This study has two aims: (1) identify the methods that consistently yield the most descriptive and accurate concept maps; (2) validate these methods and technologies using a reference method. The most complex condition can be expressed in this combination: natural language + open-ended + directional adjacency data + pronoun-edited. As to the network analysis method, the social network analysis (SNA) is considered as an alternative in terms of obtaining a more descriptive concept map and diverse network information. However, considering that a complex approach is only required when it offers a greater benefit over a less complex one, more complex approaches were compared to less complex approaches. In addition, this comparison process serves to validate applied technologies as well.

Participants

Participants included 20 students and seven experts. The original student data were gathered from 136 undergraduate students enrolled in a course at a university in the southern United States. The course aimed to educate students on knowledge and skills for integrating technology in teaching and learning. In the class, students made written responses to a specific complex problem. For this study, from the original group of students whose responses contained more than 350 words, a random selection of 20 students was made; the reason for restricting the selection to responses with more than 350 words is that one of the selected technologies, T-MITOCAR, requires at least 350 words for an analysis. Fifteen students were female and five were male. Of the 20, 10 were in their junior year, five were sophomores, and five were at the senior level. Seven expert responses were gathered from seven professors teaching at six major universities in the United States. It was assumed that using expert responses would enable us to investigate the technologies' ability to detect higher-level responses.

The Problem-solving Task

All participants were asked for responses to a complex problem situation using natural language. The task provided a simulated situation in which students were assumed to be participating in an evaluation project, the purpose of which was to investigate an unsuccessful project that had as its goal adapting a technology (i.e., a tablet PC) for classroom teaching. In order to elicit students' knowledge in detail, the questions asked them to explicitly describe the concepts, issues, factors, and variables likely to have contributed to the result that the introduction of tablet PCs had very little effect on the instructional practices employed in the classes.

Reference Modeling via a Delphi Survey

This study included a reference model for the problem situation. The model was created using a Delphi survey procedure (Goodman, 1987; Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). The Delphi survey involved three iterations to develop a refined reference model that the seven experts accepted. In the first round, the participating professors created their own responses to the problem; then, all the panel's responses were consolidated. Next, a document including all statements from the professors and a list of concepts identified from the panel's responses was sent to the panel again. The professors were asked to add their comments regarding the listed statements and concepts and rank them. After gathering the second round of surveys, the researcher created a final list of ranked statements and concepts. Based on this summary, a draft of a reference model was created. In the final round, the results of the second survey were sent to the panel and revised according to their comments as necessary. Through this procedure, a reference model containing 23 key concepts was developed.

Benchmarks

Initially, four types of benchmarks were prepared according to the combinations of noun-only vs. pronoun-edited and directional vs. non-directional. To prepare for pronoun-edited responses, all responses were reviewed. Directional relations were determined when retrieving paired concepts from text.

To distill concepts and relations from text responses, according to the author's (2011) semantic relation approach, the researcher manually distilled the semantic relations that are the underlying relations between two concepts expressed by words or phrases. The approach involves diverse types of relations of concepts beyond the typical noun-verb-noun relation form including genitives (e.g., teachers' participation), prepositional phrases attached to nouns (e.g., technology in school classrooms), or sentence (e.g., Emerging new media has always led to instructional changes.). Thereafter, all distilled concepts and relations were summarized in adjacency data. The distilled concepts and relations were cross-checked by a doctoral student. There were no significant issues regarding the data. All concepts and relations were retained without changes; potential issues would have had no significant impact on the results considering the number of concepts and relations in each concept map. The adjacency data were processed via the selected network analysis package, NetMiner (http://www.netminer.com/), which provides the capability for creating concept maps and generating a variety of concept map information.

Finally, in order to obtain a list of similarity measures in the form of concept map scores, student models were compared to the reference model. The similarity measures were calculated using the similarity tool developed for this study. The tool was developed using C++ and validated by comparisons between randomly selected tool-generated and manually calculated sample data.

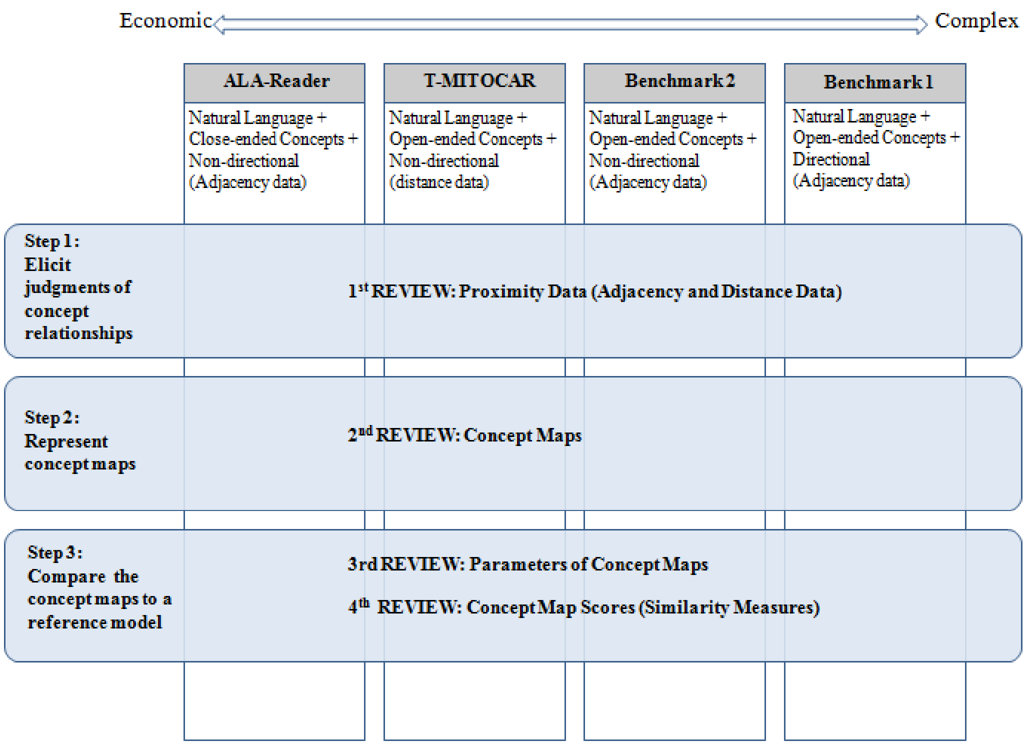


Figure 3.2. Cross-validation procedure.

Cross-Validation Procedure

Along with multiple benchmarks, two natural language technologies, T-MITOCAR and ALA-Reader, were selected and validated against the benchmark model. Figure 3.2 illustrates the cross-validation procedure. The concept mapping approach in which each method is embedded is matched as follows:

- Benchmark 1: open-ended + directional (adjacency data)
- Benchmark 2: open-ended + non-directional (adjacency data)
- T-MITOCAR: open-ended + non-directional (distance data)
- ALA-Reader: close-ended + non-directional (adjacency data)

Regarding the benchmarks, there are two types of models; noun-only and pronoun-edited. Each model can be reclassified as directional or non-directional. In the Figure 3.2, only the second-level classification was included. At the beginning of the analysis, the competing models of the benchmarks such as noun-only versus pronoun-edited and directional versus non-directional were compared so that a benchmark model could be determined for comparisons amongst different methods and technologies.

As Figure 2 shows, there were four comparisons made across the outcomes of each selected method through the concept mapping procedures (steps 1 through 3). The cross-validation includes two comparisons of outcomes from different sets. In the first review, the correlation and similarity were analyzed for comparisons amongst the proximity data. In the second review, visual inspections of the concept maps were conducted as a qualitative analysis. The third review involved a second correlation and similarity study among concept map parameters obtained from each method: (a) number of relations; (b) diameter; (c) gamma (cluster coefficient; see Table 3.2); and (d) the number of cycles (cohesive subgroups; see Table 3.1).

The last review was a comparison across various concept map scores, which are computed similarities between the student model and the reference model (see Pirnay-Dummer & Ifenthaler, 2010): (a) surface matching; (b) graphical matching; (c) concept matching; (d) gamma matching; (e) propositional matching; and (f) balanced semantic matching.

Data Analyses

Data comparison. This study validated the outcomes from the competing concept map methods such as proximity data, concept map parameters, and concept map scores. The validation methods included the numerical and pattern reviews. In the following, a numerical similarity measure was applied to see how far or near the numbers of the two models are:

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

where f_1 and f_2 denote the numerical frequency of each method compared. The similarity ranges from 0 to 1, $0 \leq s \leq 1$. To review associated patterns, Pearson correlation-coefficients were calculated between the benchmark and competing methods.

Modification of similarity measures (concept map scores). This study modified and adjusted the formulas that calculated the concept map scores. The modified formulas used for obtaining concept map scores of the benchmark and ALA-Reader.

On the whole, a similarity formula assumes each part of a pair is equally significant. In the case of a concept model comparison, the reference model and student model are not equal in terms of maturity. A reference model acts as criteria and a student model is expected to progress toward the reference model. It is assumed that a reference model is likely to contain a greater number of concepts and relations and to build a larger knowledge structure than a novice model (Chi, Glaser, & Farr, 1988; Spector & Koszalka, 2004).

As for numerical similarity, a modified algorithm was applied except for the gamma similarity (refer to Table 3.2). In case f_1 is smaller than f_2, $f_1 < f_2$, the original numerical similarity formula was used so that

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

where the frequency of a student model is f_1 and that of a reference model is f_2. Otherwise, if f_1 is not less than f_2, $f_1 \geq f_2$, the similarity value was set as ‘1’ because the student value is greater than that of the reference. That is, it indicates that the student model exceeds the reference model according to the relevant criteria.

Similarly, concerning the conceptual similarity as applied to the concept and propositional matching score, two adjustments were made. Just as a picture resembles an object rather than an object resembling a picture of it, a student model to some degree resembles the reference model that is more salient. In this asymmetric relation, the features of the student model are weighted more heavily than those of the reference (Colman &Shafir, 2008; Tversky & Shafir, 2004). When the conceptual similarities of the benchmark were calculated by Tversky's (1977) formula,

$$\mathrm{s} = \frac{f(\mathrm{A} \cap \mathrm{B})}{f(\mathrm{A} \cap \mathrm{B}) + \alpha \cdot f(\mathrm{A} - \mathrm{B}) + \beta \cdot f(\mathrm{B} - \mathrm{A})}$$

α was weighted more heavily than β ($\alpha = 0.7$ and $\beta = 0.3$). However, in the case of the ALA-Reader, since it has a predefined set of concepts and relations, the reference mode, f (B), always includes a student model, f (A). Therefore, α was set as '0,' whereas β was set as '1.'

Results

Determining a Benchmark

Two comparisons amongst the methods (the noun-only versus pronoun-edited and the directional versus non-directional) for creating benchmark models were implemented so that we could decide a reliable and economical way to establish a benchmark. The first review was of the noun-only and pronoun-edited using the numerical similarity between the two approaches.

As Table 3.3 summarizes, very high average numerical similarities (s > 0.93) were observed between the noun-only and pronoun-edited data of concepts and relations in both the directional and non-directional models. Similar to Clariana and colleagues' (2007, 2009) conclusion, there was such little difference that it was determined that the noun-only model was more economical.

Table 3.3

Numerical Similarities of Concepts and Relations between the Noun-Only and Pronoun-Edited

	Noun-Only vs. Pronoun-Edited (Directional)		Noun-Only vs. Pronoun-Edited (Non-Directional)	
	Concept	Relation	Concept	Relation
Min	0.74	0.70	0.74	0.70
Max	1.00	1.00	1.00	1.00
Average	0.97	0.93	0.97	0.93

Note. Sample size *N* = 28.

Next, the similarities between the directional and non-directional data were investigated (see Table 3.4). The directional and non-directional data have the same set of concepts in a given setting, either the noun-only or pronoun-edited. Only the numerical similarities of relations were calculated. The very high average numerical similarities (s = 0.98) demonstrated that there was little difference between the directional and non-directional data of relations in both noun-only and pronoun-edited models.

In order to certify the aforementioned findings, matrix correlations amongst the four types of benchmarks were reviewed for the selected four samples, including two randomly selected samples (students 78 and 83), the reference model, and an extreme case (student 20). As an outlier, student 20 frequently used pronouns resulting in a relatively greater difference between noun-only and pronoun-edited analyses.

Table 3.4

Numerical Similarities of Relations between the Directional and Non-Directional

	Directional vs. Non-Directional (Noun-Only)	Directional vs. Non-Directional (Pronoun-Edited)
Min	0.92	0.92
Max	1.00	1.00
Average	0.98	0.98

Note. Sample size *N* = 28.

For all samples, matrix correlations between the directional and non-directional were around 0.5 because the relation R_{ij} is the same as R_{ji} in non-directional data, $R_{ij} = R_{ji,}$ whereas R_{ij} is different from $R_{ji,}$ in directional data. In a specified directional or non-directional condition, even student 20 had a higher correlation of 0.703 between the noun-only and pronoun-edited data (see Table 3.5). It was concluded that a noun-only and non-directional data model is sufficient to describe a useful concept map. Next, the proximity data (adjacency data) of the benchmark was created.

Table 3.5

Correlations Matrix among the Four Types of Data of the Selected Samples

	Reference Model			
	ND	NN	PD	PN
ND	0	0	0	0
NN	0.516	0	0	0
PD	0.985	0.512	0	0
PN	0.508	0.985	0.515	0
	Student 20			
	ND	NN	PD	PN
ND	0	0	0	0
NN	0.5	0	0	0
PD	0.703	0.413	0	0
PN	0.351	0.703	0.5	0
	Student 78			
	ND	NN	PD	PN
ND	0	0	0	0
NN	0.52	0	0	0
PD	0.895	0.49	0	0
PN	0.464	0.891	0.518	0
	Student 83			
	ND	NN	PD	PN
ND	0	0	0	0
NN	0.531	0	0	0
PD	0.895	0.522	0	0
PN	0.486	0.914	0.543	0

Note. The ND denotes the Noun-Only & Directional data; the NN is the Noun-Only & Non-Directional data; the PD indicates the Pronoun-Edited & Directional data; and the PN means the Pronoun-Edited & Non-Directional data.

1st Review: the Proximity Data

The proximity data obtained from the benchmark, ALA-Reader, and T-MITOCAR were compared. As Table 3.6 shows, the benchmark model had the greatest number of concepts and relations followed by those of T-MITOCAR and ALA-Reader. Interestingly, in the case of ALA-Reader, one and three cases both of which are expert' responses, deviated from the range of the concept and relation, respectively (see Figure 3.3). This result implied that ALA-Reader is more sensitive to the assessment context in terms of shared terms.

Table 3.6

The Numbers of Concepts and Relations of the Benchmark Model, ALA-Reader, and T-MITOCAR

	Concept			Relation		
	B	A	T	B	A	T
Min	16	0	8	18	0	7
Max	54	23	19	64	35	46
Average	33	5	13	39	6	23

Note. Note. Sample size *N* = 28. The B denotes the benchmark model; the A is the ALA-Reader; and the T means the T-MITOCAR

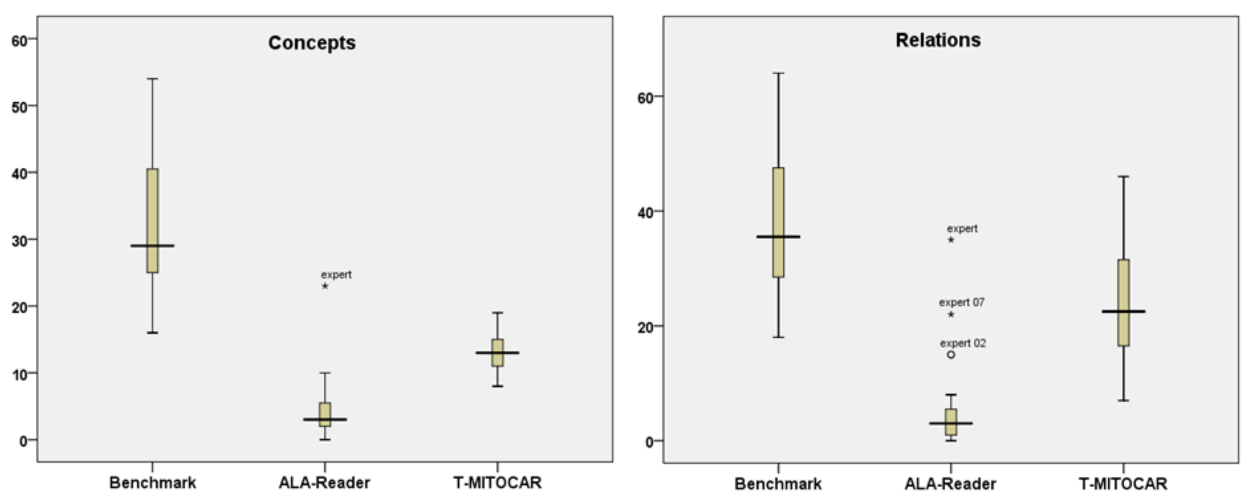


Figure 3.3. Boxplots of the numbers of concepts and relations.

As can be seen, the numerical similarities of concepts and relations revealed that ALA-Reader and T-MITOCAR provided less information in terms of the number of concepts and relations (see Table 3.7).

Table 3.7

Similarities of the Numbers of Concepts and Relations of ALA-Reader or T-MITOCAR with those of the Benchmark

	ALA-Reader		T-MITOCAR	
	Concept	Relation	Concept	Relation
Min	0.000	0.000	0.200	0.179
Max	0.426	0.547	0.773	0.969
Average	0.126	0.117	0.432	0.584

Note. Sample size N = 28.

Table 3.8

Correlations among Concepts and Relations in the Benchmark, ALA-Reader, and T-MITOCAR and Word Count of Each Sample

	WC	BC	BR	AC	AR	TC	TR
WC	-						
BC	0.526**	-					
BR	0.572**	0.949**	-				
AC	0.224	0.730**	0.735**	-			
AR	0.365	0.701**	0.746**	0.966**	-		
TC	-0.044	0.174	0.206	0.099	0.051	-	
TR	-0.014	0.033	0.130	-0.033	-0.009	0.876**	-

Note. Sample size N = 28. ** $p < .01$.
WC (Word Count); BC (Benchmark Concept); BR (Benchmark Relation); AC (ALA-Reader Concept); AR (ALA-Reader Relation); TC (T-MITOCAR Concept); and TR (T-MITOCAR Relation).

To examine the associations among the methods, correlation-coefficient analyses were conducted using word count, concepts, and relations. The benchmark and ALA-Reader had high correlations in their concepts and relations, r = 0.730 and 0.746, $p < .01$, respectively (see Table 3.8). In spite of a large difference in the numbers of concepts and relations, the ALA-Reader model was highly associated with the benchmark.

It was assumed that the larger volume of text response in general represents more concepts and relations. A reliable tool should be sensitive to the volume of text representation. Although the ALA-Reader and T-MITOCAR limit the number of concepts to no more than 30, it was expected that the numbers of distilled concepts and relations of the tools are to some extent associated with the volume of words in responses. Therefore, correlations with word count were investigated. The results showed that only the benchmark model had associations with word count in terms of concepts and relations, $r = 0.526$ and 0.572, $p < .01$, respectively (see Table 3.8). Subsequent regression analyses demonstrated that word count explained a significant proportion of the variance in the number of concepts and relations in the benchmark model, $R^2 = 0.25$, $F(1, 26) = 9.92$, $p < .01$ and $R^2 = 0.30$, $F(1, 26) = 12.62$, $p < .01$, respectively (see Tables 3.9 and 3.10).

Table 3.9

Simple Regression Analyses Investigating Linear Associations between Word Count and Concept

	Benchmark			ALA-Reader			T-MITOCAR		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Word Count	4.77	1.52	.53*	4.55	3.89	.22	-1.27	5.64	-.04
R^2		.25			.05			.00	
F		9.92*			1.37			.051	

* $p < .05$. ** $p < .01$.

Table 3.10

Simple Regression Analyses Investigating Linear Associations between Word Count and Relation

	Benchmark			ALA-Reader			T-MITOCAR		
	B	*SE B*	β	*B*	*SE B*	β	*B*	*SE B*	β
Word Count	4.20	1.18	.57*	4.36	2.18	.37	-.13	1.75	-.01
R^2		.30			.13			.00	
F		12.62**			3.99			.00	

* $p < .05$. ** $p < .01$.

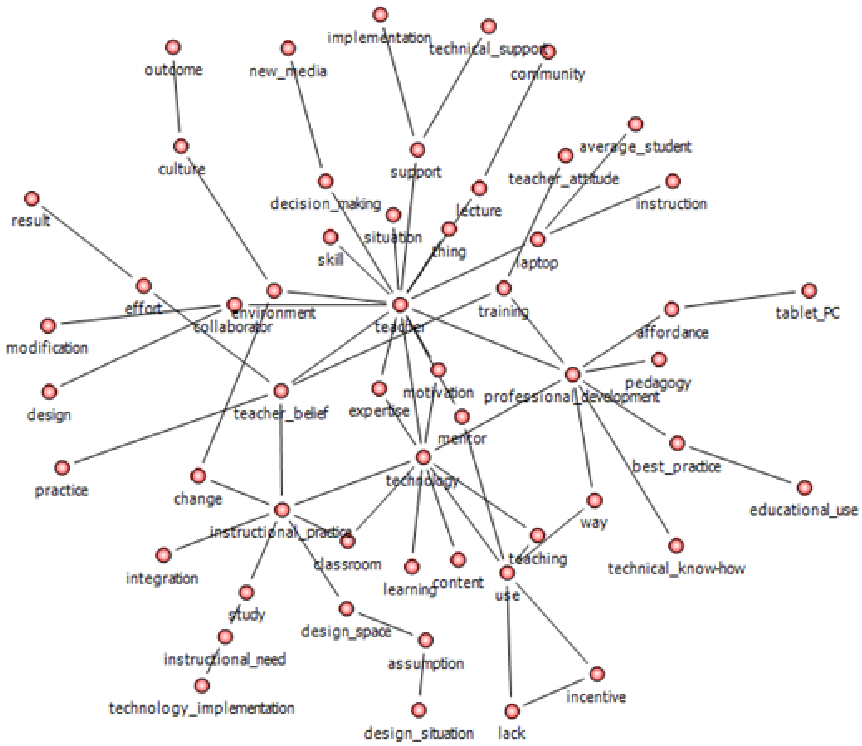


Figure 3.4a. Concept maps of the reference model created by the benchmark

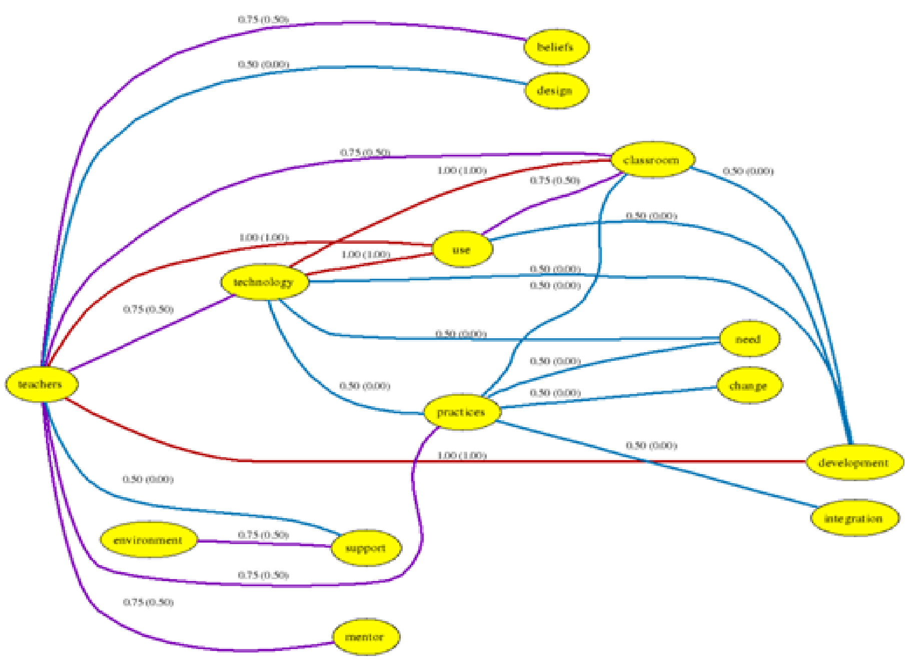


Figure 3.4b. Concept maps of the reference model created by T-MITOCAR

Figure 3.4c. Concept maps of the reference model created by ALA-Reader

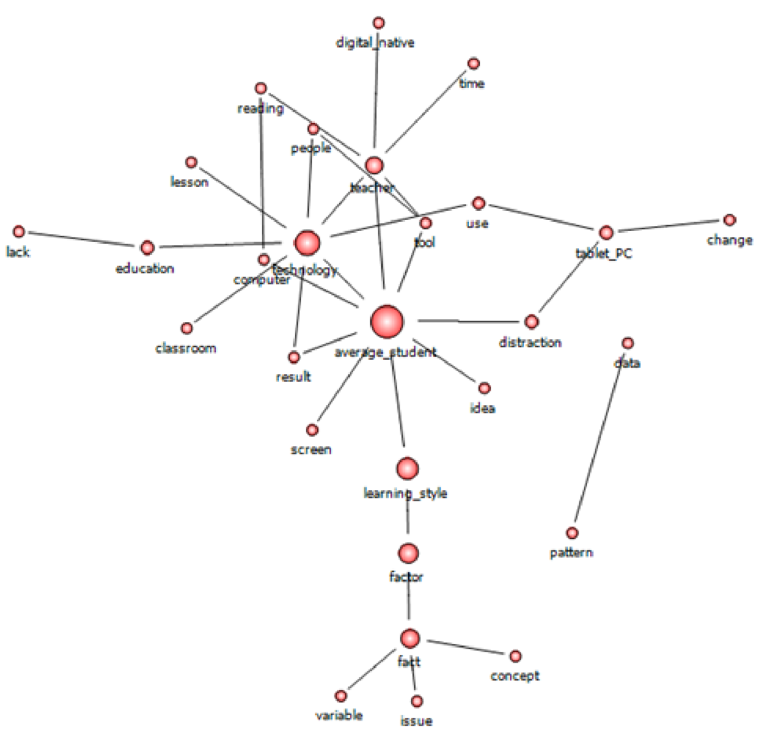


Figure 3.5a. Concept maps of the student 83 created by the benchmark

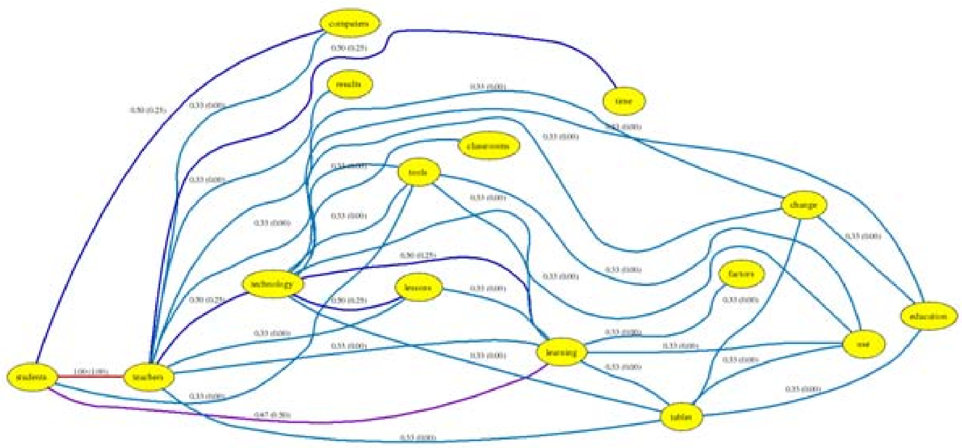


Figure 3.5b. Concept maps of the student 83 created by T-MITOCAR

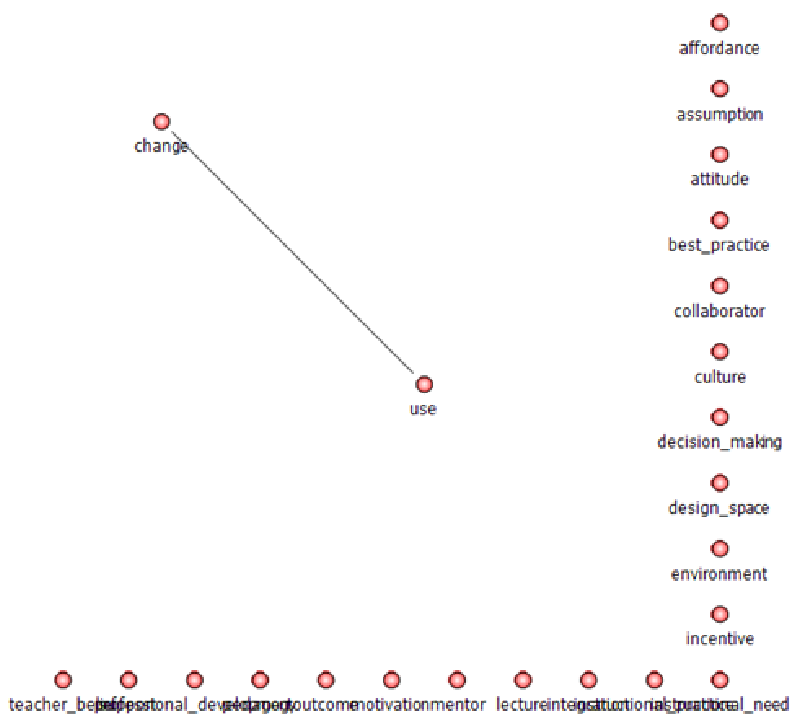


Figure 3.5c. Concept maps of the student 83 created by ALA-Reader

2nd Review: the Concept Maps

For the reference model and student 83, concept maps drawn from the competing methods were visually investigated. Regarding the reference, although the benchmark created a more cohesive and informative concept map (see Figure 3.4a), for two reasons, all other concept maps were highly connected (see Figure 3.4b and 3.4c): (a) the reference response was written carefully in a cohesive manner; and (b) T-MITOCAR and ALA-Reader let all elements of the concept maps connect technically, assuming all concepts are linked in the mind.

In contrast, the concept maps of student 83 substantially differed from one another. The benchmark differentiated the reference model from the concept map of student 83 (see Figure 3.5a). In contrast, T-MITOCAR yielded a concept map (see Figure 3.5b) more complex than that of the reference model, while the concept map produced by ALA-Reader was simple, capturing only two concepts (see Figure 3.5c).

3rd Review: Concept Map Parameters

Another numerical similarity and correlation study was performed regarding the structural parameters of concept maps: (a) diameter; (b) gamma (cluster coefficient); and (c) cohesive subgroups. In reference to Table 3.11, the benchmark model produced the largest and the most complex structure (average diameter = 7.04 and subgroup = 14.82). The T-MICOCAR model had an average diameter and subgroup value similar to those of the benchmark (average diameter = 6.00 and subgroup = 13.07). In contrast, the highest average gamma was identified in the ALA-Reader model, which was probably affected by the smaller structure (gamma = 0.34 and subgroup = 1.61). According to Table 3.12, parameter values of the T-MITOCAR model were much closer to those of the benchmark than those of the ALA-Reader model.

Table 3.11

Descriptive Statistics of Structural Parameters of Concept Maps

	Min	Max	Mean	SD
Benchmark				
Gamma	0.00	0.32	0.18	0.09
Diameter	5.00	12.00	7.04	1.60
Subgroup	6.00	29.00	14.82	5.70
ALA-Reader				
Gamma	0.00	1.0	0.34	0.43
Diameter	0.00	7.00	2.04	0.10
Subgroup	0.00	10.00	1.61	2.20
T-MITOCAR				
Gamma	0.12	0.59	0.25	0.10
Diameter	3.00	9.00	6.00	1.44
Subgroup	8.00	19.00	13.07	3.21

Note. Sample size *N* = 28.

Table 3.12

Numerical Similarities of Structural Parameters with those of the Benchmark Model

	ALA-Reader			T-MITOCAR		
	Min	Max	Average	Min	Max	Average
Gamma	0.00	0.80	0.15	0.00	0.98	0.58
Diameter	0.00	0.86	0.28	0.33	1.00	0.81
Subgroups	0.00	0.39	0.09	0.50	1.00	0.73

Note. Sample size *N* = 28.

Correlation analyses were conducted (see Table 3.13). In contrast to concepts and relations, most structural parameters had no relation with the word count. This result implied that the structure of a concept map has features distinctive from the frequencies of concepts and relations. The gamma score was in a negative relationship with the diameter and subgroup scores across all models. The gamma score did not appear simple to interpret in an assessment situation. The diameter and subgroup parameters were closely correlated across the benchmark, ALA-Reader, and T-MITOCAR.

Similar to concepts and relations, in spite of the ALA-Reader model's lower numerical similarity values, its structural parameters correlated more highly with the benchmark than did those of the T-MITOCAR. For example, BC was correlated with AC, $r = 0.630$, $p < .01$, while the r-value of TS was 0.419, with a $p < .01$.

Table 3.13

Correlations among Structural Parameters of the Benchmark, ALA-Reader, and T-MITOCAR Including Word Count

	WC	BG	BM	BC	AG	AM	AC	TG	TM	TS
WC	-									
BG	.020	-								
BM	-.026	-.300	-							
BC	.363	-.126	.648**	-						
AG	.386*	.442*	-.215	.065	-					
AM	.054	-.079	.493**	.626**	-.026	-				
AC	.076	.005	.331	.630**	.092	.864**	-			
TG	.003	-.041	.078	-.034	.147	-.020	-.022	-		
TM	.074	-.029	.209	.334	-.165	.063	.094	-.623**	-	
TS	-.044	-.317	.375*	.410*	-.397*	.210	.072	-.445*	.617**	-

Note. Sample size $N = 28$. * $p < .05$. ** $p < .01$.
WC (Word Count); BG (Benchmark Gamma); BM (Benchmark Diameter); BC (Benchmark Cohesive subgroups); AG (ALA-Reader Gamma); AM (ALA-Reader Diameter); AC (ALA-Reader Cohesive subgroups); TG (T-MITOCAR Gamma); TM (T-MITOCAR Diameter); and TS (T-MITOCAR Structure Measure).

4th Review: Concept Map Scores (Similarity Measures)

The final review was of six concept map scores obtained by measuring similarities between the student model and the reference model. Table 3.14 summarized the descriptive statistics of six measures. The average surface, graphical, and gamma scores were above 0.5 in the benchmark and T-MITOCAR, while the other three (concept, proposition, and balance) had low similarities, ranging from 0.061 to 0.284. That is, overall, the samples were above the half levels of the reference in terms of surface, graphical, and gamma scores but were not in concept, proposition, and balance score. All scores of the ALA-Reader were very low except for the balance measure (m = 0.635). Those scores resulted from the very small concept map sizes, in particular many of the student concept maps, affected by the constraint on analyzed concepts.

Table 3.14

Descriptive Statistics of Similarity Measures of Concept Maps

	Min	Max	Mean	SD
Benchmark				
Surface	0.281	0.984	0.590	0.184
Graphical	0.714	1.000	0.921	0.107
Gamma	0.000	0.942	0.644	0.245
Concept	0.119	0.570	0.284	0.100
Proposition	0.000	0.277	0.073	0.067
Balance	0.000	0.529	0.222	0.151
ALA-Reader				
Surface	0.029	0.629	0.142	0.151
Graphical	0.167	1.000	0.347	0.208
Gamma	0.000	0.777	0.234	0.286
Concept	0.087	0.435	0.190	0.112
Proposition	0.000	0.425	0.127	0.116
Balance	0.000	1.292	0.635	0.466
T-MITOCAR				
Surface	0.318	1.000	0.707	0.188
Graphical	0.500	1.000	0.835	0.138
Gamma	0.440	0.965	0.751	0.155
Concept	0.000	0.519	0.197	0.118
Proposition	0.000	0.360	0.061	0.087
Balance	0.000	0.697	0.223	0.214

Note. Sample size N = 28.

Similar to the earlier investigations, the scores of T-MITOCAR had a high numerical similarity of at least more than 0.7 with those of the benchmark in terms of surface, graphical, and gamma (see Table 3.15). As for the concept, proposition, and balance scores, their numerical similarities were moderate, ranging from 0.409 and 0.626. In contrast, the scores of the ALA-Reader had a low similarity, ranging from 0.018 to 0.332, with the exception of concept, which was 0.571. When the ranges of distributions were reviewed, the ALA-Reader and T-MITOCAR models appeared similar. A majority of the samples yielded relatively small-sized concept maps when processed via the ALA-Reader.

Table 3.15

Numerical Similarities of the Similarity Measures between the Benchmark and ALA-Reader or T-MITOCAR

	ALA-Reader			T-MITOCAR		
	Min	Max	Average	Min	Max	Average
Surface	0.004	0.639	0.018	0.345	0.973	0.721
Graphical	0.009	1.000	0.332	0.500	1.000	0.840
Gamma	0.000	0.967	0.293	0.000	0.994	0.738
Concept	0.003	0.991	0.571	0.000	0.993	0.626
Proposition	0.000	0.903	0.286	0.000	0.900	0.409
Balance	0.000	0.937	0.218	0.000	0.990	0.443

Note. Sample size *N* = 28.

Table 3.16

Ranks of the seven experts in the Benchmark Data

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7
Surface	8 (5, 8)	4 (2, 4)	5 (2, 1)	18 (7,20)	3 (6,21)	2 (4,16)	1 (1,2)
Graphical	1 (1, 1)	1 (3, 8)	1 (3, 8)	1 (3,26)	1 (3,21)	1 (2,21)	1 (3,1)
Gamma	21 (12, 9)	3 (3, 23)	17 (2, 6)	8 (12,16)	25(12,17)	6 (1,13)	11 (4,10)
Concept	3 (4, 17)	2 (1, 6)	1 (3, 1)	8 (7,14)	4 (6,12)	6 (5,21)	4 (1,1)
Proposition	5 (11, 10)	7 (3, 6)	1 (2, 1)	2 (4,8)	2 (19,16)	6 (12,18)	2 (1,2)
Balance	5 (18, 8)	11 (11, 5)	2 (9, 2)	1 (7,6)	3 (19,16)	7 (17,18)	3 (8,3)

Note. Sample size *N* = 28. In the parentheses, the first is the rank in the ALA-Reader data and the second is the rank in the T-MITOCAR.

To investigate the measurement accuracy of the three approaches, all 27 samples were ranked according to a single concept map score for the individual samples. Expert responses were expected to be ranked at the top. Overall, in the benchmark model, expert responses fell into the upper ranks of the list. The ranks of the ALA-Reader and T-MITOCAR varied but the ALA-Reader provided rankings closer to those of the benchmark than T-MITOCAR. There was no pattern in the gamma ranks of all three models.

Table 3.17

Correlations of the Concept Map Scores between the Benchmark and ALA-Reader or between the Benchmark and T-MITOCAR

	ALA-Reader	T-MITOCAR	ALA-Reader and T-MITOCAR[a]
Benchmark			
Surface	.724**	.093	.416*
Graphical	.412*	-.222	.057
Gamma	.277	-.145	-.096
Concept	.815**	.696**	.520**
Proposition	.555**	.654**	.634**
Balance	-.048	.489**	.025

Note. Sample size $N = 28$. * $p < .05$. ** $p < .01$. Pearson r was applied to the similarity measures
a. The correlations of each measure between ALA-Reader and T-MITOCAR.

Despite the lack of numerical similarity, the surface and graphical scores had a significant correlation only between the benchmark and ALA-Reader, $r = 0.724$ and 0.412, $p < .05$ (see Table 3.17). The correlations of concept and propositional scores were significant overall across the three approaches. The balance score had only a moderate association between the benchmark and T-MITOCAR, $r = 0.489$, $p < .01$. On the whole, the ALA-Reader generated concept map scores better associated with those of the benchmark than the T-MITOCAR, while the T-MITOCAR had a better association in terms of proposition and balance scores (see Table 3.17).

Lastly, in regard to the benchmark, correlations of its concept map scores were examined (see Table 3.18). The surface score was moderately correlated with the concept and proposition score, $r = 0.583$ and 0.512, p < .05, respectively. In contrast, no significant correlation was identified with structural parameters such as the graphical and gamma scores. The balance score was highly associated with the propositional score, *r* = 0.924, p < .01. Those results demonstrated that concept map scores account for different features of concept maps.

Table 3.18

Correlations of the Similarity Measures of Concept Maps in the Benchmark

	1	2	3	4	5	6
1. Surface	-					
2. Graphical	.338	-				
3. Gamma	-.233	-.278	-			
4. Concept	.583**	.093	-.034	-		
5. Proposition	.512*	.272	-.013	.870**	-	
6. Balance	.400*	.264	-.027	.723**	.924**	-

Note. Sample size $N = 28$. * $p < .05$. ** $p < .01$. Pearson Correlation Coefficients in the lower diagonal.

Conclusion

Research Findings

This study assumed that an individual student's understanding is meaningfully elicited via a natural language approach. Two state-of-the-art technologies, ALA-Reader and T-MITOCAR, were selected because they were consistent with the initial assumption. In order to validate the technologies, an alternative method was established as a benchmark.

It was believed that linguistic knowledge representation should be open-ended in terms of concepts and should be directional in terms of relations. In addition, it was assumed that editing pronouns in text responses helps obtain more accurate and descriptive concept maps. Creating a benchmark drew on distilling semantic relations from responses because eliciting linguistic semantic structure was assumed to be a better way to visually represent concept maps.

The noun-only data was not significantly different from the pronoun-edited data, which was aligned with Clariana and colleagues' (2007, 2009) suggestion. In addition, there was little difference between the directional and non-directional approaches Thus, it was concluded that a simple noun-only and non-directional approach was sufficient for creating a benchmark.

Two findings are summarized through the numerical similarity and correlation analyses amongst the benchmark, ALA-Reader, and T-MITOCAR:

- The benchmark model created the most descriptive concept maps in terms of the number of concepts and relations, followed by the T-MITOCAR and ALA-Reader.
- The ALA-Reader model had smaller numerical similarities, although it yielded higher correlations with the benchmark model than the T-MITOCAR model in terms of proximity data, concept map parameters, and concept map scores.

Those results are probably affected by the constraints introduced by each technology. The ALA-Reader and T-MITOCAR can only analyze at most 30 concepts in a concept map. Moreover, ALA-Reader pre-defines a particular set of terms to be used in text analyses. Accordingly, the ALA-Reader yielded the smallest concept maps. The higher correlation between the ALA-Reader and benchmark model is in part explained by their methodological similarity. That is, both use adjacency rather than distance data.

When considering the sharp drop in the numbers of concepts in student samples, ALA-Reader will be more appropriately applied in a setting in which a set of key terms are intentionally introduced or adequately exposed before and during assessment activities, which may yield a closer association with the benchmark model than occurred in this study. T-MITOCAR is only applicable to cases having more than 350 words. Accordingly, the technology seems on the whole adaptable to data reduction utilizing a large volume of text. Although T-MITOCAR yielded data numerically more similar to those of the benchmark, their associations with the benchmark were somewhat lower than the ALA-Reader. This result requires further studies in regard to the association between adjacency data (the benchmark) and distance data (T-MITOCAR).

New Opportunities

The benchmark model was explored initially, and concept maps were processed manually. In spite of those limitations, the benchmark model provided some opportunities: First, the concept maps were much more descriptive than those of the other two models. When the purpose of assessment is to provide formative feedback and instructional supports, descriptive information of students' status is essential.

Second, the benchmark model was able to more capably distinguish better concept maps from maps of lesser quality. For example, expert responses were accurately identified and ranked at the top, which to a large degree resulted from the modified similarity formula suggested in this study.

Third, the benchmark model has no constraints on the number of words and used terms. Thus, this approach can deal with diverse assessment contexts. For example, when gathering data, the instructor clearly asked the students to write responses of more than 350 words. However, only one third of the students met the requirement. It is natural that diverse levels of students in a classroom provide diverse volume of responses. Accordingly, an assessment technology should cover all responses.

Suggestions

In regard to further development of the benchmark model, there are two suggestions for future studies: First, the methods for drawing semantic relations from written responses need to be more specific and algorithmic. Second, a set of combined rules is required to assess the progress of student learning based on multiple concept map scores. For example, the six concept map scores were not often at the same level. They were not correlated in some of the pairs because concept map scores account for the different features of concept maps. Lastly, developing an automated assessment technology embedding the benchmark model is required. That technology would enable a teacher to have a better sense of students' learning and provide them with elaborated feedback and support.

References

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton, N.J.: Princeton University Press.

Brown, A. N. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Science*, *2*(2), 141-178.

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces*, *70*, 215-225.

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clariana, R. B. (2010). Multi-decision Approaches for Eliciting Knowledge Structure. In *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 41-59). Retrieved from http://dx.doi.org/10.1007/978-1-4419-5662-0_4

Clariana, R. B., & Wallace, P. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research*, *37*(3), 211-227. doi:10.2190/EC.37.3.a

Clariana, R. B., & Wallace, P. (2009). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge in information systems courses. *International Journal of Instructional Media*, *36*(3), 287-302. doi:Article

Clariana, R., Wallace, P., & Godshalk, V. (2009). Deriving and measuring group knowledge structure from essays: The effects of anaphoric reference. *Educational Technology Research and Development*, *57*(6), 725-737. doi:10.1007/s11423-009-9115-z

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

Colman, A. M., & Shafir, E. (2008). Tversky, Amos. In N. Koertge (Ed*.), New dictionary of scientific biography* (Vol. 7, pp. 91-97). Farmington Hills, MI: Charles Scribner's Sons.

Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural Comparison of Cognitive Associative Networks in Two Populations. *Journal of Applied Social Psychology*, *37*(9), 2097-2129. doi:10.1111/j.1559-1816.2007.00253.x

Curtis, M. B., & Davis, M. A. (2003). Assessing knowledge structure in accounting education: an application of Pathfinder Associative Networks. *Journal of Accounting Education*, *21*(3), 185-195. doi:10.1016/S0748-5751(03)00024-1

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, *13*(5), 533-568. doi:10.1016/S0959-4752(02)00025-7

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chicheser: Ellis Horwood Ltd.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.

Gijbels, D., Dochy, F., den Bossche, P. V., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*(1), 27-61.

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

Goldsmith, T. E., & Kraiger, K. (1997). Applications of structural knowledge assessment to training evaluation. In J. K. Ford (Ed.), I*mproving training effectiveness in organizations* (pp. 73–95). Mahwah, NJ: Lawrence Erlbaum Associates.

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, *12*(6), 729-734.

Greeno, J. G. (1989). Situations, mental models, and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285–318). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Grotzer. T. A., & Perkins, D. N. (2000, April). *A taxonomy of causal models: The conceptual leaps between models and students' reflections on them*. Paper presented at the National Association of Research in Science Teaching (NARST), New Orleans, LA, 28 April - 1 May 2000.

Hage, P., & Harary, F. (1983). *Structural models in anthropology*. Cambridge: Cambridge University Press.

Hsu, Chia-Chien & Sandford, Brian A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation, 12*(10). Available online: http://pareonline.net/getvn.asp?v=12&n=10

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A micro-analytic review. *Psychonomic Bulletin and Review*, *10*, 785–813.

Ifenthaler, D. (2006). *Diagnose lernabhängiger Veränderung mentaler Modelle. Entwicklung der SMD-Technologie als methodologische Verfahren zur relationalen, strukturellen und semantischen Analyse individueller Modellkonstruktionen*. Freiburg: Universitäts-Dissertation.

Ifenthaler, D. (2007, October). Relational, structural, and semantic analysis of graphical *representations and concept maps. Introducing the SMD-Technology*. Paper presented at the Annual Meeting of the Association for Educational Communications and Technology, Anaheim, CA.

Ifenthaler, D., Masduki, I., & Seel, N. M. (2009). The mystery of cognitive structure and how we can detect it: tracking the development of cognitive structures over time. *Instructional Science*. doi:10.1007/s11251-009-9097-6

Jeong, A. (2008). *jMap*. Retrieved May 5, 2010, from http://garnet.fsu.edu/~ajeong

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185-208). Cambridge University Press.

Johnson-Laird, P.N. (2005b). The history of mental models. In K. I. Manktelow, & M. C. Chung (Eds), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179-212). Psychology Press.

Johnson, T. E., O'Connor, D. L., Spector, J. M., Ifenthaler, D., & Pirnay-Dummer, P. (2006). Comparative study of mental model research methods: Relationships among ACSMM, SMD, MITOCAR & DEEP. In J. Canas & J. D. Novak (Eds.), (pp. 87-94). San Jose, Costarica: Universidad de Costa Rica.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.*

Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology*, *20*, 333–342. doi:10.1002/acp.1249

Author (2011). ----

Knoke, D., & Kuklinski, J. H. (1982). *Network analysis*. Newbury Park: Sage.

McClure, J., Sonak, B., & Suen, H. (1999). Concept map assessment of classroom learning: reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, *36*, 475–492.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.

Narayanan, V. K. (2005). Causal mapping: An historical overview. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for research in information technology* (pp.1-19). Hershey, PA: Idea Group Publishing.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Novak, J. D., & Cañas, A. J. (2006). *The origins of the concept mapping tool and the continuing evolution of the tool*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.3382

O'Connor, D. L., & Johnson, T. E. (2004). *Measuring team cognition: Concept mapping elicitation as a means of constructing team shared mental models in an applied setting*. First International Conference on Concept Mapping, September 14-17, 2004, Pamplona, Spain.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15-29.

doi:10.1016/j.im.2003.11.002

Phelan, J., Kang, T., Niemi, D. N., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessments in middle school mathematics*. (CRESST Report 750). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Pirnay-Dummer, P. (2006). *Expertise und Modellbildung – MITOCAR*. Freiburg: Universitäts-Dissertation.

Pirnay-Dummer, P. (2007). *Model inspection trace of concepts and relations. A heuristic approach to language-oriented model assessment*. Paper presented at the AREA 2007, Chicago, IL.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. (2010). Highly integrated model assessment technology and tools.*Educational Technology Research & Development, 58*(1), 3-18. doi:10.1007/s11423-009-9119-8.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216–270). Cambridge, MA: MIT Press.

Ruiz-Primo, M. A., Schultz, S., Li, M., & Shavelson, R. J. (1999). *On the cognitive validity of interpretations of scores from alternative concept mapping techniques.* (CSE Tech. Rep. No. 503). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010a). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment, 8*(4). Available online at http://escholarship.bc.edu/jtla/vol8/4

Rupp, A. A., Sweet, S., & Choi, Y. (2010b). *Modeling learning trajectories with epistemic network analysis: A simulation-based investigation of a novel analytic method for epistemic games*. Presented at the annual meeting of the International Society for Educational Data Mining (EDM), Pittsburgh, PA.

Schlomske, N., & Pirnay-Dummer, P. (2008). Model based assessment of learning dependent change during a two semester class. In Kinshuk, Sampson, D., & Spector, M. (Eds.), *Proceedings of IADIS International Conference Cognition and Exploratory Learning in Digital Age 2008*, Freiburg, Germany, pp. 45-53.

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

Schvaneveldt, R.W., Durso, F.T., Goldsmith, T.E., Breen, T.J., & Cooke, N.M. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies*, *23*, 699-728.

Schvanevldt, 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* (pp. 249-284). San Diego, CA: Academic Press.

Seel, N. M. (1999). Educational diagnosis of mental models: Assessment problems and technology-based solutions. *Journal of Structural Learning and Intelligent Systems, 14*(2), 153–185.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science*, *29* (4-5), 403-427.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation*, *23*(4), 373-398.

Shaffer, Hatfield, Svarovsky, Nash, Nulty, & Bagley et al. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, *1*(2), 1-21.

Shute, V. J., Jeong, A. C., Spector, J. M., Seel, N. M., & Johnson, T. E. (2009). Model-based methods for assessment, learning, and instruction: innovative educational technology at Florida State University. In M. Orey, R. M. Branch, & V. J. McClendon (Eds.), *Educational Media and Technology Yearbook* (pp. 61-80). New York: Springer.

Shute, V. J., & Zapata-Rivera, D. (2008). Using an evidence-based approach to assess mental models. In D. Ifenthaler, P. Pirnay-Dummer, J. M., Spector (Eds.), *Understanding Models for Learning and Instruction: essays in honor of Norbert M. Seel* (pp. 23-41). New York: Springer.

Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2006). Causal maps, mental models and assessing acquisition of expertise. *Technology, Instruction, Cognition and Learning, 3*(2), 167–183.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Taricani, E. M., & Clariana, R. B. (2006). A Technique for Automatically Scoring Open-Ended Concept Maps. *Educational Technology Research & Development*, *54*(1), 65-82.

doi:Article

Thomas, R. M. (2005). *High-stakes testing: coping with collateral damage*. Mahwah, NJ: Psychology Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Tversky, A., & Shafir, E. (2004). *Preference, belief, and similarity: selected writings*. London, England: MIT Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge University Press.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477-501.

CHAPTER 4

DEVELOPMENT OF AN ASSESSMENT TECHNOLOGY FOR MEASURING KNOWLEDGE STRUCTURES USING NATURAL LANGUAGE RESPONSE TO A COMPLEX PROBLEM SCENARIO[4]

[4] Kim, M. Submitted to *Contemporary Educational Psychology*, 04/07/2012.

Abstract

There is strong interest in and emphasis on learner preconceptions in learning and instruction (Bransford, Brown, & Cocking, 2000; Spector, 2004). This paper reports an exploratory study using natural language responses to a complex problem scenario as the basis for generating and analyzing problem conceptualizations. In particular, this study focused on the potential uses of descriptive concept maps for formative assessment and feedback. Based on a review of research and theory on cognition and linguistic comprehension, this study proposed the semantic relation (SR) approach to explore the salient role of the semantic relations among concepts extracted from natural language responses to problem scenarios. A set of methodologies was introduced to distill concepts and relations and to quantify the attributes in the form of concept maps for both learner and expert responses. A comparison with other approaches suggests that the SR approach is consistent with theories of cognition and provides an effective basis for formative feedback.

Keywords: assessment technology, concept map, knowledge structure, natural language processing, semantic relation

Teachers need a way to precisely but efficiently assess student application of understanding and complex problem solving. A simple knowledge-based (e.g., multiple choice, short answer, etc.) test is not sufficient to test a learner's ability to solve complex problems (Champagne, Kouba & Hurley, 2000; Duschl, 2003; Quellmalz & Haertel, 2004), yet it is challenging to assess cognitive models of complex problem solving knowledge and skills such as scientific inquiry. The goal of this study was to (a) explore concept map technologies used to assess complex problem solving and (b) propose a new approach aimed at developing a real-time representation of a student's understanding that can serve as the basis for meaningful formative feedback.

Concept map techniques have been used to represent a student's understanding, and a commonly used technique is to portray the propositional relations among concepts found in a body of text such as a student essay or response to a problem situation (Clariana, 2010; Jonassen, Beissner, & Yacci, 1993; Narayanan, 2005; Novak & Canãs, 2006; Spector & Koszalka, 2004). This study is based on the assumption that a concept map re-representing a student's cognitive processes can reflect different states in a student's learning progress.

Language plays a critical role in building and mediating an individual's internal representations of the external world (Wittgenstein, 1922). As a consequence, a second assumption is that using natural language as the basis for constructing concept maps is likely to be closer in meaning and structure to the targeted internal mental models (Pirnay-Dummer, Ifenthaler, & Spector, 2010). Detecting the qualities of internal mental models as accurately as possible should help to provide more meaningful and productive instructional feedback suited to individual learning needs (Phelan, Kang, Niemi, Vendlinski, & Choi, 2009; Shute & Zapata-Rivera, 2007; Yorke, 2003).

This exploratory study involved comparing a new approach utilizing natural language responses with two recognized technologies for assessing students' written responses to a problem: T-MITOCAR (Text Model Inspection Trace of Concepts and Relations; Pirnay-Dummer & Ifenthaler, 2010) and ALA-Reader (Analysis of Lexical Aggregates-Reader; Clariana & Wallace, 2007, Clariana, Wallace, & Godshalk, 2009).

First, this study began with classifying state-of-the-art technologies. Based on an analysis of the state-of-the-art in concept map assessments, a new methodology was developed. Some concept map technologies do elicit natural language input, but it is rare to come across a study that illustrates and compares their underlying linguistic assumptions and analytical methods - in fact no such studies were found. Typically, a study focuses on just one methodology and describes the linguistic assumptions and analytical methods of that particular methodology. The lack of comparative studies makes progress in this area problematic.

Second, this study explored a new concept map methodology that can use any concepts and relations in a text elicited in response to a problem scenario. The methodology then can contrast the result with two other promising concept map assessment technologies that constrain the analytic process in various ways. More specifically, T-MITOCAR requires no fewer than 350 words and limits the number of analyzable concepts to 30 or fewer. ALA-Reader has an equal limitation on the number of concepts that are pre-defined by expert judgment. Those constraints on assessment methods probably have an impact on the resulting analysis, although whether the effect is significant has yet to be determined.

Lastly, this study assumes that concept maps believed to represent a student's mental models or cognitive process should closely reflect a student's linguistic responses to a problem situation. It is, consequently, imperative to assure cognitive fidelity between a concept map and a student's linguistic representation in order to obtain adequate descriptive and formative information about a student's mental model. In short, this study introduces a new technology to elicit a student's conceptual model through natural language (i.e., written text) used by a student in response to a problem scenario. The text is an initial re-representation of a student's beliefs and thinking about a problem, and the constructed concept map is a second representation of that mental model (see Figure 4.1). The reason to construct concept maps is that they can be more effectively and efficiently analyzed for purposes of diagnosing a student's cognitive model to a problem scenario.

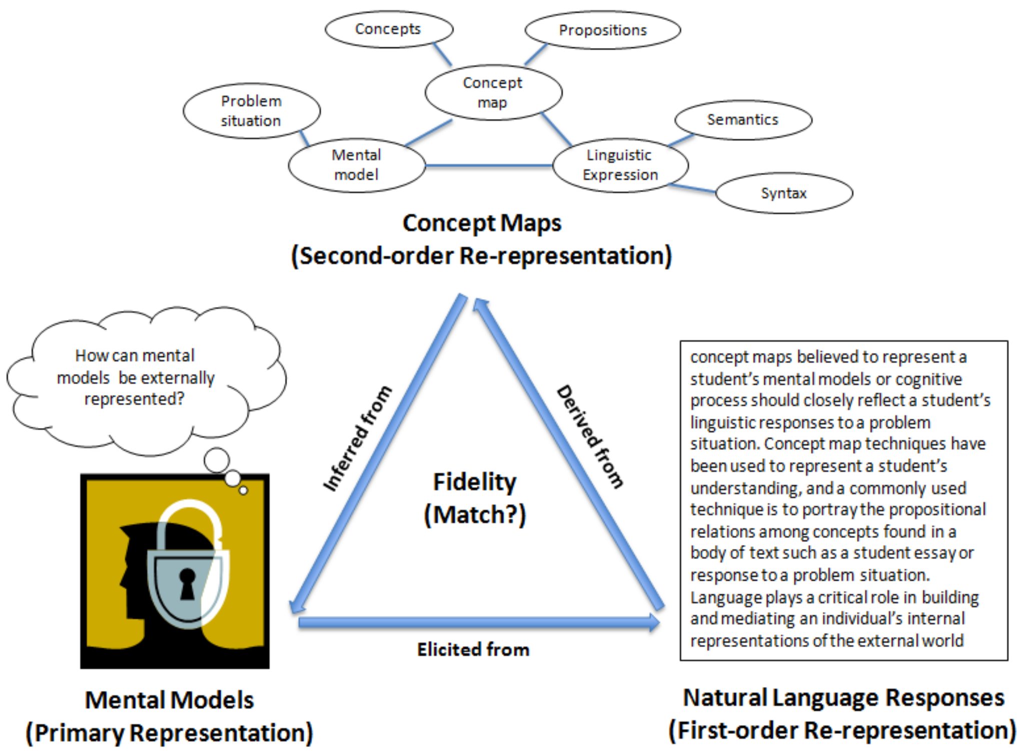


Figure 4.1. The Relations of Mental models, natural language responses, and concept maps.

Backgrounds

Mental Models, Natural Language Representations, and Concept Maps

The theory of mental models explains that a person builds his/her understanding by mentally representing certain aspects of external situations corresponding to his/her preconceptions (Johnson-Laird, 2005a, 2005b; Seel, 2001, 2003). There is a common belief that mental models are structural cognitive artifacts that contain symbols and their relations (Newell, 1990; Johnson-Laird, 2005a, 2005b). In that sense, the progress of mental models within an individual can be considered as the changing of knowledge structures toward an expected or desired state (Anzai & Yokoyama, 1984; Collins & Gentner, 1987; Johnson-Laird, 1983; Seel, 2001, 2003, 2004; Seel & Dinter, 1995; Smith, diSessa, & Roschelle, 1993; Snow, 1990).

Problem solving includes conceptualizing a problem space as a more structured understanding and integration of many ideas and concepts related to a problem (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Jonassen et al., 1993; Newell & Simon, 1972; Segars, 1997; Spector & Koszalka, 2004). It is reasonable to say that a conceptualizaiton of a problem space is a kind of mental model of a certain problem situation.

A concept map as a structural knowledge representation consisting of concepts and relations (Clariana, 2010; Narayanan, 2005; Novak & Canãs, 2006; Spector & Koszalka, 2004), can be an effective tool to use in assessing a student's conceptualization of a problem space as a structural mental representation to a given problem. Concept maps have been used to elicit cognitive representations of an individual's knowledge of a domain in which concepts are interrelated (Funke, 1985; Narayanan, 2005; Novak & Canãs, 2006; Schvaneveldt, 1990). The data used for concept maps are generally collected from interviews or texts. Text-based data collection is economical in terms of time and effort (Brown, 1992), and is based on techniques that avoid recall bias and potentially leading or misleading questions (Axelrod, 1976).

Language is a symbol system, and mental models result from both perceptual and linguistic comprehension (Johnson-Laird, 2005a, 2005b; Seel, 2001). Symbols play a central role in representing ideas and thoughts (Seel, 1999); model building is essentially a process of symbolic notation and representation (Greeno, 1989). That is, mental models are partially or completely construed by situational inputs described in text, discourse, diagrams, and so on (Garnham, 1987, 2001). In contrast, a student's linguistic representations (external re-representations) reflect that student's mental models (internal primary representations). Since it is the internal representation that is directly linked to learning and understanding, it is important that any external representation used to make inferences about learning be as closely linked to internal representations as possible.

According to Miller and Johnson-Laird's (1976) study, the world people talk about is the world they perceive. In other words, expressions in language denote elements of a model, and the relations between those elements decide whether sentences are true or false (Garnham, 1987). Those elements and relations of a model in linguistic representations constitute concept maps representing a part of mental models. It is arguable that using natural language responses as the basis for concept maps are descriptive and likely to resemble the underlying mental model (Pirnay-Dummer et al., 2010).

A central issue is to explain how elements and relations of a model are identified from the linguistic expressions of a problem situation. Until the middle of the previous century, it was a common belief that thoughts represented logical pictures of facts (Wittgenstein, 1922). Accordingly, a mental model is represented by a sentence which adheres to the grammatical and syntactical rules of the relevant language (Chomsky, 1957; Wittgenstein, 1922). More recently, linguists have come to recognize that a sentence includes both surface and an underlying deep structure, and that denotation and connation are relevant to semantic analysis. According to Katz and Postal (1964), the surface structure (syntax) characterizes the shape of the sentence, while the semantic information of the deep structure account for a substantial part of the meaning (Bransford & Franks, 1971; Bransford, Barclay, & Franks, 1972; Bransford & Johnson, 1972; Fodor, Bever, & Garrett, 1974). This study assumes that mental models are effectively depicted in integrated semantic networks − deep structures − constituted by semantic relations − surface structures − that are presumed to be mostly nested within the syntactic structure of sentences.

Approaches to Elicit Concept Maps from Natural Language Expressions

Approaches to construct concept maps can be characterized by the ways to identify concepts and determine the *associatedness* of the relations among concepts. This study centers on the notion of propositions as units of meaning - the semantic contents of sentences that typically include two or more concepts with relations indicated by verbs (Cañas, 2009; McGrath, 2011). The following questions can serve as guides for classification of concept mapping approaches:

- What are the ways to distill concepts from a written response?
- How can key concepts be identified from a list of concept in a response?
- What are the ways to identify relations between concepts?
- How do we judge the strengths of the relations?

To address the above questions, this study groups concept-mapping methods into three classifications based on how propositional relations are treated. The three approaches are labeled Adjacent Relation (AR), Proximity Relation (PR), and Semantic Relation (SR). In particular, the

SR approach was designed as an alternative way to create more descriptive and complex concept models when natural language is used for learner input.

Table 4.1

Classification of Concept Mapping Approaches

	Adjacent Relation (AR)	Proximity Relation (PR)	Semantic Relation (SR)
Approach	Abstractive	Abstractive	Descriptive
Origin of Relation	Spatial Model	Spatial Model	Discrete Model
Relation Judgment	Adjacency	Distance	Syntactic/Semantic
Direction of Relation	Non-directional only	Non-directional only	Directional/Non-directional
Tool	ALA-Reader	T-MITOCAR	-

Adjacent Relation (AR). Adjacent Relation (AR) is a spatial model because AR takes a position on the associated nature of concepts (words), which is that concepts that are closely connected tend to be presented as physically closer within a text (see Table 4.1). This perspective believes verbs do not matter in a knowledge structure. Certainly, identifying concept-concept within a text is far easier than counting concept-verb-concept structures. Compared to the other two methods, AR employs the simplest method in which any two concepts adjacent to each other are regarded as associated (i.e., adjacency in Table 4.1).

ALA-Reader (Analysis of Lexical Aggregates-Reader) (Clariana & Koul, 2008) is a tool designed to capture knowledge structure according to the adjacent relations of concepts (nouns) in text. This technology identifies a predefined set of concepts (nouns) in a student's written response. In an AR analysis, it does not matter whether two concepts are located in the same sentence; two concepts adjacent to each other are considered to be associated. Consequently, the whole text is treated as one corpus without separations between sentences. Two adjacent concepts in a text are regarded as strongly associated with each other, while two concepts farther from each other are considered not to be directly associated. The former relationship has the strength value 1 and the latter is coded as 0 in the data (Clariana et al., 2009). Accordingly, although AR stems from the spatial model, its data take the form of a discrete model. AR defines the strengths of nodes and relations based on the whole structure. However, AR cannot have directional information of relations because it is not concerned with linking words (e.g., verbs).

ALA-Reader uses no more than 30 key terms for analysis. The restriction on the number of terms is partly affected by prior research arguing that more meaningful concept map structures are found when a smaller number of terms are used in knowledge structure analysis (Clariana & Taricani, 2010). ALA-Reader does not include a function to create a concept map. To construct a concept map, one must employ a network graph tool. In this study, NetMiner tool (http://www.netminer.com/) was used when creating the concept map.

Proximity Relation (PR). Although proximity is commonly used as a general term representing similarity, relatedness, and distance between concepts (Schvaneveldt, 1990), a proximity relation (PR) is an approach to measure the strength of relations according to the distance between concepts in a text (see Table 4.1). In the sense that relations are weighted by geometric distance between a pair of concepts, the PR approach is grounded on the spatial model (Schvanevldt et al, 1989). It is assumed in the spatial model that the more two concepts are associated with each other, the closer the concepts present within or across sentences (Pirnay-Dummer & Ifenthaler, 2010). PR measures the distance of each paired concepts from a text response and then creates a whole network using distance data. The spatial model assumes that all concepts in a written artifact are basically associated with one another and have different levels of distances. Since all possible pairs of concepts can occur in a network, it is necessary that a PR approach include a procedure to reduce information of a network in order to project meaningful networks. That is, PR is abstractive due to its selection of concepts and relations. In addition, due to the relations determined by spatial distance rather than semantic relation, PR lacks the directional information of relations.

T-MITOCAR (Text-Model Inspection Trace of Concepts and Relations) (Pirnay-Dummer & Ifenthaler, 2010) is chosen as a technology representing the Proximity Relation (PR) approach because it generates distance data to create concept maps using students' natural language responses. The algorithms of T-MITOCAR generate distance data constituted with proximity vectors. Detailed information of T-MITOCAR can be found in Pirnary-Dummer and Ifenthaler (2010). A brief introduction to the analytic methods of T-MITOCAR follows:

- Sentences in a text response are parsed into syntactic components. Among tokenized components, only nouns and names are distilled and listed. The current version of T-MITOCAR can deal with no more than 30 nouns in a text, and a text response requires more than 350 words.
- The frequency of each noun in a written response is calculated and the most frequent concepts (nouns), up to 30 nouns, are selected.
- The selected *n* concepts results in *n(n-1)/2* pairs. All pairs of concepts are considered possible. When a pair of concepts exists in a sentence, the distance of the pair is calculated with the minimum number of words between two concepts.
- The sum of distances of each pair determines the strength of association. Now T-MITOCAR obtains a distance data. Drawing on the distance data, a set of algorithms transforms the distance values into relative weights. Lastly, a concept map is created with weights for each pair of concepts.

Semantic Relation (SR). This study proposes a new Semantic Relation approach to distill concepts and relations from written responses so that one can obtain more detailed descriptive features of concept models. In this Semantic Relation (SR) approach, the atomic units of meaning in sentences are clearly indicated. According to Beamer and colleagues (2008), semantic relations are the underlying relations between two concepts expressed by words or phrases. SR is similar to the general meaning of proposition in the sense of focusing on the meaning of words, but SR involves diverse types of relations of concepts beyond the typical noun-verb-noun relation form (Adrian, Moldovan, Badulescu, Tatu, Antohe, & Girju, 2004; Cañas, 2009; Girju, Nakov, Nastase, Szpakowicz, Turney, & Yuret, 2009). The types of SR can include complex noun compounds (e.g., ‘knowledge analysis’), genitives (e.g., ‘teachers’ participation’), prepositional phrases attached to nouns (e.g., ‘community of practice’), or sentences (e.g., ‘Emerging new media has always led to instructional changes.’).

For instance, a proposition is generally considered as similar to a sentence or statement that is often used instead of a proposition (McGrath, 2011; “Proposition,” n.d.). In a sentence, the relation between concepts is defined by the verb (e.g., ‘Success of a community of practice is determined by individuals’ active engagement and contributions.’). This example consists of two concepts (i.e., ‘success of a community of practice’ and ‘individuals’ active engagement and contributions’) and a verb connection (i.e., ‘is determined by’).

When a constrained notion is applied to a proposition as ‘the relations of paired nouns’ (Girju et al., 2009), the example sentence is decomposed into a greater number of units of meaning (semantic relations). The following relations in the former example can be identified: (a) <u>success</u> of a <u>community</u>; (b) <u>community</u> of <u>practice</u>; (c) <u>individual</u>’s <u>active engagement</u>; and (d) <u>engagement</u> and <u>contributions</u>.

The SR approach allows concept maps to use whatever concepts and linking terms students use in domain learning without constraints on the number of concepts. The elaborated semantic relations help create descriptive and complex concept maps involving delays and feedback loops that would be represented in a causal influence diagram. As Table 4.1 summarizes, SR is descriptive in the sense that it includes all semantic relations found in student responses using both syntactic and semantic analysis. A discrete model accounts for SR relations since semantic relations are represented by binary values indicating whether there is an association between concepts. In addition, the meaning of relations is a core feature in the SR approach, which makes it possible to define the direction of each relation. For example, in a phrase, 'learning progression toward expert level,' the first concept (i.e., learning progression) is in the direction to the second concept (i.e., expert level). In the following sections, guidelines for extracting semantic relations from a text response are described and applied to the analysis of a few representative responses to a complex problem.

Introduction to the Semantic Relation (SR) Approach

What Should Be Measured Or Inferred?

This study agrees that a cognitive model is a possible configuration of the underlying structure (Norman, 1986). It is believed in this study that a mental model (conceptual structure) is probably not the same as the semantic structure of lexicons but embeds the properties of semantic representations (Bierwisch & Schreuder, 1992; Kamp, 1981). An appropriate theoretical account of the relations between the conceptual structure and semantic structure of natural language can lead to assessing students' mental models based on written responses. This study begins with modeling the relations of internal and external structures of representations on the grounds of Bierwisch and Schreuder (1992) (see Figure 4.2).

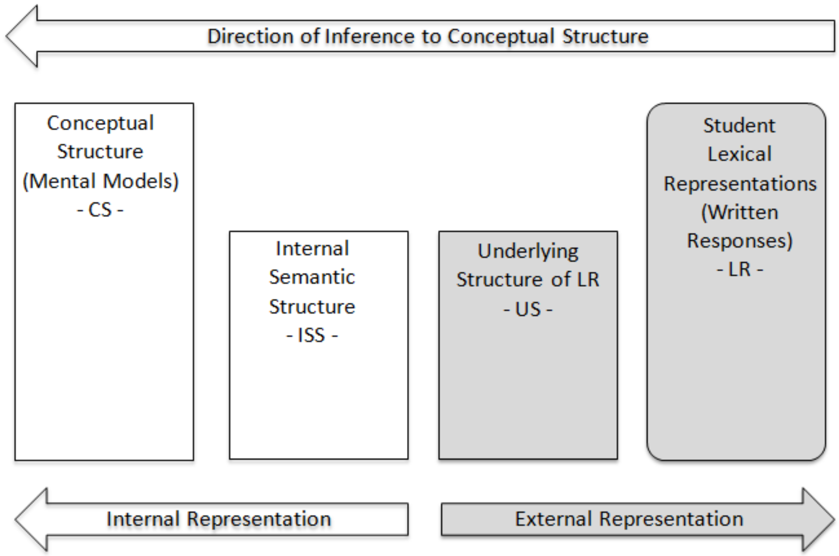


Figure 4.2. The relations of structures of internal and external lexical representations.

a. Conceptual structure (CS): CS indicates the student's actual mental model in which a problem situation is specified. It is assumed that CS is highly associated with but not dependent on the system of language.

b. Internal Semantic structure (ISS): ISS is the internally represented linguistic semantic structure that is determined by a "representational system of lexical meaning and their combination" (Bierwisch & Schreuder, 1992, p. 26). It is assumed that ISS belongs to a different mental domain from that of CS, so ISS is not necessarily identical to CS.

c. Underlying structure of lexical representation (US): US, as labeled in this study, indicates the underlying structure of external lexical representation (LR). US is represented as a deep structure constituted by the surface structure (i.e., individual concepts and relations) (Bransford & Johnson, 1972; Katz & Postal, 1964; Kintsch & van Dijk, 1978).

d. Lexical representation (LR): LR is a linguistic expression such as a student's verbal or written response to a problem situation.

Externalization process. Bierwisch and Schreuder (1992) explained that CS is the level of pre-verbal message structure. CS becomes the linguistic structure when a student chooses lexical items in accordance with his/her CS and then establishes an ISS that is determined by the representational system of a language. Finally the internal SS (ISS) is externalized when a student writes (or verbalizes) his/her response (LR).

CS inferred via ISS. Conceptual structure (CS) and Internal semantic structure (ISS) are internal representations that are not directly observable but can only be inferred based on observable representations. CS is believed to be substantially associated with representations of ISS (Bierwisch & Schreuder, 1992; Kintsch, 1974, 1994; Kintch & von Dijk, 1978). ISS is interpreted as a cognitive artifact of the mental lexical systems of a language (Levelt, 1989). Thus, it is presumed that ISS is directly inferred from the visually represented underlying structure (US) of a natural language representation (LR).

This study argues that the assessment efforts using concept map techniques conceptually aim at investigating CS but actually elicit ISS as a partial substructure of CS from externalized students' natural language writing. ISS is considered very similar to the visually represented underlying structure (US) of a lexical expression. In particular, this study underscores a technique for obtaining a coherent and descriptive concept map: that is, to extract meaningful concepts and relations from text according to the rules and meanings of the language. The described approach is called semantic relation (SR).

What Algorithms Can Be Applied to Distill Semantic Relations?

Although natural language researchers have tried to set algorithms and classifications for analyzing semantic relations (Beamer et al., 2008), it is rare to find concept map technologies that employ an analysis of semantic relations. This study suggests procedures and principles for distilling semantic relations including concepts from text responses.

Step 1. Process the sentences. The first step is to analyze the syntax of sentences, which includes three sub-steps: tokenizer, tagger, and syntactic parser.

- Step 1-1. The tokenizer is to break a document into lexical items called tokens (e.g., Sunlight/ melted/ a/ cake/ down/./).
- Step 1-2. The Tagger is a step to label each word with its corresponding part of speech in context (e.g., technology/ NN).
- Step 1-3. The syntactic parser is to group words into phrases using tagging names (e.g., A veterinarian investigated the cat: DT – NNP – VBZ – DT – NN).

Step 2. Distill concepts. Basically a noun is characterized as a concept in a discourse. In order to identify and distill concepts from a written response, a set of rules are established.

- Rule 1. A concept takes diverse forms of noun (Girju, 2011; Girju, Beamer, Rozovskaya, Fister, & Bhat, 2010; Girju et al., 2009; Moldovan & Girju, 2001, Murphy, 2003; Rijkhoff, 2002): (a) one-word noun, (b) noun compounds, and (c) noun and adjective pre-modifier.
- Rule 2. Distilled concepts (nouns) are primarily stored as singular.
- Rule 3. Pronouns are not replaced with the nouns they represent.

The distilled concepts include three types of nouns (Girju, 2008): one-word nouns (e.g., practice, technology, and classroom); noun compounds that consist of the head noun and noun modifier(s) (e.g., ‘bus station’ and ‘technology implementation'); and noun and adjectives pre-modifiers (e.g., technological intervention).

Step 3. Build sets of synonyms. Once concepts are distilled from the text, the concepts are grouped in synonym sets. Although each concept is basically regarded as having a unique meaning, concepts sharing a same or very similar meaning in the domain belong to a single category (Moldovan & Girju, 2001). For example, ‘chalkboard’ is synonymous with ‘blackboard’ and ‘whiteboard.’ This study aims to view those three concepts as one.

Step 4. Identify semantic relations (the pairs of concepts). Identifying semantic relations is the most important step because it provides information used for creating concept maps. This step is to select pairs of concepts, C_i and C_j, linked by a particular semantic relation. The principles determining the pairs of concepts are established according to linguistics studies (Downing, 1978; Girju, 2008; Hearst, 1998; Levi, 1978; Moldovan & Girju, 2004). The semantic relations determined by syntactic patterns that are classified as phrase-level patterns and sentence-level patterns (Girju, 2008; Hearst, 1998).

First, phrase-level patterns include prepositional phrases attached to nouns (noun phrases) or s-genitives. For example, ‘the library of the school’ is interpreted as having a semantic relation of ‘part–whole.’ A list of eight prepositions (of, for, in, at, on, from, with, and about), as defined by Lauer (1995), plays a critical role in determining the semantic relations based on algorithmic patterns.

Second, the semantic relation is also determined by the sentence. For example, in the sentence (the school has a new technology.), the relation of ‘school’ and ‘technology’ is categorized as ‘possession’. Twenty-two types of semantic relations are defined by Moldovan and Girju (2004). However, a natural language expression is not a simple sentence, and multiple patterns can exist in a sentence. Thus, 14 rules to determine semantic relations were specified in this study (see Table 4.2).

Table 4.2

Rules to Determine the Pairs of Concepts from Complex Lexico-Syntactic Patterns

Complex Lexico-Syntactic Patterns	Pairs of Concepts–$R(C_i , C_j)$
N0 is N1 and N2 (e.g. technology is hardware and software.)	(N0, N1); (N0, N2); (N1, N2)
N0 of N1 and N2 verb N3 (e.g., The use of technology and access to internet allow student to…)	(N0, N1); (N0, N2): (N0, N3); (N2, N3)
N0 such as N1, N2,…, Nn (e.g, classroom technologies such as laptops, Internet, and electric whiteboard)	(N0, N1); (N0, N2);…; (N0, Nn)
Such N0 as N1, N2,…, Nn (e.g., such new technologies as Web 2.0, cloud computing, and mobile internet)	(N0, N1); (N0, N2);…; (N0, Nn)
Np are N1, N2,…, Nn or other N0 (e.g., Magnetism is the positive or negative.)	(Np, N1); (Np, N2);…; (Np, Nn); or (Np, N0)
N0 include N2 and N3 (e.g., internal representation includes conceptual structure and linguistic semantic structure)	(N0, N1); (N0, N2)
N0, especially N1, verb… (e.g., supportive environments, especially leadership support is the most important.)	(N0, N1)
N1 of N2 in N3 of N4 (e.g, the use of technology in the classrooms of participating schools)	(N1, N2); (N3, N4); (N1, N3)
By –ing N1 and N2, Np verb N3[a] (e.g., By using the Internet and Smartphone, students can access learning materials any time, any where.)	(Np, N1); (Np, N2); (Np, N3)
N1 provide N2 with N3 (e.g., the Internet provides us with	(N1, N2); (N1, N3)
N1 and N2 verb N3 (e.g, Teachers and students are not used to using a computer.)	(N1, N2); (N1, N3); (N2, N3)
N1 verb that-clause (e.g., A witness hated that the boy attacked the victim.)	(N1, the first N in that-clause)[b]
N1 between N2 and N3 (e.g., discrepancy between boys and girls)	(N1, N2); (N1, N3)
N1 that N2 verb N3[c] (e.g, Teachers maintain the belief that these efforts will have positive results.)	(N1, N2); (N2, N3)

Note. a. Subordinate clause in which subject is omitted. b. It is to make a connection between N1 and that-clause. c. Conjunction clause.

Step 5. Determine the direction of relations. The last step is to determine the directional relations between two paired concepts. The relation in which the direction begins from the first concept and ends to the second concept is classified as follows: subject to object; source to target; from A to B; cause to effect; mutual relation (the first to the second); A belongs to B or B includes A; superior to inferior; A exists for B; A serves for B; tool to object; people to object; nouns linked with eight prepositional modifiers (the first to the second).

Finally, in order to construct a concept map, all concepts distilled from a text response are listed and paired with one another in a matrix. The paired concepts having semantic relations are given the vector value of 1 in an *n* by *n* concept array where *n* is the number of concepts; otherwise, the vector value is 0. In addition, to include directional information, the first concept (C_i) is considered as the source and the second concept (C_j) becomes the target in a pair.

Methods for Comparisons of SR, PR, and AR Approaches

The semantic relation (SR) is assumed as a way to elicit a descriptive and complex concept map from a text response. In order to test that assumption this study involves a comparison study amongst concept maps constructed by semantic relation (SR), proximity approach (PR) and adjacency (AR) approaches.

Participants

Participants included seven professors teaching at six major universities in the United States. The professors participated in a Delphi survey to obtain a reference model for a complex problem in terms of a technology implementation problem in K-12 schools. It was assumed that using expert responses would enable us to investigate the technologies' ability to detect higher-level responses.

The Problem-solving Task

As part of the Delphi survey, the panel was asked for responses to a complex problem situation using natural language. The task provided a simulated situation in which professors were assumed to be participating in an evaluation project, the purpose of which was to investigate an unsuccessful project that had as its goal adapting a technology (i.e., a tablet PC) for classroom teaching. In order to elicit professors' knowledge in detail, the questions asked them to explicitly describe the concepts, issues, factors, and variables likely to have contributed to the result that the introduction of tablet PCs had very little effect on the instructional practices employed in the classes.

Reference Modeling via Delphi Survey

This study included a reference model for the problem situation. The model was created using a Delphi survey procedure (Goodman, 1987; Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). The Delphi survey involved three iterations to develop a refined reference model that the seven experts accepted. In the first round, the participating professors created their own responses to the problem; then, all the panel's responses were consolidated. Next, a document including all statements from the professors and a list of concepts identified from the panel's responses was sent to the panel again. The professors were asked to add their comments regarding the listed statements and concepts and rank them. After gathering the second round of surveys, the researcher created a final list of ranked statements and concepts. Based on this summary, a draft of a reference model was created. In the final round, the results of the second survey were sent to the panel and revised according to their comments as necessary. Through this procedure, a reference model containing 23 key concepts was developed.

Analysis Procedure

Data manipulation. A total of eight responses involving seven professors' initial responses and the reference model were used for the study. Concepts and relations in SR approach were manually distilled in accord with the procedure described earlier. In PR approach, T-MITOCAR tool was used to process the data. As to AR, ALA-Reader tool, as a representative tool of AR approach, requires predefined key concepts so as to distill relations of the concepts from the text. The 23 terms (concepts) defined by the expert panel were used in the analysis.

Construct concept maps. Two kinds of tools were used for constructing concept maps. Netminer software was used for creating concept maps of SR and AR because the software processes adjacent matrices with 0 or 1 values. Concept maps of PR were created by T-MITOCAR software that has an embedded function to generate concept maps using proximity data.

Data comparison. Comparisons of the three approaches (SR, PR, and AR) included quantitative and qualitative reviews. Concepts and relations were distilled from eight natural language responses based on the three approaches. The descriptive statistics of the results were first compared in terms of the number of concepts and relations.

PR and AR include only key concepts determined by the frequencies of occurrence of concepts or expert judgments, while SR contends that key concepts emerge from the concept map as a whole. This study uses the centrality measure based on Affected by Anthonisse (1971) and Freeman's (1977) assertion that a concept can exert control over the interaction between other pairs of concepts in a network.

$$C_B(n_i) = \sum_{j\langle k} g_{jk}(n_i) / g_{jk}$$

where g_{jk} is the number of geodesics that are the shortest path between two concepts $d(n_j, n_k)$ in the network, $g_{jk}(n_i)$ indicates g_{jk} that contains a certain concept *i*, and $g_{jk}(n_i)/g_{jk}$ is the probability that a concept *i* is included in the geodesics between concept *j* and *k*. By using the equation above one can obtain the centrality value of concept *i*, $C_B(n_i)$, and then the value is standardized as a measure of $0 \leq C \leq 1$. This study suggests that concepts having values larger than 0 were considered as key concepts in the concept maps.

Key concepts identified by centrality values in SR were compared to the 23 terms experts determined for AR and the terms used in PR. For those comparisons, two types of similarity measures were applied: (a) numerical similarity and (b) conceptual similarity. The comparisons of the number of key concepts are derived from the equation:

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

Conceptual similarity indicating the extent to which the paired models share the same concepts and relations is calculated by the Tversky's (1977) formula:

$$\mathrm{s} = \frac{f(\mathrm{A} \cap \mathrm{B})}{f(\mathrm{A} \cap \mathrm{B}) + \alpha \cdot f(\mathrm{A} - \mathrm{B}) + \beta \cdot f(\mathrm{B} - \mathrm{A})}$$

where α and β are weights for differentiating the quantities between A and B. This study assumes that there is no difference in weights because of the two reasons: (a) the three approaches derived key concepts from the same data; (b) each part of a pair is assumed to be equally significant. Thus, the weights of α and β were set as equal to 0.5 (α = β = 0.5). As for the qualitative review, a visual inspection of the elicited concept maps was conducted.

Comparisons of Concept Maps

Judgments of Concept Relations

SENTENSE 1: **Instructional [need]** was likely not fully identified due to insufficient **study** of how **instructional [practices]** in the **[classroom]** were being conducted already without the **[technology]**.
SENTENSE 2: One big **issue** is defining what a successful **[integration]** or **[change]** in **instructional [practice]** actually is.

Semantic Relation (SR)			Proximity Relation (PR)		Adjacent Relation (AR)	
Concept *i*	**Concept *j***	**Relation *ij***	**Concept *i***	**Concept *j***	**Concept *i***	**Concept *j***
Instructional need	Study	Not identified due to	Need	Practice	Instructional need	Instructional practice
Study	Instructional practice	Of	Need	Classroom	Instructional practice	Integration
Instructional practice	Classroom	In	Need	Technology	Integration	Change
Instructional Practice	Technology	Conducted without	Practice	Classroom	Change	Instructional practice
			Practice	Technology		
			Classroom	Technology		

Figure 4.3. Extracted concepts and relations in a sample text.

Figure 4.3 describes how differently the three approaches distill relations of paired concepts from the same text. The bold words are the concepts used for SR, and the underlined words used for AR are key concepts determined by the seven experts. The words in square brackets denote key concepts determined by PR. Every semantic relation including concepts in the SR approach is regarded as a necessary component (i.e., surface structure) that constitutes an individual's current knowledge structure (i.e., deep structure). For example, in the first sentence, SR distilled five concepts (i.e., instructional need, study, instructional practice, classroom, and technology) and built four semantic relations (instructional need & study, study & instructional practice, instructional practice & classroom, and instructional practice & technology).

Judgment of relations in PR and AR depends on the distance between the two concepts. In the same sentence, T-MITOCAR as a PR tool identified four terms (i.e., need, practice, classroom, and technology), and all pair-wised relations were calculated valued with the number of words between each pair of concepts in a sentence, while other pairs that are not seen in the sentence get maximum distance value. The T-MITOCAR detected only the single noun type of concepts, which is likely problematic when interpreting a concept map due to ontological differences (e.g., difference of meaning between 'instructional need' and 'need').

ALA-Reader as an AR tool suggests the linear aggregation method in which word relations are identified across sentences. For instance, in Figure 4.3, the term 'instructional practice' in the first sentence is considered to be associated with the term 'integration' in the second sentence. Linear aggregation renders concepts associated across sentences. However, it is possible to link two key terms with little or no semantic relations.

Table 4.3

Descriptive Statistics: the Number of Concepts and Relations in SR, PR, and AR

Model	Word count	Concepts (Nouns)			Relations (Links)		
		SR	PR	AR	SR	PR	AR
Reference	397	54	14 (26%)	23 (43%)	64	22 (34%)	35 (55%)
Expert 1	319	41	13 (32%)	8 (20%)	43	18 (42%)	7 (16%)
Expert 2	411	42	11 (26%)	10 (24%)	54	24 (44%)	15 (28%)
Expert 3	331	37	13 (35%)	9 (24%)	50	22 (44%)	15 (30%)
Expert 4	297	29	10 (34%)	5 (17%)	30	12 (40%)	5 (17%)
Expert 5	436	48	19 (40%)	6 (13%)	55	41 (75%)	6 (11%)
Expert 6	481	52	17 (33%)	7 (13%)	62	33 (53%)	8 (13%)
Expert 7	751	50	13 (26%)	10 (20%)	63	23 (37%)	22 (35%)

Note. The values in the parentheses indicate the proportions of the numbers in PR or AR to those in SR.

Distilled Concepts and Relations

The descriptive statistics obtained from the three approaches were investigated in terms of the numbers of concepts and relations (see Table 4.3). The numbers of concepts and relations in SR are much greater than those in PR and SR. For example, compared to SR, the proportion of

PR ranged from 26 % to 40%, while the proportion of AR fell between 13% and 43%. As for the proportion of relations, PR reached the number of SR ranging from 34% to 75% and AR ranged from 11% to 55%. Overall, the numbers of concepts and relations in PR were greater than those of AR. The strict function of ALA-Reader that extracts only concepts exactly matched with the predefined concepts, on the whole, resulted in the smallest numbers of concepts and relations. However, a reverse result was observed in the reference model. The numbers in AR were closer to those of SR than those of PR. The reason is because the reference model included all 23 key concepts and was written in a cohesive manner.

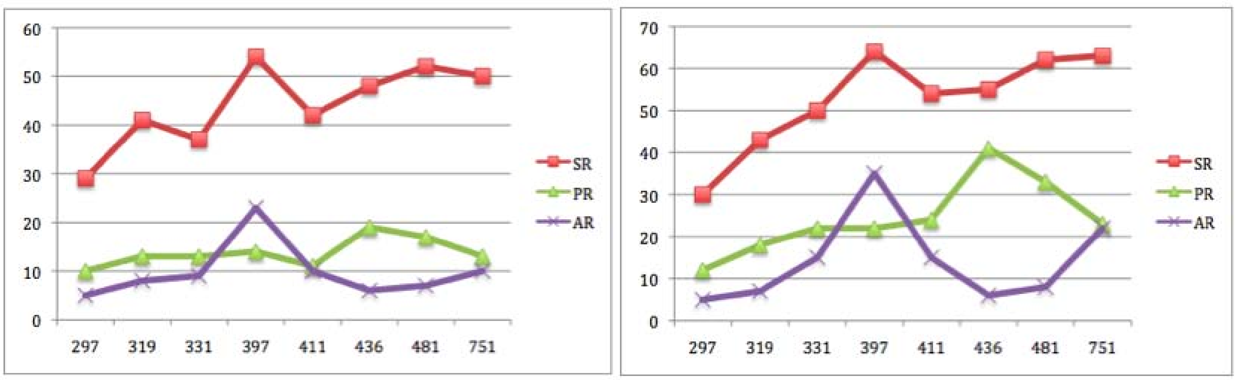


Figure 4.4. Association of word counts with the numbers of concepts or relations in SR, PR, and AR. The left represents the associations between the number of concepts and the word counts of responses and the right represents those between the number of relations and the word counts of responses.

Graphical investigations were conducted on the association of word counts and the numbers of concepts or relations. As Figure 4.4 depicts, high variations were observed in all three approaches. Nonetheless, SR showed a moderate positive line with word count. In contrast, for PR and AR (i.e., T-MITOCAR and ALA-Reader), there was little relation with word count. Those results implied that PR and AR approaches tend to be abstractive in ways that describe knowledge structure and are likely losing the semantic information constituting a whole structure. In particular for AR, some sharp drops were observed, even in responses having a higher word count. That implied AR is more sensitive to writing in terms of respondents' background knowledge, selection of words, and writing style (e.g., frequent use of pronouns).

Key Concepts Derived from the Deep Structure

This study proposed the SR approach on the basis of two assumptions: (a) SR provides descriptive information such as adequate concepts and relations to create complex concept maps; and (b) a substantial part of the meaning can be derived from the complex concept maps as deep structure (Fodor, Bever, & Garrett, 1974; Katz & Postal, 1964). The latter assumption was investigated using the reference response.

Table 4.4

The number of Key Concepts Similar or Dissimilar among the Three Approaches in the Reference Model

Pair (A with B)	f (A−B)	f (B−A)	f (A∩B)
SR with AR	7	6	17
SR with PR	13	3	11
AR with PR	14	5	9

Note. The numbers of key concepts of SR, AR, and PR are 24, 23, and 14 respectively.

The concept map of the reference response drawn from SR involved 24 key concepts that had a centrality value no less than zero calculated with the aforementioned Freeman's (1977) formula. T-MITOCAR (as a PR tool) selected 14 nouns based on frequencies of the nouns, and 23 key concepts determined by experts were used for ALA-Reader (as an AR tool). The concepts used for AR were considered the standards. Key concepts for the three approaches were compared using the numerical and conceptual similarities between the pairs of approaches (see Table 4.4). As for the nouns of PR, due to the limitation of T-MITOCAR that distills only a single-noun type of concepts, the nouns of PR similar to the complex nominals used in the other two approaches were regarded as the same. For example, 'need' is considered similar to 'instructional need.' The results showed that a substantial part of key concepts of SR were overlapped with those of AR and PR, 17 and 11 among 24 respectively.

Table 4.5

Similarities of Key Concepts in the Reference Model among SR, AR, and PR

	1	2	3
SR	-	0.96	0.58
AR	0.72	-	0.61
PR	0.68	0.64	-

Note. Numerical similarities are in the upper diagonal, and conceptual similarities are located in the lower diagonal.

Similarity measures showed clear functionality of centrality measures in SR (See Table 4.5). The numerical similarity between SR and AR was very high, 0. 96. The conceptual similarity of SR was high across AR and PR, 0.72 and 0.68 respectively, whereas the similarity between PR and AR was somewhat lower than that of SR and AR. Those results proved that (a) SR is a descriptive and informative approach because it includes all possible semantic relations in a concept map; (b) in addition, SR can provide abstractive information in terms of key concepts and relations using centrality measures; and (c) thus, SR accommodates the theory of semantic structure including both surface structure and deep structure.

Visual Inspection

A visual inspection of the concept maps was conducted. Figures 4.5a, 4.5b, and 4.5c illustrate the diverse features of concept maps drawn from the three approaches. The concept maps of SR were much more complex and descriptive in terms of the numbers of concepts and relations than those of T-MITOCAR and ALA-Reader, across the samples. As for the reference model, the concept maps of the three approaches were highly cohesive and connected due to two facts: (a) the reference response was written very carefully to make it cohesive, with the writing connecting key concepts as much as possible; and (b) T-MITOCAR and ALA-Reader technically let all elements of the concept maps connect, assuming all concepts are linked in mind.

In contrast, the concept maps of the expert 06 differed substantially from one another. The concept map of SR contains two components that are subsets of a network having no connection with the other subsets. This study claims that this feature of the concept map attests to the fact that SR elicits concept maps akin to mental models because it is hard to believe that all concepts are always represented as connected without exception.

The concept maps of PR and AR looked insufficient to be used for an instructional purpose such as providing formative feedback and instructional remedy. In the concept maps, all concepts are connected without information about their relations, which make it hard to interpret the meaning of the concept maps. Moreover, the data reduction process contained by PR and AR possibly causes problems in interpreting concept maps. As to PR, the expert 06 concept map was more complex than the reference model in terms of the numbers of concepts and relations (see Figure 4.5b), while the expert 06 model of AR is too simple compared to that of the other two approaches (see Figure 4.5c).

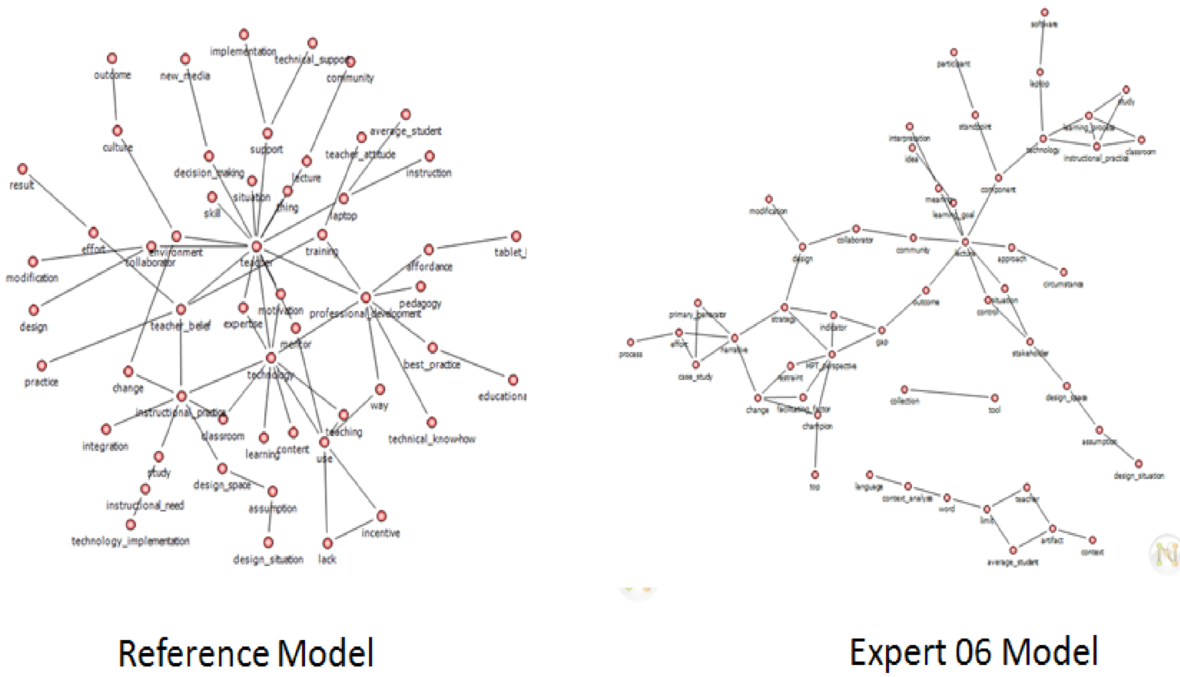


Figure 4.*5a*. Concept map of the reference model and expert 06 drawn from SR.

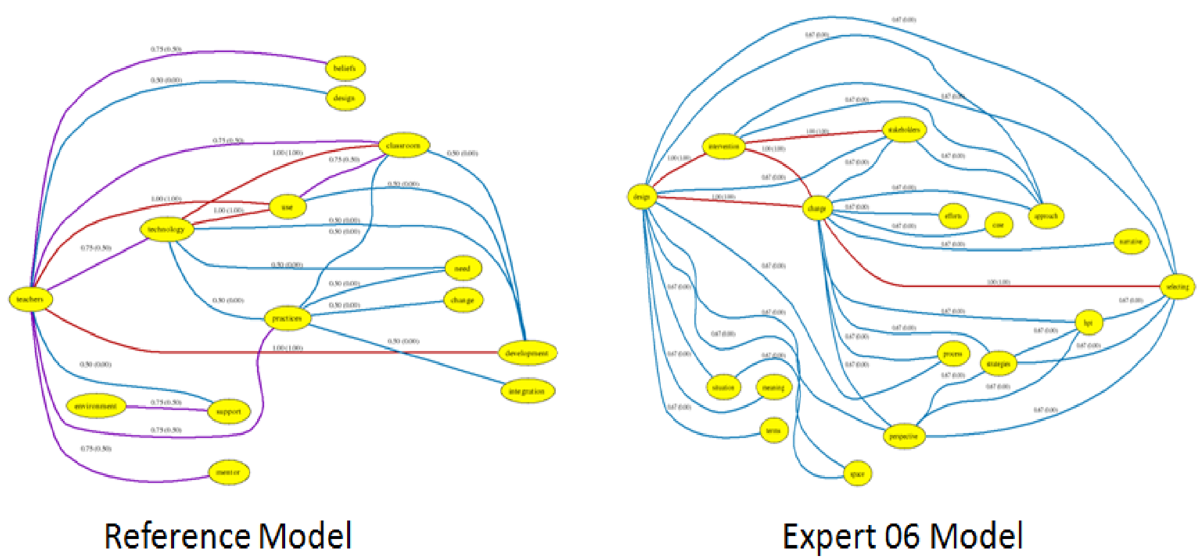


Figure 4.*5b*. Concept map of the reference model and expert 06 drawn from T-MITOCAR (PR).

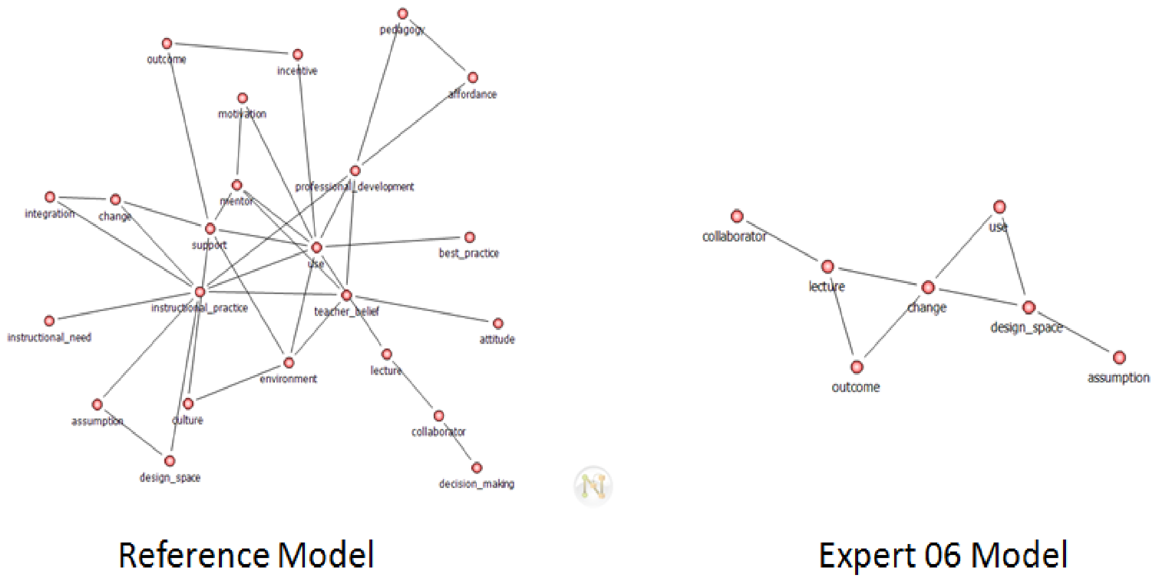


Figure 4.5*c*. Concept map of the reference model and expert 06 drawn from AR (ALA-Reader).

Conclusion

This study is an initial effort to devise new methodologies and technologies for eliciting concept maps from students' written responses that reflect their internal representations to a problem situation(s). In particular, it is based on the argument that semantic relations together with concepts (as in an SR approach) obtained from a written response so that a concept map are more likely to represent a student's internal mental representation and thinking than approaches based primarily on spatial and surface features of the written response (as in PR and AR approaches).

Underlying theories and assumptions of relations between language and cognition were intensively reviewed in terms of internal and external representations via the lexical systems of a language. It was argued that the internal semantic structure is externalized as the linguistic semantic structure of a linguistic expression. Eliciting the semantic structure was assumed to be a better way to visually represent concept maps. That approach was termed ‘Semantic Relation’ in contrast with the current approaches, ‘Proximity Relation (adopted in T-MITOCAR)’ and ‘Adjacent Relation (employed in ALA-Reader).’

Just as the belief that a semantic structure is represented as a whole, that the macrostructure (deep structure) is constituted by the microstructure (surface structure), the data investigation proved that SR is an alternative way of enabling concept maps to be more descriptive and authentic. In particular, centrality measures showed that key concepts can be detected by the authentic microstructure. Moreover, SR was more robust against variations in contexts and writing styles.

The primary conclusion of this study is that the SR approach can be an effective approach when the goal of the concept map is to obtain more meaningful (and thus formative) information about students’ cognitive changes. In contrast, this study does not intend to argue that the proposed SR approach is always superior to other methods and technologies. PR and AR represented by T-MITOCAR and ALA-Reader tool, respectively, can be ways to provide information about cognitive status, succinctly and economically focusing on key concepts and relations. For example, when key concepts are explicitly defined in conjunction with learning goals, and the goal of instruction is to help students correctly internalize them along with their prior knowledge, the ALA-Reader tool can be a good method for monitoring students’ cognitive changes on the condition that the concepts are introduced and explained to some extent or at least sufficiently.

As a matter of course, there are many issues for future development and study. For example, although SR is proposed as possibly automated, it is true that complex structured sentences are not easily interpreted in a concept map in terms of distilling correctly paired concepts. Thus, in order to elicit concept models from multiple quantified language inputs, technological supports in terms of natural language processing are required. Some suggestions include the following:

- Elaborate the algorithm to identify semantic relations from a text, including developing automated natural language processing.
- Build diverse measures that capture the attributes of the knowledge structure.
- Elaborate the methodology to compare concept maps to the reference model(s) and to monitor structural changes as learning trajectories.

The SR approach embedded in automated technologies could be an effective tool to cut across disciplinary boundaries such as traditional language comprehension study in linguistics. In learning and instruction, SR is applicable to a wide range of areas: automated essay evaluation; expertise modeling; competency diagnosis in adult learning; technology-enhanced adaptive learning systems (e.g., intelligent tutoring system); longitudinal study of learning progress; and formative assessment and feedback. For example, the SR technology will enable teachers to identify individual student progression on a complex problem solving based on the whole structure of the concept map and to provide personalized feedback in terms of missing key areas and relations.

References

Adriana, D. M., Moldovan, D., Badulescu, A., Tatu, M., Antohe, D., & Girju, R. (2004). Models for the semantic classification of noun phrases. *In HLT-NAACL 2004: Workshop on Computational Lexical Semantics*, 60-67.

Anthonisse, J. M. (1971). *The rush in a graph.* Amsterdam: Mathematische Centrum.

Anzai, Y., & Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction, 1*, 397-450.

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton, N.J.: Princeton University Press.

Beamer, B., Rozovskaya, A.,& Girju, R. (2008) Automatic Semantic Relation Extraction with Multiple Boundary Generation. *Proceedings of AAAI*, 824-829.

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*(1-3), 23-60. doi:16/0010-0277(92)90039-K

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3,* 193-209.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bransford, J. D., & Franks, J. J. (1972). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 331-350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717-726. doi:16/S0022-5371(72)80006-9

Brown, D. (1997). *An introduction to object-oriented analysis: Objects in plain English*. New York: John Wiley & Sons.

Cañas, A. J. (2009). What are Propositions? ... from a Concept Mapping Perspective. Retrieved July 31, 2011, from http://cmap.ihmc.us/docs/Proposition.html

Champagne, A. B., Kouba, V. L., & Hurley, M. (2000). Assessing inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiry into inquiry learning and teaching in science* (pp. 456-461). Washington, DC: American Association for the Advanced of Science.

Chi, M.T.H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5* (April/June), 121-152.

Chi, M.T.H., Glaser, R., & Rees, E. (1981).Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, N. J.: Lawrence Erlbaum.

Chomsky, N. (1957). *Syntactic structures*. London: Mouton and Company.

Clariana, R. B. (2010). Multi-decision Approaches for Eliciting Knowledge Structure. In *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 41-59). Retrieved from http://dx.doi.org/10.1007/978-1-4419-5662-0_4

Clariana R. B., & Koul, P. (2008). The effects of learner prior knowledge when creating concpet maps from a text passage. *International Journal of Instructional Media*, *35*(2), 229-236.

Clariana, R. B., & Taricani, E. M. (2010). The consequences of increasing the number of terms used to score open-ended concept maps. *International Journal of Instructional Media*, *37*(2), 218–226.

Clariana, R. B., & Wallace, P. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research*, *37*(3), 211-227. doi:10.2190/EC.37.3.a

Clariana, R., Wallace, P., & Godshalk, V. (2009). Deriving and measuring group knowledge structure from essays: The effects of anaphoric reference. *Educational Technology Research and Development*, *57*(6), 725-737. doi:10.1007/s11423-009-9115-z

Collins, A. & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought*, (pp. 243-265). Cambridge, MA: Cambridge University Press.

Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural Comparison of Cognitive Associative Networks in Two Populations. *Journal of Applied Social Psychology*, *37*(9), 2097-2129. doi:10.1111/j.1559-1816.2007.00253.x

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, *13*(5), 533-568. DOI:10.1016/S0959-4752(02)00025-7

Downing. (1977). On the creation and use of English compound nouns. *Language, 53*(4), 810-842.

Duschl, R. (2003). Assessment of inquiry. In J. M. Atkins & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41-60). Arlington, VA: NSTA Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Freeman, L. C. (1977). A set of measures of centrality based on betweeness. *Sociometry, 40*, 35-41.

Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle [Control of dynamic systems by building up and using subjective causal models]. *Zeitschrift für Psychologie, 193*, 443-466.

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chicheser: Ellis Horwood Ltd.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove, UK: Psychology Press.

Girju, R. (2008, August). *Semantic relation extraction and its applications*. Paper presented at 20th European Summer School in Logic, Language and Information, Freie und Hansestadt Hamburg, Germany.

Girju, R. (2011). The Syntax and Semantics of Prepositions in the Task of Automatic Interpretation of Nominal Phrases and Compounds: A Cross-Linguistic Study. *Computational Linguistics*, *35*(2), 185-228. doi:i: 10.1162/coli.06-77

Girju, R., Beamer, B., Rozovskaya, A., Fister, A., & Bhat, S. (2010). A knowledge-rich approach to identifying semantic relations between nominals. *Information Processing & Management*, *46*(5), 589-610. doi:10.1016/j.ipm.2009.09.002

Girju, R, Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., & Yuret, D. (2009). Classification of semantic relations between nominals. *Language Resources and Evaluation*, *43*(2), 105-121. doi:10.1007/s10579-009-9083-2

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, *12*(6), 729-734.

Greeno, J.G. (1989). Situations, mental models, and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285–318). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings on 14th international conference on computational linguistics* (COLING-92) (539–545).

Hsu, Chia-Chien & Sandford, Brian A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation, 12*(10). Available online: http://pareonline.net/getvn.asp?v=12&n=10

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A micro-analytic review. *Psychonomic Bulletin and Review*, *10*, 785–813.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185-208). Cambridge University Press.

Johnson-Laird, P.N. (2005b). The history of mental models. In K. I. Manktelow, & M. C. Chung (Eds), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179-212). Psychology Press.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.*

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277-322). Amsterdam: Mathematisch Centrum.

Katz, J. J., & Postal, P. M.(1964). *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294-303. doi:10.1037/0003-066X.49.4.294

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394. doi:10.1037/0033-295X.85.5.363

Larkin, J. H. (1981). Enriching formal knowledge: A model for learning to solve problems in physics. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Lauer, M. (1995).Corpus statistics meet the noun compound: Some empirical results. *Proceedings of the Association for Computational Linguistics Conference* (ACL) (47–54), Cambridge, MA.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

McGrath, M. (2011). Propositions. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved July 31, 2011, from http://plato.stanford.edu/entries/propositions/

Miller. G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge. MA: Harvard University Press.

Miller, K. (1998). Modifiers in WordNet. In C. Fellbaum (Ed.), *WordNet: an electronic lexical database* (pp.61-64). Cambridge, MA: MIT Press.

Moldovan, D. I., & Girju, R. C. (2001). An interactive tool for the rapid development of knowledge bases. *International Journal on Artificial Intelligence Tools*, *10*(1/2), 65-86.

Murphy, M. L. (2003). *Semantic relations and the lexicon: antonymy, synonymy and other paradigms*. Cambridge, MA: University Press.

Narayanan, V. K. (2005). Causal mapping: An historical overview. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for research in information technology* (pp.1-19). Hershey, PA: Idea Group Publishing.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the POP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2) (pp. 531-546). Cambridge, MA: MIT Press.

Novak, J. D., & Cañas, A. J. (2006). *The origins of the concept mapping tool and the continuing evolution of the tool*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.3382

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15-29. doi:10.1016/j.im.2003.11.002

Phelan, J., Kang, T., Niemi, D. N., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessments in middle school mathematics*. (CRESST Report 750). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. (2010). Highly integrated model assessment technology and tools.*Educational Technology Research & Development, 58*(1), 3-18. doi:10.1007/s11423-009-9119-8.

Proposition (n.d.). Retrieved July 31, 2011, from Wikipedia: http://en.wikipedia.org/wiki/Proposition

Quellmalz, E. S., & Haertel, G. D. (2004). *Use of technology-supported tools for large-scale science assessment: implications for assessment practice and policy at the state level*. Commissioned paper prepared for the National Research Council's Committee on Test Design for K-12 Science Achievement, Washington, DC.

Rijkhoff, J. (2002). *The noun phrase*. Oxford, UK: Oxford University Press.

Rumelhart, D. E., McClelland, J. L., & Group, P. R. (1987). *Parallel Distributed Processing, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

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

Schvanevldt, 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* (pp. 249-284). San Diego, CA: Academic Press.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: studies in knowledge organization*. Ablex Publishing Corp. Retrieved from http://portal.acm.org/citation.cfm?id=SERIES9012.119801

Seel, N.M. (1999). Semiotics and structural learning theory. *Journal of Structural Learning and Intelligent Systems, 14* (1), 11–28.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science*, *29* (4-5), 403-427.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition, and Learning, 1* (1), 59-85.

Seel, N. M. (2004). Model-centered learning environments: Theory, Instructional design, and ethics. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instruction design: International perspectives* (pp. 49-74). NJ: Lawrence Erlbaum Associates, Inc.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation*, *1*(1), 4-35.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation, 23*(4), 373-398.

Shute, V. J., & Zapata-Rivera, D. (2007). Adaptive technologies. In J. M. Spector, M. D. Merill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 227-294). Routledge: Taylor & Francis Group.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115-163.

Snow, R.E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14* (5), 455–473.

Spector, J. M. (2004). Problems with problem-based learning: Comments on model-centered learning and instruction in Seel (2003). *Technology, Instruction, Cognition and Learning, 1*(4), 359-374.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Taricani, E. M., & Clariana, R. B. (2006). A Technique for Automatically Scoring Open-Ended Concept Maps. *Educational Technology Research & Development*, *54*(1), 65-82. doi:Article

van Dijk, T. A. (1977). Macro-structures, knowledge frames, and discourse comprehension. In M. A. Just & P.Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 3-32). Hillsdale, N.J.: Erlbaum.

Wasserman, S., & Faust, K. (1994). Social networks analysis: Methods and applications. Cambridge: Cambridge University.

Wittgenstein, L. (1922). *Philosophical investigations*. London: Blackwell.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477-501.

CHAPTER 5

INVESTIGATION OF A MODEL OF STAGE-SEQUENTIAL LEARNING PROGRESS IN PROBLEM SOLVING[5]

[5] Kim, M. Submitted to the *Contemporary Educational Psychology*, 12/08/2011.

Abstract

It is generally accepted that students understand a complex problem based on their own personal knowledge base in which concepts and relations are embedded (Chi, 2008; Johnson-Laird, 1983; Seel, 2003; Vosniadou, Vamvakoussi, & Skopeliti, 2008). This study argues that learning is a process of transitioning from one stage to another stage within this knowledge base. The theoretical framework of learning progress, titled *a model of stage-sequential learning progress*, provided a diagnostic framework of learning progress in a problem-solving situation. In addition, this study involved the development of diagnostic methodologies that include: (a) devising parameters to quantify knowledge structure; (b) defining multidimensional constructs of knowledge structure; (c) creating measures to capture cognitive changes; and (d) constructing a diagnostic model adaptable to the theoretical framework. Drawing on these methods, a validation study was conducted with empirical data. The analysis demonstrated only two stages have a probability of presence in the data rather than the four stages assumed in the framework. It is argued that either two stages or four stages would be theoretically acceptable. Although further studies are required, this study as an initial effort can provide a diagnostic model of learning progress and research methods for future studies of cognitive change in complex problem solving.

*Keywords***:** learning progress, cognitive change, problem solving, assessment, latent class model, mental models

This study aimed at validating[6] a model of learning progress in a complex problem-solving situation so as to utilize that model for assessment and instructional support in classrooms. The discussions are on (a) conceptualizing, (b) parameterizing, and (c) validating a model of learning progress. More specifically, a theoretical model of stage-sequential learning progress was posited. To parameterize that model, this study draws on concept maps as re-represented learners' knowledge structures through which measures possibly able to describe students' cognitive states in learning were explored. Afterward, the assumptions suggested by the model were investigated.

Many current assessment methods that tend to focus on knowledge outcomes only reflect partial aspects of knowledge (Grotzer & Perkins, 2000; Thomas, 2005), but problem-solving assessments have received less attention in national testing systems (NRC, 2001, 2003, 2005). Learning problem solving is not a simple task because problem solving mostly includes higher-order thinking (e.g., critical thinking, abstract reasoning, meta-cognition etc.) (de Vries, Lund & Baker, 2002; Glassner, Weinstock, & Neuman, 2005; Hammer, 2000; Karoly & Panis, 2004; Kuhn , Black, Keselman, & Kaplan, 2000; NRC, 2005; Sandoval, 2003) as well as ample domain knowledge (Bransford, Brown, & Cocking, 2000; Chi, Glaser, & Farr, 1988; Newell & Simon, 1972; Ericsson, 2006). Therefore, assessing problem solving must investigate problem-solving knowledge and skills as a whole.

Cognitive change takes place when people confront unfamiliar, challenging situations (diSesa, 2006; Festinger, 1962; Piaget, 1964). When striving to resolve problem situations, learners experience changes in their mental representations by which these problem situations are recognized, defined, and organized (Seel, 2003, 2004). The theory of mental models explains

[6] Validation in this study indicates the practice of confirming a predictive model by estimating model-fitness to the empirical data; this study involves two validation practices for (a) a multi-dimensional model of knowledge structure and (b) a model of learning progress.

that problem solving involves a process of building mental representations to a problem situation (Johnson-Laird, 1983). It is believed that mental models are structurally represented as a whole including facts, concepts, variables, objects, and their relations involved in a problem situation (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Jonassen, Beissner, & Yacci, 1993; Segars, 1997).

Learners possibly experience qualitatively different levels of knowledge structure when engaged in problem solving. Just as Piaget (1964) argued that children experience qualitatively-distinct sequential knowledge states, developmental psychologists have supported the idea that learning and development evolve as the learner constructs a qualitatively-distinct knowledge structure (Alexander, 2003, 2004; Flavell & Miller, 1998; Siegler, 2005; Siegler, Thompson, & Opfer, 2009; Werner, 1957; Vygotsky, 1934/1978). A number of experimental studies proved that qualitatively different cognitive changes take place when learners respond to problems (e.g., Alexander & Murphy, 1998; Chen & Siegler, 2000; Opfer & Siegler, 2004; Siegler et al., 2009).

So far, there are accounts of the existence of qualitatively distinct states of learning progress, yet there are few specific models of qualitative levels of learning progress along with assessment methodologies applicable to problem-solving situations. The purpose of this study is to identify the theoretical foundations of cognitive changes, to build a diagnostic model of learning progress in problem solving, called *a model of stage-sequential learning progress*, and to explore particular parameters and statistical methods pertinent to testing hypothesized learning stages. This study ultimately proposes ways to diagnose cognitive development in problem-solving situations, which are essential to provide a best-fit learning experience to students with precision and confidence (Grow-Maienza, Hahn, & Joo, 2001; Hattie, 2009; Stigler and Stevenson, 1991).

Stage-Sequential Learning Progress

Theoretical Foundation

Learning progress can be defined as a series of qualitatively distinct stages in a learner's understanding which experiences both gradual and sudden changes through instruction. That is, learning progress is *stage-sequential*. The theory of mental models provides a conceptual foundation accounting for these cognitive changes.

Mental models are cognitive artifacts which a learner constructs in order to understand a given problem situation (Anzai & Yokoyama, 1984; Kieras & Bovair, 1984; Mayer, 1989). A problem situation is mentally represented in a learner's mind when a learner is involved in a problem-solving process. Mental models depend primarily on a learner's prior knowledge (that is, prior mental models) and change over time (Seel, 1999, 2001). Mental model changes in a learning situation are not simple shifts without any directional goal, but transformations toward goals; therefore, such changes can be considered progress. Thus, the progress of mental models involves learning-dependent and developmental transitions between preconceptions and causal explanations (Anzai & Yokoyama, 1984; Carley & Palmquist, 1992; Collins & Gentner, 1987; Johnson-Laird, 1983; Mayer, 1989; Seel, 2001, 2003, 2004; Seel & Dinter, 1995; Shute & Zapata-Rivera, 2008; Smith, diSessa, & Roschelle, 1993; Snow, 1990).

Mental models change over time while developing mastery in problem situations (Johnson-Laird, 2005a, 2005b; Seel, 2003, 2004). The progress of mental models involves qualitative change as well as quantitative trends in the frequency of existing models. While intellectual growth goes on in a learning context, mental models may also evolve through a variety of qualitative stages. For instance, mental models may be enlarged when a learner adopts new concepts into an existing model (quantitative increase), or a mental model may experience fundamental changes to adapt to a new situation (qualitative change).

A Model of Stage-Sequential Learning Progress

A conceptual model of stage-sequential learning progress can be derived from the studies of knowledge structure and the development of expertise. Mental models can be hypothesized as progressing through different levels of structural knowledge representing certain aspects of external situations in specific domains (Johnson-Laird, 2005a, 2005b). When problem solving is defined as a process of mental activity relying on structurally represented mental models (Dochy et al., 2003; Segars, 1997), it is necessary that the assessment of problem-solving knowledge and skills is sensitive to the structural characteristics of the knowledge base (Gijbel, Dochy, Van den Bossche, & Segers, 2005).

The 3S knowledge structure. Author (2011a) argued that structural characteristics of mental models can be illustrated with three features of knowledge structure (labeled "3S") involving (a) surface, (b) structure, and (c) semantic features. First, the surface feature denotes the descriptive information of knowledge components in terms of the numbers of concepts and their relations. The surface feature is compatible with the surface level of mental models explained as relevant objects and aspects of the context (Holyoak & Koh, 1987; Simon & Hayes, 1976). In linguistics comprehension studies, the surface feature including semantic relations with concepts (e.g., nouns) in text is also argued to characterize the shape of linguistic representations as re-represented mental models (Fodor, Bever, & Garrett, 1974; Katz & Postal, 1964).

Second, the structural feature describes the levels of size, complexity, and cohesiveness of mental models as a whole. That is, the structural feature indicates a deep level in terms of a well-organized knowledge structure within a particular context in which underlying causal principles, including key variables and their relations, are subsumed (Bransford & Johnson, 1972; Gentner & Medina, 1998; Katz & Postal, 1964; Kintch & von Dijk, 1978).

The third indicator is the semantic feature that shows the levels of understanding of concepts and their relations in a knowledge structure. While the surface and structure feature inform us of generic information of a whole structure (e.g., the number of concepts, the complexity of a knowledge structure), the semantic feature is related to individual concepts and propositional relations of a particular pair of concepts. In particular, the semantic feature includes principle variables that are believed to emerge from information integrated from the whole structure (Katz & Postal, 1964; Kintch & von Dijk, 1978).

Stages of learning progress. Stages of learning progress in problem-solving situations can be explained by studies of the development of expertise. In addition, the three features of knowledge structures (3S) are able to characterize the knowledge structure of each stage of learning progress.

Learners proceed through a similar progression of qualitatively distinct stages in both the short term and the long term (Siegler et al., 2009; Vosniadou et al., 2008; Vygotsky, 1934, 1978; Werner, 1957). In accord with that argument, current studies of the development of expertise focus more on expertise evolving through learning and instruction in academic domains (e.g., Alexander, 2003, 2004; Chi, 2006). In that perspective, Author (2011a) characterized the stages of learning progress as drawing on five stages of expertise development (Dreyfus & Dreyfus, 1986). The stages of learning progress in this model can also be defined by the characteristics of knowledge structures, the 3S features of knowledge structure (see Table 5.1).

Table 5.1

Knowledge Structures of the Stages of Learning Progress

Stage	Three Features of Knowledge Structure (3S)
Novice	(a) All features (surface, structure, and semantic) are quite dissimilar to those of experts; or (b) the structure feature could be seen to be mastered because mental models consisting of a small number of concepts and relations are likely to look cohesive and connected.
Advanced beginner	(a) Knowledge structures have similar surface features with those of expert models but not with structure and semantic features; or (b) there is a high similarity of semantic features but dissimilarity in surface and structure features between a student model and an expert model.
Competent Learner	(a) Structure feature shows enough complexity along with a proper surface feature, which is not necessary to guarantee a semantic fit, however; and (b) the other consists of an appropriate number of contextual and principle concepts (surface and semantic), but that are not well-structured in a proper manner (structure).
Proficient Learner	(a) Knowledge structures are well-featured at all levels (surface, structure, and semantic); or (b) significant number of principles (semantic) creates a cohesive structure (structure) but with a small total number of concepts (surface).
Intuitive Expert	Intuitive decision-making takes place as an advanced semantic structure; unlikely to theorize its measurable structure at this point.

Note. Summarized based on Author (2011a).

The *novice* is a stage in which a learner is new to a domain and has little prior knowledge. A beginner starts to learn context-free abstract knowledge and face situations in a new domain. Knowledge structures at this stage are assumed to show the following characteristics: (a) the lack of all three features of knowledge structure characterizes his or her mental model; or (b) due to a small number of concepts and relations, the mental models looks highly structured (high structure feature with low surface and semantic feature).

The *advanced beginner* stage indicates a mental structure in which a learner recognizes an adequate amount of situational and non-situational knowledge, but the knowledge is still compartmentalized. Ill-structured knowledge results in a lack of sense of what is important in a particular situation. It is anticipated that there are two types of knowledge structure: (a) a common observation would be an adequate number of concepts and relations (i.e., surface feature) but with a lack of structure and semantic features; or (b) affected by instruction in key concepts, the learners internalize a number of key concepts (good-fit in semantic feature) but not well-organized with a proper number of contextual concepts (poor-fit in surface and structure feature).

In the *competent learner* stage, the learners become familiar with given problem situations and then identify concepts underpinning the situations. However, the mental structure is not fully organized, so the mental state may be inadequate in helping him or her to develop a proper solution. With increasing experience, (a) a learner develops complex knowledge structure in which they are likely to determine which elements of a situation are critical (good-fit in surface and structure feature but not sure of good semantic feature); or (b) most key concepts are set in a learner's mental model, but the propositional relations among concepts are somewhat different from the expert model (good-fit in surface and semantic feature but poor-fit in structure feature).

The next stage is the *proficient learner* stage. Learners in this stage conceptualize a best-fit problem space which meets the real traits of the given problem situation. The probability of resolving a problem may markedly increase in the proficient learner stage. A learner approaches a problem holistically and immediately recognizes a problem situation. Two types of knowledge structures are assumed: (a) the expected concepts and relations are represented as organized properly (good-fit in all features); or (b) the learners represent relatively a small size of but efficient knowledge structure in which sufficient key concepts are well-structured (good-fit in structure and semantic feature but not in surface fit). That knowledge structure is accordant with the claim that sometimes experts create mental models having an ‘optimal’ rather than ‘maximum’ number of concepts and relations that are very efficient (Glaser, Abelson, & Garrison, 1983; Glaser, 1992).

The final stage is the *intuitive expert* level. Using a vast repertoire of situational cases, a problem solver is able to make a more subtle discrimination and then intuitively decide how to act while solving a problem. That is, intuitive decision-making characterizes the stage. However, a set of measureable features of a knowledge structure in the expert level are not identified yet, which render diagnosis of the intuitive expert level challenging. In addition, formal instruction may help novices at most become a proficient learner because intuitive experts may be fostered more by experiences both before and after instruction.

Exploring Measures for Cognitive Changes in Learning Progress

A theoretical framework of learning progress has been discussed based on the theory of mental models as structural knowledge. Admittedly, the human mind is not easily observable but may be indirectly inferred from externalized representations such as written responses to a problem situation. The study in this section explores a list of possible measures derived from external representations such as concept maps so as to use those measures for diagnosing cognitive changes.

Parameters for the 3S Knowledge Structure

The three features (3S: Surface, Structure, and Semantic) of a knowledge structure can be quantified by parameters obtained from a concept map that reflect an individual’s mental model (see Table 5.2). Concept mapping is a method that elicits cognitive representations of an individual's structural knowledge in a domain which is constituted with concepts and relations (Axelrod, 1976; Clariana, 2010; Narayanan, 2005; Novak & Canãs, 2006; Spector & Koszalka, 2004).

Table 5.2

Parameters of the 3S Feature of Knowledge Structure

Measure	3S	Technical Definition[a]	Operationalization
M1. Concept	Surface	The total number of nodes (vertices)	The overall number of concepts[b]
M2. Relation	Surface	The total number of links (edges)	The overall number of relations of paired concepts[b]
M3. Average Degree	Structure	The average number of links of a node ranging from 0 and *g*-1 (*g* is the total number of nodes)	As the number of incoming and outgoing relations grows, the complexity of the cognitive structure is taken as more increases[b].
M4. Density	Structure	The density of a graph is the proportion of possible lines that are actually present in the graph.	The density of a concept map indicates how cohesive a concept map is.
M5. Mean Distance	Structure	The average geodesic[c] distance between any pair of nodes in a network	An indicator represents how close the concepts are to one another.
M6. Diameter	Structure	The length of the largest geodesic between any pair of nodes (1 to *g*-1)	It represents how broad the understanding of a domain is[b].
M7. Clustering Coefficient	Structure	The clustering coefficient[d] of the entire network is the average of the clustering coefficients for all the nodes.	This parameter indicates the extent to which a cognitive structure is clustered.
M8. Connectedness	Structure	This measure is to calculate ratio of pairs; it can be reached mutually each other in the graph.	This parameter describes the extent to which the concepts are connected[b].
M9. Cohesive subgroups (= n-clan)	Structure	Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties.	It is assumed that more complex cognitive structure has more numerous subgroups that intermediate the whole connection[b].
M10. Centrality	Semantic	A particular node might be able to control interactions between pairs of other nodes in the network.	Principle concepts might be able to control connections between pairs of other concepts in the network.

Note. a. Wasserman and Faust (1994)
b. Those parameters were also introduced by Ifenthaler (2010).
c. A shortest path between two nodes is referred to as a geodesic.
d. This is a ratio of (the number of connections observed) to (the number of the maximum possible connections) between its neighbor nodes.

Concept maps are visually represented through a set of network analysis techniques mostly involved in graph theory (Rupp, Sweet, & Choi, 2010; Schvaneveldt, Durso, Goldsmith, Breen, & Cooke, 1989; Wasserman & Faust, 1994). Ifenthaler (2010) introduced parameters of graph theory applicable to educational diagnostics of knowledge representation. This study extended the number of potential parameters derived from network analysis methods (Coronges, Stacy, & Valente, 2007; Wasserman & Faust, 1994) and then related the parameters to the 3S knowledge structure. 10 parameters are assumed to portray a knowledge structure. Table 5.2 describes the measures listed below:

- M1 (*Concept*) indicates the overall number of concepts involved in a concept map.
- M2 (*Relation*) indicates the overall number of relations of paired concepts.
- M3 (*Average Degree*) means the average number of relations incoming to or outgoing from a concept that ranges from 0 to g-1 (g is the total number of nodes); As average degree grows, the knowledge structure is taken as more complex.
- M4 (*Density*) is the proportion of possible links in a graph that indicates how much a concept map is cohesive.
- M5 (Mean Distance) is an indicator representing how close the concepts are to one another in a concept map.
- M6 (*Diameter*) indicates how broad the understanding of a knowledge structure is.
- M7 (*Clustering Coefficient*) indicates the extent to which a cognitive structure is clustered.
- M8 (*Connectedness*) describes the extent to which the concepts are connected in a network.
- M9 (*Cohesive Subgroups*) subgroups are subsets of actors among whom there are

relatively strong, direct, intense, frequent, or positive ties; it is assumed that more complex cognitive structure has a more number of subgroups that intermediate the whole connection.

- M10 (*Centrality*) is used to identify key concepts in a concept map based on an assumption that a particular node might be able to control interactions between pairs of other nodes in the network.

Amongst 10 parameters, two parameters (M1. *Concept* and M2. *Relation*) are related to the surface feature, seven parameters (M3. *Average Degree*, M4. *Density*, M5. *Mean Distance*, M6. *Diameter*, M7. *Clustering Coefficient*─*Gamma*, M8. *Connectedness*, and M9. *Cohesive subgroups*) are associated with the structure feature, and one parameter (M10. *Centrality*) is connected to the semantic feature.

Similarity Measures as Indicators of Cognitive Change

Similarity measures. Evaluation of a student's concept map is often done by comparison with a reference model, which is usually elicited from an expert (Curtis & Davis, 2003; Goldsmith & Kraiger, 1997; Coronges et al., 2007; Taricani & Clariana, 2006). Comparison between concept maps is indicated by similarity measures assessed by overlaying network patterns with the concept map information (Coronges et al., 2007; Monge & Contractor, 2003). Similarity measures at each measurement occasion can indicate a level of closeness of a learner model to a reference model. It is assumed that an individual's learning trajectory can be monitored by similarity measures implemented multiple times in a longitudinal manner. This study proposes 13 similarity measures applicable to the study of cognitive change (see Table 5.3).

Table 5.3

Similarity Measures

Similarity Measure	Definition	3S Feature[a]			T-MITOCAR[b]
		S1	S2	S3	
1. Number of concepts	Compare the number of concepts (nodes) between two models	○			
2. Number of relations	Compare the number of links (edges)	○			Surface
3. Average Degree	Compare the average number of degrees		○		
4. Density of graphs	Compare the density of two models		○		
5. Mean Distance	Compare the mean distances of two models		○		
6. Diameter	Compare the largest geodesics of two models		○		Graphical matching
7. Clustering coefficient	Compare the clustering coefficient (%) of the two entire networks		○		Gamma
8. Connectedness	Compare the ratios of pairs that reach each other in each graph		○		
9. Subgroups	Compare the number of cohesive subgroups in each graph		○		
10. Concept matching	Compare semantically same concepts including both contextual and principle variables			○	Concept Matching
11. Principle Matching	Compare fully identical principle concepts			○	
12. Propositional Matching	Compare fully identical propositions (edges) between two concept maps			○	Propositional Matching
13. Balanced Semantic Matching	Compare the balances calculated by dividing Propositional Similarity with Concept Similarity			○	Balanced Semantic Matching

Note. a. It is assumed that each observed similarity is explained by some of three latent attributes of the concept map, such as Surface fit (S1), Structural fit (S2), and Semantic fit (S3).
b. Some similarity measures were matched with or cited from T-MITOCAR similarity measures (Pirnay-Dummer & Ifenthaler, 2010).

As described in Table 5.3, the similarity measures suggested in this study involve those obtained by comparisons of the 10 parameters introduced as indicators of 3S features of a knowledge structure as well as three measures used in the Text Model Inspection Trace of

Concepts and Relations (T-MITOCAR) (Pirnay-Dummer & Ifenthaler, 2010). The measures adopted from T-MITOCAR are: concept matching (that compares semantically same concepts including both contextual and principle variables), propositional matching (that compare fully identical propositions—edges—between two concept maps), and balanced semantic matching (compares the balances calculated by dividing Propositional Similarity with Concept Similarity).

Each of the similarity measures is considered indicating at least one of the 3S of knowledge structure. Amongst 13 similarity measures, on the whole two (*number of concepts*; and *number of relations*) belong to the surface, seven (*average degree*; *density*; *mean distance*; *diameter*; *clustering coefficient*; *connectedness*; and *subgroups*) pertain to the structure, and four (*concept matching*; *principle matching*; *propositional matching*; and *balanced semantic matching*) are related to the semantic. This association implies that gaining a high value of a similarity measure requires a good match with feature(s) of the knowledge structure associated with the measure.

Modification of similarity measures. This study modified and adjusted the formulas that calculated the similarity measures. As to numerical similarity, amongst the similarity measures, density, mean distance, and clustering coefficient measures used the original formula because an optimal value indicates a good condition for those three measures rather than a greater value (refer to Table 5.2):

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

where f_1 and f_2 denote the numerical frequency of each method compared. The similarity ranges from 0 to 1, $0 \leq s \leq 1$.

On the whole, a similarity formula assumes each part of a pair is equally significant. In the case of a concept model comparison, the reference model and student model are not equal in terms of maturity. A reference model acts as criteria and a student model is expected to progress toward the reference model. It is assumed that a reference model is likely to contain a greater number of concepts and relations and is comprised of a larger knowledge structure than a novice model (Chi, Glaser, & Farr, 1988; Spector & Koszalka, 2004). Thus, a modified algorithm was applied except for the density, mean distance, and clustering coefficient similarity. In case f_1 is smaller than f_2, $f_1 < f_2$, the original numerical similarity formula was used so that:

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

where the frequency of a student model is f_1 and that of a reference model is f_2. Otherwise, if f_1 is not less than f_2, $f_1 \geq f_2$, the similarity value was set as '1' because the student value is greater than that of the reference. That is, it indicates that the student model exceeds the reference model according to the relevant criteria.

Similarly, concerning the conceptual similarity as applied to the four similarity measures (concept, principle, propositional, and balanced semantic matching score), an adjustment was made. Just as a picture resembles an object rather than an object resembles a picture of it, a student model to some degree resembles the reference model that is more salient. In this asymmetric relation, the features of the student model are weighted more heavily than those of the reference (Colman &Shafir, 2008; Tversky & Shafir, 2004). When the conceptual similarities were calculated by Tversky's (1977) formula, α was weighted more heavily than β (α = 0.7 and β = 0.3).

$$s = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A - B) + \beta \cdot f(B - A)}$$

Validation Methods

Participants

Participants included 136 undergraduate students and seven experts. The students were enrolled in a course at a university in the southern United States. The course aimed to educate students on knowledge and skills for integrating technology in teaching and learning. In the class, students made written responses to a specific complex problem. Female students formed 83% of the study participants, while 17% were male. 7% were freshman, 39% were in their junior year, and sophomore and senior levels occupied 27% each.

Seven experts were professors teaching at a major university in the United States. They participated in a Delphi survey to obtain a reference model. The panel members were selected based on pre-set criteria: (a) professors in Instructional Technology or related fields; (b) professors teaching a course titled Instructional Design or Technology Integration in Learning; (c) professors who research technology-integration in classroom learning; and (d) professors whose doctorates were received at least three years ago.

The Problem-solving Task

All participants were asked for responses to a complex problem situation using natural language. The task provided a simulated situation in which students were assumed to be participating in an evaluation project, the purpose of which was to investigate an unsuccessful project that had as its goal adapting a technology (i.e., a tablet PC) for classroom teaching. In order to elicit students' knowledge in detail, the questions asked them to explicitly describe the concepts, issues, factors, and variables likely to have contributed to the result that the introduction of tablet PCs had very little effect on the instructional practices employed in the classes.

Reference Model

This study includes a reference model to obtain similarity measures for student concept maps. The reference model was created according to the Delphi survey procedures (Goodman, 1987; Hsu & Sandford, 2007; Okoli & Pawlowski, 2004). The Delphi survey included three rounds to refine the reference model. In the first round, the participating experts created their own responses to the problem and all responses from the panel were consolidated. After that, a document including all statements and a list of identified concepts was sent to the panel again. The experts were asked to add their comments on the listed statements and concepts and rank them. After gathering the second surveys, the researcher created a final list of ranked statements and concepts. Based on the summary, a draft of a reference model was created. In the final round, the results of the second survey were sent to the panel and revised according to their comments, if needed. Throughout these procedures, a written reference model containing the 23 key concepts identified by the panel was developed.

Data Collection

This study gathered manifest indicators resulting from a network analysis approach that consists of a three-step procedure (Curtis & Davis, 2003; Taricani & Clariana, 2006): (a) elicit judgments about concept relationships; (b) construct concept maps; and (c) compare the concept maps to a reference model.

The first step is to elicit judgments about concept relations in students' written responses to the problem. This study made judgments in compliance with Author's (2011a) Semantic Relation (SR) approach which argues that semantic relations together with concepts obtained from a written response create a concept map which is more likely to represent a student's internal mental representation.

The second step is to construct concept maps. This study employed the network analysis software, NetMiner (http://www.netminer.com/) in order to visualize concept maps and generate concept map parameters. This tool automatically rendered the parameters of a concept map suggested as indicators for educational diagnostics.

The final step is to compare student concept maps to a reference model so that we can obtain similarity measures that are eventually used for the validation study. The aforementioned 13 similarity measures were calculated using the similarity tool developed by Author (2011b). The author (2011b) developed the tool using C++ programming language, and the tool was validated by random comparisons between tool-generated and manually calculated data.

Similarity measures are continuous variables ranging from 0 to 1. For the following validation analyses, the similarity measures were transformed to discrete variables (exactly to the dichotomous variables). The transformation of similarity variables can involve two conditions: one is a generous condition, where a value over 0.5 is transformed to a dummy value ("1" indicates that the subject's parameter is similar to that of the reference model), and the other is a rigorous condition, where similarity is decided if a value is over 0.75. A cut point of 0.5 in this study was applied considering that the measurement was implemented at the earlier time of the course at which students were unlikely to be familiar with the topic.

Latent Class Analysis for the Model of Learning Progress

Stages in learning progress are inferred rather than directly observed. Accordingly, qualitative stages of the learning progress can be labeled latent classes because of their psychometric characteristics. Latent class modeling (LCM) is a method used to test the hypothesized latent categorical variables against the theorized latent classes (Heinen, 1996; Kaplan, 2008). In LCM there are corresponding latent classes associated with the other stages in the model so that at any given occasion of measurement each individual has an array of latent class memberships (Collins & Cliff, 1990; Collins, Graham, Rousculp, Fidler, Pan et al., 1994; Rost & Langeheine, 1997).

This study considers latent class models adaptable to test the latent stages of learning progress. The main assumption is that a learner belongs to a certain stage of learning progress that is cumulative but exclusive in such a way that discrete ordinal latent classes (a learner's status of mental model progress) range from novice to proficient. Considering that there are a variety of LCMs, to validate the model of learning progress this study located a best-fit model that is illustrated in the following section.

Log-linear Classification Diagnostic Model (LCDM). In general, LCMs are used to determine whether the hypothesized model explains the number of latent classes in the data. Restricted latent class models make it possible to test the hypotheses of the structure of a measurement model and place fewer demands on the data (Heinen, 1996). Log-linear models, in the context of categorical data analysis, allow latent class models to place linear restrictions on the log-linear parameters (Agresti, 2007).

Diagnostic Classification Models (DCMs), as a special case of constrained LCMs, are "multidimensional latent variable models with multiple latent variables" (Rupp et al., 2010, p. 149). DCM includes the latent variables, or *attributes*, through which DCMs define an individual's ability as the probability of a correct response (Henson, Templin, & Willse, 2009). According to Templin (2004, p. 8): "commonly, attributes are the atomic components of ability, the specific skills that together comprise the latent space of general ability." Ultimately, the attributes that have or have not been mastered define an individual's mastery profile. DCMs have a set of attribute mastery profiles in that the set of K latent attributes can be considered a latent class model with 2^k classes (Henson et al., 2009; Templin, 2004).

In the context of the model of learning progress, the three features of knowledge structure (K = 3 latent attributes: surface, structure, and semantic) can be the attributes in DCMs. Each feature (attribute) of knowledge structure can be profiled as obtained (1) or not obtained (0). Those mastery profiles provide stages (i.e., $2^3 = 8$ latent classes) of learning progress at which an individual is classified. The latent classes associated with the stages of learning progress are detailed in Table 5.4.

Table 5.4

Class-to-Profile Table

Latent Class	3S Attribute [a]			Stage of Learning Progress [c]
	S1	S2	S3	
Class 1 (C_1)	0[b]	0	0	Novice (L1)
Class 2 (C_2)	0	0	1	Advanced Beginner (L2)
Class 3 (C_3)	0	1	0	Novice (L1)
Class 4 (C_4)	0	1	1	Proficient Learner (L4)
Class 5 (C_5)	1	0	0	Advanced Beginner (L2)
Class 6 (C_6)	1	0	1	Competent Learner (L3)
Class 7 (C_7)	1	1	0	Competent Learner (L3)
Class 8 (C_8)	1	1	1	Proficient Learner (L4)

Note. a. 3S attributes involve the three features of knowledge structure—Surface (S1), Structural (S2), and Semantic (S3). b. 0=absent and 1=present. c. The stages of learning progress use the first four stages labeled in Table 5-1, originally defined by Dreyfus and Dreyfus (1986).

For instance, class 4 has its mastery profile (S1: 0, S2: 1, and S3: 1) that is matched to the proficient learner (L4) level. This match was determined by the theoretical assumption of the knowledge state of the level: The learners represent relatively a small in size (S1: 0 = surface feature is absent) but efficient knowledge structure (S2: 1 = structure feature is present) in which sufficient key concepts are well-structured (S3: 1 = semantic feature is present) (good-fit in structure and semantic feature but not in surface fit, as discussed in Table 5.1).

LCDMs are log-linear models used to represent DCMs (Rupp, Templin, & Henson, 2010). LCDM requires all item responses and all attributes to be dichotomous. A key assumption of LCDM is the local independence (or, in other words, conditional independence) which means that "the responses of respondents are independent given the number of discrete latent variables included in the model, which create the latent classes" (Rupp et al., 2010, p. 325). This assumption is congruent with the conjecture of qualitatively different stages of learning progress in that the features of knowledge structure (attributes) capture the state of mental stages.

Results

This section provides validation results that involve (a) preliminary review of similarity measures; (b) selection of similarity measures; (c) validation of the three features of knowledge structure (the 3S knowledge structure model); and (d) validation of the model of stage-sequential learning progress using LCDMs. For the first three analyses, the original similarity measures ($0 \leq s \leq 1$) as continuous variables were used and for the LCDMs, the transformed similarity measures as binary variables (0 = not mastered, 1= mastered).

Data Review

A pre-analysis data inspection that included examination of descriptive statistics and correlations was conducted. As Table 5.5 describes, most similarity measures showed biased distributions. The similarity measures M10 to M12 were distributed within the lower areas of similarity band, ranging from 0 to 0.68 with lower means ranging from 0.05 to 0.24, while the distributions of the measures M3, M5, M6, to M8 stayed within higher band ranging from 0.15 to 0.99 with means ranging from 0.73 to 0.82. Other measures such as M1, M2, M4, and M9 had means between 0.31 and 0.35. Although some measures (M4. density, M7. clustering coefficient, and M9. subgroups) classified into the structure feature (see Table 5.3) showed trends more like those in the surface feature, the descriptive results implied that the similarity measures likely indicate different constructs that are assumed to be the features of knowledge structure.

Table 5.5

Descriptive Statistics of the Similarity Measures

	N	Minimum	Maximum	Mean	SD
M1. Concept	143	0.04	0.96	0.35	0.18
M2. Relation	143	0.02	0.98	0.31	0.20
M3. Average Degree	143	0.42	1.00	0.82	0.14
M4. Density	143	0.05	0.96	0.40	0.19
M5. Mean Distance	143	0.28	0.99	0.74	0.17
M6. Diameter	143	0.14	1.00	0.76	0.22
M7. Clustering Coefficient	143	0.00	0.98	0.35	0.35
M8. Connectedness	143	0.15	1.00	0.73	0.25
M9. Subgroups	143	0.00	1.00	0.31	0.20
M10. Concept Matching	143	0.05	0.57	0.24	0.08
M11. Principle Matching	143	0.00	0.38	0.12	0.07
M12. Propositional Matching	143	0.00	0.28	0.05	0.04
M13. Balanced Matching	143	0.00	0.68	0.18	0.15

Next, as shown in Table 5.6, correlations of the similarity measures were calculated. Concepts (M1) had a very high correlation with the relations (M2), $r = 0.97$, $p < .01$.Due to that strong correlation, it was determined to only use the relations measure (M2), indicating the surface feature of knowledge structure. Amongst the measures classified in the structure feature (see Table 5.3), clustering coefficient (M7) and connectedness (M8) were removed because they had a low correlation with other measures in the same category (average $r = 0.28$ and 0.26, respectively). Remaining measures were retained because they mostly moderately related with one another, ranging from 0.38 to 0.85, and discriminated from the measures (M10 to M13) within the semantic feature category. The retained measures were: average degree (M3), density (M4), mean distance (M5), diameter (M6), and subgroups (M9). As for the measures within the semantic feature category, even though the balanced semantic matching (M13) had low correlations with concept matching (M10) and principle matching (M11), it was retained to have enough number of measures for the construct, semantic feature of knowledge structure.

Table 5.6

Correlations of the Similarity Measures

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
M1	1												
M2	.97**	1											
M3	.60**	.71**	1										
M4	.95**	.85**	.38**	1									
M5	.64**	.62**	.65**	.59**	1								
M6	.66**	.62**	.62**	.63**	.95**	1							
M7	.36**	.41**	.55**	.25**	.31**	.31**	1						
M8	0.00	0.15	.51**	-.22**	.35**	.21*	.18*	1					
M9	.92**	.92**	.63**	.85**	.71**	.73**	.28**	0.14	1				
M10	.52**	.54**	.47**	.44**	.41**	.43**	.30**	0.06	.48**	1			
M11	.26**	.30**	.35**	.17*	.24**	.25**	.18*	.17*	.27**	.45**	1		
M12	.37**	.42**	.35**	.27**	.20*	.21*	.23**	0.13	.36**	.61**	.32**	1	
M13	0.16	.20*	.22**	0.09	0.07	0.07	0.13	0.11	0.15	.27**	0.15	.87**	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

In particular, the relations (M2) measure was highly related to the average degree (M3), Density (M4), and subgroups (M9), $r > 0.9$, $p < .01$. Those correlations implied that the three structural measures (M3, M4, and M9) are possibly explained by the construct, surface feature, of relations (M2). That conjecture was justified by the theoretical assumption that three structural measures (M3, M4, and M9) necessitate adequate surface feature. For example, higher density requires an adequate number of relations (links) in the graph.

Validation of the 3S Knowledge Structure

As a part of the preliminary review, the three-factor model of knowledge structure that was composed of the surface, structure, and semantic feature was investigated using CFAs (Confirmatory Factor Analyses). For the CFAs, M-plus software was used. A multivariate normality test showed univariate and multivariate distributions were significantly non-normal. To deal with this violation of test assumption, MLM (Maximum Likelihood Method) estimation was applied based on a suggestion in the M-Plus manual. Based on a significance level of .05, three cases were determined to be problematic outliers and thus were removed from the data. Finally, the total number of samples became 140 for the CFAs.

Figure 5.1 depicts the three-factor CFA model. The validation procedure includes sequential evaluation of a single factor model with the proposed factor model. A single factor model provided poor-fit indices (CFI < 0.90, NNFI < 0.90, RMSEA > 0.05), while overall fit for the three-factor was good (CFI > 0.90, NNFI > 0.90) (see Table 5.7).

Table 5.7

Summary of Fit Indices

Models	$\chi^2(df)$	$\chi^2/(df)$	CFI	NNFI	RMSEA	SRMR	AIC	ABIC
Single factor	699(35)	19.90	0.49	0.34	0.36	0.14	-2518	-2430
3 factor	96(26)	3.69	0.94	0.90	0.14	0.18	-3157	-3042

Note. Indices (their expected values) are: $\chi^2(df)$ = Chi-square statistics and degrees of freedom for test of model fit (equal to 0); $\chi^2/(df)$ = the ratio of Chi-square statistic to the degrees of freedom (equal to 1), CFI = comparative fit index(> 0.90), NNFI = non-normal fit index (a.k.a., Tucker-Lewis index)(> 0.90), RMSEA = root square error of approximation (<0.05), SRMR = standardized root mean square residual (< 0.05), AIC (Akaike Information Criterion (close to 0, smaller the better), ABIC (Adjusted Bayesian Information Criterion) (close to zero, smaller the better)

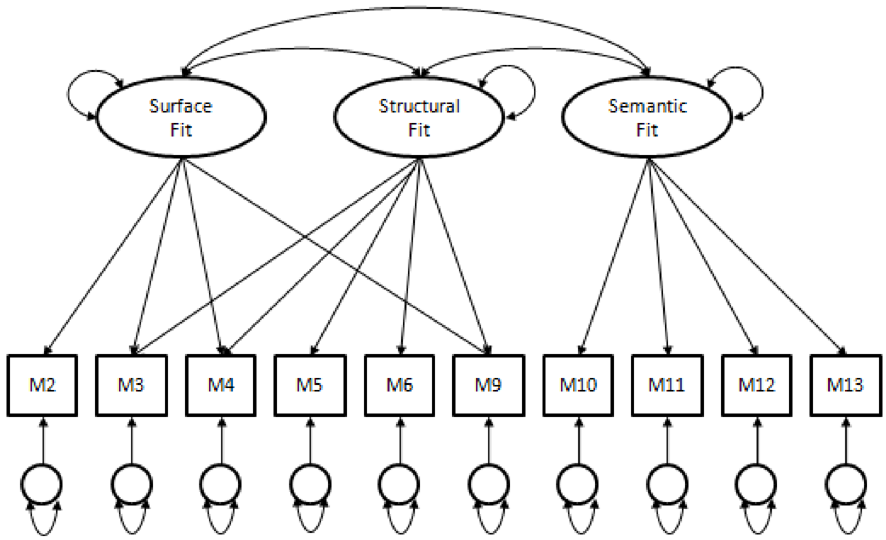


Figure 5.1. Three-factor CFA model. 10 similarity measures were used as items in the model: M2 (Relations); M3 (Average degree); M4 (Density); M5 (Mean distance); M6 (Diameter); M9 (Subgroups); M10 (Concept matching); M11(Principle Matching); M12 (Propositional Matching); and M13 (Balanced semantic matching).

LCDMs for the Model of Learning Progress

So far it was identified that: (a) the similarity measures as continuous variables indicate a set of constructs like the 3S knowledge structures; (b) their trends somewhat differ from one another depending on the traits of the constructs by which the measures are explained; and (c) the proposed 3S knowledge structures are present in the data. However, the findings were not able to provide comprehensive information to diagnose an individual's learning stages. LCDMs were employed as confirmatory methods for the diagnosis of learning progress.

Q-matrix. It is important to note that LCDM requires a substantive theoretical model so that researchers can interpret statistical classifications as meaningful latent classes. The hypothetical model in LCDM is called Q-matrix, which defines a limited relationship between a set of attributes and a set of test items (Templin, 2004). It is imperative to specify the Q-matrix applicable to the context of the model of learning progress. In addition, notably, LCDM requires all item responses and all attributes to be dichotomous.

Following the results from the correlations and CFA analyses, ten similarity measures were selected and transformed as binary variables. Those transformed similarity measures and the three features of knowledge structure are defined as the items and the attributes in the LCDMs, respectively. The Q-matrix postulates that correctly answering each item (i.e., the subject's model value is similar to that of the reference model) requires mastering the designated attributes (that is to say, obtaining the required knowledge features)). Table 5.8 shows the hypothesized relationships between the ten measures (items) and three features of knowledge structure (attributes). In the table, it is coded '1' when the attribute is required to correctly answer the item. For example, correctly answering the balanced semantic matching is assumed to require the semantic feature (refer to Table 5.3 and Figure 5.1). It is notable that the patterns of mastering attributes define an individual's latent classes that were listed in Table 5.4.

Table 5.8

Similarity Measures Associated with Attributes of Knowledge Structure (Q-Matrix)

Similarity Measure	3S Attribute [a]		
	S1	S2	S3
I1:Number of relations	1	0	0
I2:Average Degree	1	1	0
I3:Density of graphs	1	1	0
I4:Mean Distance	0	1	0
I5:Diameter	0	1	0
I6:Subgroups	1	1	0
I7:Concept matching	0	0	1
I8:Principle Matching	0	0	1
I9:Propositional Matching	0	0	1
I10:Balanced Semantic Matching	0	0	1

Note. a. the 3S knowledge attributes: surface (S1), structure (S2), and semantic (S3)

Evaluating the Model of DCMs. The transformed binary similarity measures were investigated. The data review showed that all cases of item 8 and item 9 were zero (see Table 5.9). Thus, only eight measures excluding these two were applied to the analysis of DCMs.

Table 5.9

Descriptive Statistics of the Transformed Similarity Measures

	N	Minimum	Maximum	Mean	SD.
I1. Relation	143	0.0	1.0	0.17	0.38
I2. Average Degree	143	0.0	1.0	0.97	0.17
I3. Density	143	0.0	1.0	0.22	0.42
I4. Mean Distance	143	0.0	1.0	0.91	0.29
I5. Diameter	143	0.0	1.0	0.86	0.35
I6. Subgroup	143	0.0	1.0	0.20	0.40
I7. Concept Matching	143	0.0	1.0	0.01	0.08
I8. Principle Matching	143	0.0	0.0	0.00	0.00
I9. Propositional Matching	143	0.0	0.0	0.00	0.00
I10. Balanced Matching	143	0.0	1.0	0.02	0.14

The evaluation was conducted for multiple candidate DCMs so as to determine a good-fit model. M-plus software was used for the analysis of DCMs. The models included the DINO and DINA model. The *deterministic-input, noisy-or-gate* (DINO) model (Templin & Henson, 2006) assumes that a correct/positive answer for item *i* is expected when a respondent masters at least one of the required abilities (attributes), while the *deterministic-input, noisy-and-gate* (DINA) model (de la Torre & Douglas, 2004) allows the increasing probability of correct answers for item *i* only if a respondent has mastered all required attributes.

As Table 5.10 shows, The DINA model fit better than the DINO model considering that lower values are desirable in both AIC and BIC. In both of the two models, the three-way interaction effect parameter between attributes 1, 2, and 3 had no variance with *SE* = 0. Thus, the parameter was removed from the models to reduce the complexity of the models. Additional

DINO and DINA models without the three-way effect were evaluated. As a result, the DINA model without the three-way interaction was determined to have the best-fit indices.

Table 5.10

Comparison of Relative Fit of DCMs

Model	Number of Parameters	AIC	BIC
DINO	23	639.485	707.630
DINO without 3 way interaction	22	637.485	702.668
DINA	23	551.833	619.978
DINA without 3 way interaction	22	549.935	615.118

Note. *Akaike's information criterion* (AIC); *Bayesian information criterion* (BIC)

X^2 and G statistic for the eight-item diagnostic assessment with the DINA model without the three-way interaction effect returned values with a *p*-value of 1.000, supporting the null hypothesis that the model fit the data well. The selection of the DINA model requiring the mastery of all attributes was in accord with the theoretical assumption that good structure (attribute 1) necessitates sufficient surface feature (attribute 2).

The estimated latent class membership. M-plus reports the estimated posterior probabilities and the most likely latent class for each respondent. Table 5.11 describes the posterior probability results in terms of final counts and proportions for the latent classes.

Two latent classes such as class 5 and class 6 were not present in both estimated counts and proportion. The posterior values of the latent classes were summed as each of the stages of learning progress according to the earlier theoretical discussion about knowledge features (attributes) associated with the stages. Differing from the measurable four stages suggested in the theoretical framework, in this analysis, the advanced beginner stage was not present and the proportion of the competent learner stage was extremely small. Thus, this model analysis concluded that two stages—novice and proficient learner—have a high probability to be present during a problem-solving situation. Referring to the estimated latent class membership, among the participants, 113 individuals belonged to the novice and 30 participants the proficient learner category. All seven experts were classified in the proficient learner level.

Table 5.11

The Estimated Final Class Counts and Proportions

Stage of Learning Progress	Latent Class (profile)	Counts	Proportion	Sum of Counts	Sum of Proportion
Novice	Class 1 (000)	19.9991	0.13986	110.86033	0.77525
	Class 3 (010)	90.86123	0.63539		
Advanced Beginner	Class 2 (001)	0.00009	0.00000	0.00009	0.00000
	Class 5 (100)	0.00000	0.00000		
Competent Learner	Class 6 (101)	0.00000	0.00000	0.00152	0.00001
	Class 7 (110)	0.00152	0.00001		
Proficient Learner	Class 4 (011)	0.00002	0.00000	32.13725	0.22474
	Class 8 (111)	32.13723	0.22474		

Discussion

This study attempted to validate the model of stage-sequential learning progress composed of four stages. Contrary to expectations, the data analyses using latent class modeling technique demonstrated that there are probably two stages rather than four stages in the data. Some potential implications of that result are presented here based on three categories: (a) statistical issues; (b) research context issues; and (c) theoretical issues.

First, statistical issues of this study might render the results different from the proposed model. At first, in most assessment situations, attributes are positively correlated. For example, the surface feature can be assumed to be nested in the structure feature. The correlated attributes result in classifying a large number of respondents as having mastered either none or all attributes (i.e., either the novice level or the proficient learner level) (Rupp et al., 2010).

In addition, the relatively small sample size (N = 143) of this study could be problematic. For instance, eight items created $2^8 = 256$ possible response patterns. That setting needs over five respondents at each pattern in an ideal condition (N = 256 X 5 = 1280) following Agresti and Finlay's (1997) suggestion. Small sample size probably led to many cells being sparse and reduced the chance for the two items (items 8 and 9) to have positive responses. Larger sample size and including all ten items in an analysis may provide somewhat more information in terms of the number of stages and their estimated proportions. Further studies with an adequate number of samples are required.

Second, the research context should be taken into consideration. The course was not designed to teach contents directly related to the problem situation (i.e., technology adaptation to classroom instruction). Moreover, the responses were gathered at an early part of the semester. That is, students responded to the problem with little chance of being instructed about the problem contexts. It is possible to argue that at that point there were two groups of students who either had or had not prior experience. There might be a lack of time to develop transitional knowledge structures. For future studies, a better research context can be suggested as involving: (a) providing instruction related to a problem situation prior to the data gathering; and (b) implementing data collection in the middle of an instructional period.

Third, accepting the two-stage model as true, studies of conceptual change can provide some accounts of why only two stages were present. In spite of controversy regarding the process of conceptual change, this study supports that learning is a process of reorganizing the knowledge base as a coherent structure in which concepts are embedded. It is argued that concept change often requires change of the knowledge base as a whole (Chi, 2008; Vosniadou et al., 2008). Conceptual change can be illustrated as a shift through which a student having theoretical framework *A* (a wrong structure) changes to theoretical framework *B* (an expected structure).

A shift from one model to another seems to abruptly happen at a certain point after an initial slower process. The empirical studies claim that radical conceptual changes usally happen at the end state of a slow process (Vosniadou, 2003; Hatano & Inagaki, 1994). Vosniadou and colleagues (2008) contend that the slow and gradual enrichment of knowledge is largely unconscious although the enrichment mechanism leads to conceptual changes in the long run. For example, a recent experiment conducted by Siegler and colleagues (2009) supported the claim that stage transitions suddenly take place after retaining the current model for some time; they called it a logarithmic-to-linear shift. In their experiment, children estimated the position of a number in a line representing the numeric magnitude. Their estimations showed probabilistic patterns, which moved from being stable to approximate to the actual value. This shift occurred abruptly after recurrently taking existing approaches.

In other words, a transition from the novice level to the proficient learner level is likely to abruptly take place at some point during a slow and gradual process of learning and instruction. The middle stages such as the advanced beginner and competent learner levels are probably present but very short moment just before becoming the proficient level. If that situation is true, it would be hard to capture the stages in a measurement occasion.

In summary, considering that this study as an initial attempt suggested a measureable framework of learning progress and tried to validate the model with a data, it is too early to conclude how many stages of knowledge structure can explain the process of learning progress.

In addition, there exist diverse potential theoretical accounts of the mental stages. Therefore, further studies in the conditions described earlier are requested.

Implications

Along with the theoretical model of stage-sequential learning progress, this study provides future research with methods and tools that involve: (a) a set of parameters quantifying traits of a knowledge structure; (b) multidimensional attributes of a knowledge structure; (c) a set of similarity measures applicable to the studies of cognitive changes; and (d) a statistical approach (that is LCDM approach) to diagnose the stages of learning progress. Those methods are applicable to a wide range of areas that include: the studies of conceptual change in a problem-solving situation, linguistic comprehension, the evaluation of scientific argumentations, expertise modeling, and longitudinal studies of learning progress. For example, it is required to study longitudinal stage changes of learning progress so as to evaluate effectiveness of instruction and determine proper educational supports to an individual. Collins and Wugalter (1992) pointed out that psychological research and theory is increasingly turning to longitudinal studies in which development is monitored by following individuals over a period of time suggesting a measurement theory and methods using latent variables.

The ultimate goal of applying LCDMs to the validation of the theoretical model of learning progress is to determine parameters in the LCDM model that can be generalized. In other words, based on the identified parameters, a student’s stages can be easily and quickly estimated using their responses such as similarity values. That diagnostic algorithm can be embedded in an assessment technology. It is essential to devise an assessment technology adapted for the complex, dynamic structure of mental models because assessment is a fundamental unit of instruction providing feedback, revision, and reflection on learning (Pellegrino, Chudowsky, & Glaser, 2001).

References

Agresti, A. (2007). *An introduction to categorical data analysis*. Hoboken, NJ: Wiley Inter-Science.

Agresti, A., & Finlay, B. (1997). Statistical methods for the social sciences (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273-298). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Alexander, P. A., & Murphy, P. K. (1998). Profiling the differences in students' knowledge, interest, and strategic processing. *Journal of Educational Psychology, 90*, 435-447.

Anzai, Y., & Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction, 1*, 397-450.

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton, N.J.: Princeton University Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717-726. doi:16/S0022-5371(72)80006-9

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces*, *70*, 215-225.

Chen, Z., & Siegler, R. S. (2000). Intellectual development in childhood. In R.Sternberg (Ed.) *Handbook of intelligence* (pp. 92-116). New York: Cambridge.

Chi, M.T.H. (2006). Two approaches to the study of experts' characteristics. In K.A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 121-130), Cambridge University Press.

Chi, M.T.H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York: Routledge.

Chi, M. T. H., Glaser, R., & Farr, M. (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Clariana, R. B. (2010). Multi-decision Approaches for Eliciting Knowledge Structure. In *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 41-59). Retrieved from http://dx.doi.org/10.1007/978-1-4419-5662-0_4

Collins, L. M. & Cliff, N. (1990). Using the longitudinal Guttmann simplex as a basis for measuring growth. *Psychological Bulletin*, *108*, 128-134.

Collins, A. & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought*, (pp. 243-265). Cambridge: Cambridge University Press.

Collins L. M, Graham, J. W., Rousculp S. S., Fidler, P. L., Pan, J, & Hansen, W. B. (1994). Latent transition analysis and how it can address prevention research questions. *NIDA research monograph. 142*, 81-111.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, *27*(1), 131-157.

Colman, A. M., & Shafir, E. (2008). Tversky, Amos. In N. Koertge (Ed*.), New dictionary of scientific biography* (Vol. 7, pp. 91-97). Farmington Hills, MI: Charles Scribner's Sons.

Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural Comparison of Cognitive Associative Networks in Two Populations. *Journal of Applied Social Psychology*, *37*(9), 2097-2129. doi:10.1111/j.1559-1816.2007.00253.x

Curtis, M. B., & Davis, M. A. (2003). Assessing knowledge structure in accounting education: an application of Pathfinder Associative Networks. *Journal of Accounting Education*, *21*(3), 185-195. doi:10.1016/S0748-5751(03)00024-1

de Vries. E., Lund, M., & Baker, M. (2002). Computer-mediated epistemic dialogue: Explanation and Argumentation as vehicles for understanding scientific notions. *The Journal of the Learning Sciences. 11*(1), 63-103.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika, 69*, 333-353. doi:10.1007/BF02295640

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction, 13*(5), 533-568. doi:10.1016/S0959-4752(02)00025-7

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. Feltovich, & R. R.

Hoffman, R. R. (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 685-706). Cambridge, UK: Cambridge University Press.

Festinger, L. (1962). Cognitive dissonance. *Scientific American, 207*(4), 93-102.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In W. Damon & D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 851–898). New York: Wiley.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263–297.

Glassner, A., Weinstock, M., & Neuman, Y. (2005). Pupils' evaluation and generation of evidence and explanation in argumentation. *British Journal of Educational Psychology, 75*(1), 105-118.

Goldsmith, T. E., & Kraiger, K. (1997). Applications of structural knowledge assessment to training evaluation. In J. K. Ford (Ed.), I*mproving training effectiveness in organizations* (pp. 73–95). Mahwah, NJ: Lawrence Erlbaum Associates.

Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, *12*(6), 729-734.

Grotzer. T. A., & Perkins, D. N. (2000, April). *A taxonomy of causal models: The conceptual leaps between models and students' reflections on them*. Paper presented at the National Association of Research in Science Teaching (NARST), New Orleans, LA, 28 April - 1 May 2000.

Grow-Maienza, J., Hahn, D., & Joo, C. (2001). Mathematics instruction in Korean primary schools: Structures, processes, and a linguistic analysis of questioning. *Journal of Educational Psychology, 93*(2), 363-376.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, 68* (1), 52-59.

Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition, 50*(1-3), 171-188. doi:10.1016/0010-0277(94)90027-2

Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Taylor & Francis Group.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks: Sage Publication.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika,74* (2), 191-210.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15,* 332–340.

Hsu, C. & Sandford, B. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation, 12*(10). Available online: http://pareonline.net/getvn.asp?v=12&n=10

Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. Educational Technology Research & Development, 58(1), 81-97. doi:10.1007/s11423-008-9087-4.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185-208). Cambridge University Press.

Johnson-Laird, P.N. (2005b). The history of mental models. In K. I. Manktelow, & M. C. Chung (Eds), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179-212). Psychology Press.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge. *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge.*

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential development processes. *Developmental Psychology*, *44*(2), 457-467.

Karoly, L. A., & Panis, C. W. A. (2004). *The 21st century at work: Forces shaping the future workforce and workplace in the United States*. Santa Monica, CA: RAND Corporation. Retrieved February 17, 2010, from http://www.rand.org/pubs/monographs/2004/RAND_MG164.pdf

Katz, J. J., & Postal, P. M.(1964). *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press.

Kieras, D.E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science*, *8*, 255-273.

Author (2011a). ----

Author (2011b). ----

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394. doi:10.1037/0033-295X.85.5.363

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18* (4), 495-523. Retrieved February 15, 2010, from http://www.informaworld.com/10.1207/S1532690XCI1804_3

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research, 59*(1), 43-64.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.

Narayanan, V. K. (2005). Causal mapping: An historical overview. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for research in information technology* (pp.1-19). Hershey, PA: Idea Group Publishing.

National Research Council (NRC). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

National Research Council (NRC). (2003). *Assessment in support of instruction and learning*. Washington, DC: National Academy Press.

National Research Council (NRC). (2005). *Systems for state science assessment*. Washington, DC: National Academy Press.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Novak, J. D., & Cañas, A. J. (2006). *The origins of the concept mapping tool and the continuing evolution of the tool*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.3382

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15-29. doi:10.1016/j.im.2003.11.002

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49,* 301-332.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). *Knowing what students know*. Washinton, DC: National Academy Press.

Piaget, J. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered* (pp. 7-20). Ithaca, NY: Cornell University.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.

Rost, J., & Langeheine, R. (Eds.) (1997). *Applications of latent trait and latent class models in the social sciences*. New York: Waxmann.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment, 8*(4). Available online at http://escholarship.bc.edu/jtla/vol8/4

Rupp, A. A., Sweet, S., & Choi, Y. (2010). *Modeling learning trajectories with epistemic network analysis: A simulation-based investigation of a novel analytic method for epistemic games*. Presented at the annual meeting of the International Society for Educational Data Mining (EDM), Pittsburgh, PA.

Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: The Guilford Press.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12*(1), 5-51.

Schvaneveldt, R.W., Durso, F.T., Goldsmith, T.E., Breen, T.J., & Cooke, N.M. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23,* 699-728.

Schvanevldt, 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* (pp. 249-284). San Diego, CA: Academic Press.

Seel, N.M. (1999). Semiotics and structural learning theory. *Journal of Structural Learning and Intelligent Systems, 14* (1), 11–28.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science, 29* (4-5), 403-427.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition, and Learning, 1* (1), 59-85.

Seel, N. M. (2004). Model-centered learning environments: Theory, Instructional design, and ethics. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instruction design: International perspectives* (pp. 49-74). NJ: Lawrence Erlbaum Associates, Inc.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation, 1*(1), 4-35.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation, 23*(4), 373-398.

Siegler, R. S. (2005). Children's Learning. *American Psychologist, 60*, 769-778.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education, 3*, 143-150.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology, 8*, 165–190.

Shute, V. J., & Zapata-Rivera, D. (2008). Using an evidence-based approach to assess mental models. In D. Ifenthaler, P. Pirnay-Dummer, J. M., Spector (Eds.), *Understanding Models for Learning and Instruction: essays in honor of Norbert M. Seel* (pp. 23-41). New York: Springer.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.

Snow, R.E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14* (5), 455–473.

Stigler, J. W., & Stevenson, H. W. (1991). How Asian teachers polish each lesson to perfection. *American Educator, 15*, 12-20.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains* (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Taricani, E. M., & Clariana, R. B. (2006). A Technique for Automatically Scoring Open-Ended Concept Maps. *Educational Technology Research & Development, 54*(1), 65-82. doi:Article

Templin, J. L. (2004). *Generalized linear mixed proficiency models*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287-305. doi:10.1037/1082-989X.11.3.287

Thomas, R. M. (2005). *High-stakes testing: coping with collateral damage*. Mahwah, NJ: Psychology Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Tversky, A., & Shafir, E. (2004). *Preference, belief, and similarity: selected writings*. London, England: MIT Press.

Vosniadou, S. (2003). Exploring the relationships between conceptual change and intentional learning. In G. M. Sinatra & P. R. Pintrich (Ed.), *Intentional conceptual change* (pp. 377-406). Mahwah, NJ: Erlbaum.

Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3-34). New York: Routledge.

Vygotsky, L. (1934/1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge University Press.

Werner, H. (1957). The concept of development from a comparative and organismic point of view. In D. B Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis, MN: University of Minnesota.

CHAPTER 6

CONCLUSION

The Problem Areas

This dissertation centered on the claim that dynamic formative assessments of how learners' understanding evolves in response to problem situations are widely applicable and viable approaches of adaptive instruction. Notably, creating adaptive learning environments necessitates knowing the extent to which students understand the given problem situations and the changes in their levels of understanding. The sequence of studies and papers described in this dissertation was initiated due to the lack of a robust diagnostic methodology to assess changes in student understanding to a complex problem situation. The problem areas addressed in these papers are as follows.

First, a theoretically grounded framework forms the basis for viable formative assessment methods and technologies. A theoretically sound and systematic assessment model is required to determine levels of expertise, explain learning progress, and provide adaptive instruction that meets individual requirements in terms of differences in cognitive stages of expertise as an established problem space.

Second, concept map methods and technologies are widely used for externally representing students' knowledge structures so that students' problem-solving knowledge and skills can be assessed. It was assumed that using natural language responses as a basis for concept map representations of student thinking was likely to provide a reliable foundation for use in providing formative feedback and assessment (Pirnay-Dummer, Ifenthaler, & Spector, 2010). It was required to identify which methods that use text responses to generate a concept map work best in terms of forming the basis for dynamic formative feedback.

Third, a new concept map technology constructing knowledge representations from natural language responses needed to be explored and validated. Language plays a critical role in building and mediating an individual's internal representations of the external world (Wittgenstein, 1922). The text is an initial re-representation of a student's beliefs and thinking about a problem, and the constructed concept map is a second representation of that mental model. A technology should create concept maps closer in meaning and structure to the targeted internal mental models.

Fourth, the theoretical framework suggested to account for learning progress in a problem situation needed be empirically validated so as to use that framework as a diagnostic model for a formative assessment technology. The validation process involved identifying measurable features of concept maps, selecting statistical methods pertinent to testing hypothesized learning stages, and exploring the extent to which the model fits to the data.

Summary of Results

This dissertation composed of four manuscripts focused on devising better assessment methods when natural language is used for representing students' understanding to a complex problem situation. Specific findings are as follows.

The first paper, *Theoretically grounded guidelines for assessing learning progress: Cognitive changes in problem-solving contexts*, conceptualized the levels of learning progress, associating the development of expertise in domain learning with the structural features of mental models. The theory of mental models accounts for how people conceptualize problem situations. That is, a mentally represented problem space is a structure including diverse relationships. It was necessary that assessment tools be adapted for the complex, dynamic structure of mental models so that diagnostic, formative information became more precise. In short, the proposed stage-sequential model of learning progress was theoretically justified as being able to serve as a diagnostic model of learning progress.

The second paper, *Cross-validation study of methods and technologies to assess mental models in a complex problem solving situation*, was based on the assumption that an individual student's understanding is meaningfully elicited via a natural language approach. Two state-of-the-art technologies, ALA-Reader and T-MITOCAR, consistent with that assumption were compared to an alternative method established as a benchmark. The benchmark approach was created by drawing on semantic relations distilled from responses. The results demonstrated that the benchmark approach has the potential to be a preferred way to visually represent concept maps because of three findings: (a) concept maps elicited via the benchmark were much more descriptive than those of the other two models; (b) the benchmark model was able to more capably distinguish better concept maps from those of lesser quality; and (c) the benchmark model had no constraints on the number of words.

The third paper, *Development of an assessment technology for measuring knowledge structures using natural language responses to a complex problem scenario*, elaborated on the benchmark approach explored in the second study. Eliciting the semantic structure was assumed to be a better way to visually represent concept maps. That approach was termed *Semantic Relation* (SR) in contrast with the other approaches, Proximity Relation (adopted in T-MITOCAR) and Adjacent Relation (employed in ALA-Reader). The data investigation suggests that SR is alternative more productive way of enabling concept maps to be more descriptive and authentic way of eliciting how an individual is thinking about a complex problem-solving situation.

The fourth paper, *Investigation of a model of stage-sequential learning progress in problem solving,* tested the stage-sequential model of learning progress conceptualized in the first paper. To accomplish this, a set of parameters describing the features of concept maps were defined and then matched with the structural features of mental models that account for each level of learning progress. Due to the latent characteristics of mental states, latent class model (LCM) methods were employed to validate the proposed model of learning progress. Good-fit indices proved that the students proceeded toward an expert-like knowledge structure through the suggested levels of learning progress (e.g., $\chi 2 = 48.435$, $p > .05$).. While a large number of respondents were classified at the novice level or proficient learner level, the proportion of those at the advanced beginner and competent learner levels was not significant.

Implications and Recommendations

This dissertation, as an initial effort, dealt with some of the current assessment issues pertaining to complex learning. The findings of this study should continue to inform subsequent research. Some potential topics are as follows. First, as a basis for detecting and validating stages of learning progress, this study provides a theoretical framework called stage-sequential model of learning progress. The framework can work as a diagnostic model for a formative assessment technology and be applied to further validation studies in order to find a best-fit model that accounts for changes in learner understanding in response to ill-structured complex problem situations.

Second, methods for creating reference models were theoretically (in chapter 2) and practically (in chapters 3 to 5) discussed. In an instructional setting, single or multiple reference models can be used to diagnose an individual response in near real-time, when analyzed by an automated assessment technology, in order to help that respondent consider aspects of the problem situation that might have been overlooked but that appear in one or more reference models.

Third, the studies in this dissertation can continue to developing an automated assessment technology embedding an SR concept mapping method and diagnostic model of learning progress. That technology would enable a teacher to gain a better sense of students' learning and provide them with elaborate feedback and support. For example, McKeown (2009) used HIMATT with 40 actual classroom teachers. The teachers did manage to use that technology to diagnose student understanding and provide instructional support to an individual even though the tool was new to them.

Development of the technology likely includes technological supports in terms of natural language processing. That study would require multi-disciplinary efforts including computational linguistics, computer sciences, educational statistics, and instructional science. In addition to instructional use, the automated technologies could be applicable to a wide range of areas: traditional language comprehension study in linguistics; automated essay evaluation; expertise modeling; competency diagnosis in adult learning; and longitudinal study of learning progress in problem solving.

Fourth, this study discussed the assessment model and methods to diagnose domain knowledge as an internally represented problem situation. It is generally accepted that human cognition includes meta-cognition and motivation. Necessarily, it is required to investigate how the two cognitive domains influence or interact with the proposed stages of learning progress.

Furthermore, it could be assumed that meta-cognition and motivation proceed through qualitatively different stages just as general cognition is assumed to do.

Fifth, the relationships among cognitive and non-cognitive domains need to be identified so that learning progress can be more comprehensively modeled. Following are some possible questions. What non-cognitive factors are related to learning progress? To what extent is each non-cognitive factor associated with the change of learning stages? Can the five-stage model of learning progress be a shared model that classifies non-cognitive factors?

Sixth, instructional and feedback strategies associated with each developmental stage of learning progress should be elaborated. Assessment results should be accompanied by instructional supports suited to each individual or group of students. Instructional models based on diagnostic assessment can lead to the development of diverse instructional applications such as intelligent tutoring systems.

Limitations

This sequence of studies has the following limitations. First, theoretical suggestions that include some measurable attributes for determination of expert-level mental models require elaboration. The assessment technologies designed in this study only deal with stages of learning progress that preceding the expert level. Intuitive decision-making (the expert learner) toward problem-solving is not easily discerned by investigating a single set, or even a few sets, of mental observations, such as concept models, because the measurable features of experts' mental models are still not well understood.

Second, although the semantic relation (SR) approach illustrated in the third paper can be automated, the semantic relations including concepts were manually distilled in the study. It is true that complex structured sentences are not easily rendered in a concept map in terms of automatically distilling correctly paired concepts. In order to elicit concept models from multiple quantified language inputs, firstly, the algorithm to identify semantic relations from a text should be further elaborated and then technological supports in terms of automated natural language processing would be required.

Third, the fourth study attempted to validate the model of stage-sequential learning progress composed of four stages (the novice, advanced beginner, competent learner, and proficient learner). The latent class modeling analysis demonstrated that there are probably two stages rather than four stages in the data. Two limitations might cause that result. The relatively small sample size (*N*=143) of this study is admittedly problematic. For instance, eight items created $2^8 = 256$ possible response patterns. Following Agresti and Finlay’s (1997) suggestion, in an ideal condition ($N = 256 \text{ X } 5 = 1280$) that setting needs more than five respondents at each pattern. The small sample size probably led to many cells with no case and reduced the chance for the two items (items 8 and 9) to have positive responses. A larger sample size including all ten items in an analysis may provide somewhat more information in terms of the number of stages and their estimated proportions. In addition, the research context can be taken into consideration. The course was not designed to teach content directly related to the problem situation (i.e., technology adaptation to classroom instruction). Moreover, the responses were gathered earlier in the semester. That is, students responded to the problem with little chance of having been instructed about the problem context. It is possible to argue that at that point there were two groups of students who either had or had not prior experience. There might have been a lack of time to develop transitional knowledge structures. For further studies, a better research context could be suggested as involving (a) providing instruction related to a problem situation prior to the data gathering; and (b) implementing data collection in the middle of an instructional period.

Fourth, this dissertation assumed four stages of learning progress through which a learner’s understanding of a problem evolves. The model used in this study is one a possible suggestion. Different models could very well accounting for learning progress. For example, as found in chapter 5, the number of stages might not be four, as suggested Dreyfus and Dreyfus (1986). The number of stages and the patterns of stage changes could vary depending on different problem-solving domains. In addition, considering that the concept map methodology used for analyzing learning progress was derived from English syntax and semantics, the assessment methodology might not work in other languages and cultures; the key features of concept maps and their changes when learning and instruction take place could vary.

Fifth, it is required to study the longitudinal stage changes of mental models based on the given framework so that we can apply the model of learning progress to evaluating the effectiveness of instruction and determine the proper educational supports for an individual. Collins and Wugalter (1992) pointed out that psychological research and theory is increasingly turning to longitudinal studies in which development is monitored by following individuals over time. In an investigational setting, we can suppose that researchers find a set of patterns in stage transition, their proportion, and transition probabilities among measurement points from the statistical analyses. These results can inform researchers of students’ dynamic changes of learning progress and the effects of a given instructional intervention in a longitudinal manner.

Conclusion

The ultimate goal of this sequence of studies is to propose theoretical models and methodologies to diagnose cognitive development in problem-solving situations, which are essential to providing a best-fit learning experience to students with precision and confidence. The theory of mental models associated with the development of expertise provides practical suggestions on how to elicit learner models, build reference models for both structured and ill-structured problem-solving tasks, and explain learning progress through which learners experience qualitatively distinct cognitive stages.

The SR as a new concept map approach embedded in automated technologies could be an effective formative assessment tool for classroom learning. For example, the SR technology can enable teachers to identify individual student progression on complex problem-solving based on the whole structure of the concept map and to provide personalized feedback in terms of missing key areas and relations. Moreover, the studies of this dissertation provide future research with methods and tools that involve (a) a set of parameters quantifying traits of a knowledge structure; (b) the multidimensional attributes of a knowledge structure; (c) a set of similarity measures applicable to the studies of cognitive changes; and (d) a statistical approach (that is, the LCDM approach) to diagnose the stages of learning progress. Those methods are applicable to a wide range of areas that include the following: the studies of conceptual change in a problem-solving situation, linguistic comprehension, the evaluation of scientific argumentations, expertise modeling, and longitudinal studies of learning progress.

References

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*(1), 131-157.

McKeown, J. (2009). *Using annotated concept map assessments as predictors of performance and understanding of complex problems for teacher technology integration*. (Unpublished doctoral dissertation). Florida State University, Tallahassee, FL.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. (2010). Highly integrated model assessment technology and tools.*Educational Technology Research & Development*, *58*(1), 3-18. doi:10.1007/s11423-009-9119-8.

Wittgenstein, L. (1922). *Philosophical investigations*. London: Blackwell.

APPENDICES

APPENDIX A

STUENT QUESTIONNAIRE

Personal Background Questionnaire

We would like to know some of your background. As informed earlier, any information obtained about you as a participant in this study will be held confidential. Your identity will be protected with a pseudonym or number.

Are you

- ☐ Male
- ☐ Female

What year student are you?

- ☐ Freshman
- ☐ Sophomore
- ☐ Junior
- ☐ Senior

At your last birthday, were you

- ☐ 20 or less
- ☐ 21-24
- ☐ 25-29
- ☐ 30 or more

What previous degree(s) do you have? Please include area(s) of study.

____
____

What is your current degree and area of study?

____
____

Before you take this course, have you taken any courses related to use of technology in teaching and learning?

- ☐ Yes
- ☐ No

If yes, please list up to three courses that you believe to be the most important.

____
____

What is your typical role in the educational settings that you are involved?

- ☐ Teacher
- ☐ Student
- ☐ Instructional designer
- ☐ Other_________________________

How many years have been in this role?

- ☐ 0-2
- ☐ 3-5
- ☐ 6-8

- ☐ 9-11
- ☐ 12 or more

How often do you have opportunities to do instructional design in this role?

- ☐ 1 – Never
- ☐ 2 – Seldom
- ☐ 3 – Sometimes
- ☐ 4 – Often
- ☐ 5 – Very often

APPENDIX B

THE PROBLEM-SOLVING TASK

<u>Case Study</u>

Directions: Read the case study described below and then prepare a response to the questions below (**written response with at least 350 words is required for each question**):

Assume that you have been involved in evaluating a media implementation project in an urban inner middle school. At the beginning of the school year all of the students assigned to four subject area teachers (math, language arts, social studies and science) in the seventh grade at the middle school were given tablet PCs (laptop computers also equipped with a stylus/pen and a touchscreen that can be written upon) and were also given wireless internet access at home and in school for a entire year.

The students took the tablet PCs home every evening and brought them into classes every day. The teachers were also provided with tablet PCs 24/7 (24 hours a day, every day of the week) for the entire year. The teachers and students were trained on how to use the tablet PCs. Moreover, all of the curriculum materials (textbooks, workbooks, student study guides, teacher curriculum guides, some activities, tests, etc.) were installed on the tablet PCs or were accessible through the tablet PCs.

Your job as one of the evaluators for the project was to examine how this innovation (providing teachers and students with tablet PCs 24/7) changed the way instruction was presented in the classrooms of the four teachers. Results indicated that the innovation had very little effect on the manner in which instruction took place in the teachers' classrooms.

1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

Acknowledgment

This case you are using is based on a case described by Robert Reiser for use in his Trends and Issues in ID&T course at Florida State University

APPENDIX C

THE LETTER TO THE PANEL FOR THE DELPHI STUDY

Call for the First Response to the Delphi survey

Dear Professor

I am writing to request your assistance in a research study that I am conducting. My dissertation supervisor at UGA is Professor Michael Spector. As an expert in the area of Designing and Implementing Technology-Enhanced Learning Environments in Classrooms, we know that you are well qualified to participate. The assistance I am requesting will only require a small amount of your time.

Basically, I would like you to serve as a panel member in a Delphi process to establish an expert reference model on the case study to be used in my research. What is required is simply for you to respond to a problem scenario (see attached) and respond to two questions involving the key factors influencing the problem situation and their relationships.

In accordance with a Delphi process to create consensus among a small number of expert respondents, you will be asked to respond to the case a couple of times to get a consensus on the key factors and their relationships with other panel members. The time required to respond is less than 30 minutes, and I expect that only 2 or 3 rounds will be required to reach consensus (see the Delphi process attached below). The consensus expert response will be used as a standard against which responses of less experienced persons will be assessed. In accord with the approved IRB for this effort, your response will be kept anonymous and your identity not revealed to anyone other than me.

Anticipated Delphi Process
* Round 1: Brain Storming
- Collect and consolidate all responses from experts
* Round 2: Narrowing Down
- Send refined final version of consolidated lists including statements and used concepts
- Ask expert to add comments if he/she disagrees with or has different opinion(s) on a statement
- Ask experts to rank key statements and concepts
* Round 3: Ranking
- Send each panelist ranked statements and concepts summarized by the investigators
- Ask to revise his/her judgments or to specify the reasons for remaining outside the consensus

It would be very happy to your first response to the problem scenario by the middle of January if that is possible. The total time commitment expected for this activity is between one and two hours. Please let me know if you are able and willing to participate. If more time is required, I will happily accommodate your schedule, as your expertise is very much needed for this study.

Again, I am very glad to include you on this panel and excited to learn from you.

Sincerely,
Min Kyu Kim

APPENDIX D

A SAMPLE RESPONSE TO THE FIRST ROUND OF THE DELPHI SURVEY

1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

It is possible that the teachers and students were only trained on operating the tablet PCs, which is to say that the training they received was primarily technology training. In addition to the familiarization with the functionality of the tablet PCs, the teachers required training on how to effectively integrate student of the tablet PCs into their instruction. The students also required training on how to effectively make use of the tablet PCs in support of learning goals and assignments. One factor that probably has a significant effect on instructional practice is training that aims at pedagogical uses of the technology for teachers and learning uses of the technology for students. These two kinds of training should be coordinated to ensure proper technology, pedagogical, and content knowledge. A second factor that has an effect on instructional practice is time of training. The training should be extensive and include opportunities for tryouts with feedback. Type of training is also significant – as just indicated, the training should be experiential and not just informational. Other factors that could affect instructional practice are prior training of the teachers, the length of time they had been teaching, their attitudes towards innovation, and incentives to make effective use of the new technology. Similar factors pertain to students. The longer teachers or students have developed habits of teaching and learning that are not compatible with the affordances of the new technology, the less likely are they to embrace a new technology. Similarly, if teachers or students believe that a new technology is not likely to have an impact of outcomes of interest, then they are not likely to embrace a new technology. The lack of incentives to make effective use of a new technology could also contribute to lack of use. In short, motivational concerns could account for the lack of change in instructional practices. Still other inhibiting factors include a conservative cultural environment in terms of teaching and learning. The grade level involved might be an additional consideration – 7th grade involves students who are about 12 years old. In principle, they should be able to understand the advantages of a new technology in terms of learning and performance, but they may have expected other uses of the laptop in addition to educational uses. If games were also available, they might have served as an incentive for student use of the tablet PCs. Finally, it is impossible to put much confidence in an implementation that only involved 4 teachers.

2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

First, a professional development plan for teachers could have been put into place to ensure that teachers had proper technological, pedagogical, and content knowledge to make effective use of the tablet PCs. This plan should be implemented well in advance of introducing the tablet PCs in the classroom, and it should include an experiential training approach with substantial formative feedback. Similarly, when teachers had been properly trained, they should then engage students in a similar training regimen that includes not only technology training on the use of the tablet PCs, but also how to make use of the computers in support of learning activities and exercises. Prior to introducing the tablet PCs to teachers and students, it would be a good idea to determine attitudes toward technology innovation in general as well as attitudes toward the tablet PC. Beliefs about whether use of the tablet PC is likely to improve learning and instruction should be determined and taken into account in the training to be provided. In addition, teachers and students should be given opportunities to develop their own activities and exercises to showcase the use of the tablet PC in promoting learning and enhancing instruction. Because the tablet PCs are being deployed in 7^{h} grade math, language arts, social studies and science classes, an evaluation plan should be developed to see if differences in terms of use and impact emerge for subject area sub-groups. A pilot test should be made with a few teachers to revise training and evaluation plans, and then all of the teachers should be involved. A real issue pertains to the small number of teachers and students involved, without many more teachers involved, it will be difficult to have any confidence in findings. If it is possible, all of the teachers and students in a middle school should be involved – that would still be an exploratory study, though, but it could show how to spread the innovation to other schools, and perhaps a randomized control trial could be developed as the effort scales up. As mentioned earlier, teachers and students should be tested at the outset with regard to attitudes and beliefs. If attitudes and belief are detected that are likely to inhibit effective use, these should be addressed in the training, and the training should be experiential with lots of opportunities to test a variety of educational uses of the tablet PCs.

APPENDIX E

THE ROUND #2 OF THE DELPHI STUDY

Round #2 of the Delphi Study

Aimed at Developing a Consensus Expert Model for Thinking about the Problem Situation Below

CASE STUDY

Assume that you have been involved in evaluating a media implementation project in an urban inner middle school. At the beginning of the school year all of the students assigned to four subject area teachers (math, language arts, social studies and science) in the seventh grade at the middle school were given tablet PCs (laptop computers also equipped with a stylus/pen and a touchscreen that can be written upon) and were also given wireless internet access at home and in school for an entire year.

The students took the tablet PCs home every evening and brought them into classes every day. The teachers were also provided with tablet PCs 24/7 (24 hours a day, every day of the week) for the entire year. The teachers and students were trained on how to use the tablet PCs. Moreover, all of the curriculum materials (textbooks, workbooks, student study guides, teacher curriculum guides, some activities, tests, etc.) were installed on the tablet PCs or were accessible through the tablet PCs.

Your job as one of the evaluators for the project was to examine how this innovation (providing teachers and students with tablet PCs 24/7) changed the way instruction was

Q1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

Q2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

►A summary of the expert responses to each of the two questions is presented next. In each case, you are asked to indicate your agreement or disagreement, provide any additional opinions or comments you might have, identify key concepts by highlighting selected terms, and then rank order the consensus terms identified in round #1.

Panel Responses to Question #1

Q1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

Eleven key issue areas affecting the project's results have been identified and specific descriptions associated with each issue area have been distilled from the first round responses. Please read each issue and the associated descriptions.

1. If you disagree with or have comments on it, please mark on "D (Disagree)" or "C (Comment)" and put your opinions on it using "Review>New Comment" function of MS-WORD.

2. Please find out TWO or THREE of the most important concepts (terms) associated with each key issue. As an example, three key concepts associated with Resisting Change are indicated below. You can highlight others. Please use the "Text Highlight" function of MS-WORD.

	Disagree/ **C**omment
C1. Resisting Change	D/C
C1-1. There are some teachers who will simply resist almost any innovation (about 3%). If they are not managed, they can slow the rate of change during any innovative initiative for the other teachers in the building.	D/C
C1-2. Teachers were uncomfortable with their new roles.	D/C
C1-3. Teachers were simply teaching as they were taught.	D/C
C1-4. The longer teachers or students have developed habits of teaching and learning that are not compatible with the affordances of the new technology, the less likely are they to embrace a new technology.	D/C
C2. Lack of Supportive Environment	D/C
C2-1. Inhibiting factors include a conservative cultural environment in terms of teaching and learning.	D/C
C2-2. There are concerns that the environment does not support change.	D/C
C2-3. Even teachers who have adequate skills and knowledge still may fail to enact new instructional practices because of the environment.	D/C
C2-4. Environment could include a culture that does not support the desired performance.	D/C
C2-5. Visible, sustained commitment among all constituencies in a desired environment makes the change.	D/C
C2-6. An ongoing supportive environment where teachers initially learn how to use the technology, how to use the technology with their content (including effective instructional strategies) and how to continue to develop their expertise in the technology and incorporating it to the classroom is critical.	D/C
C2-7. For example, the lack of incentives to make effective use of a new technology could also contribute to lack of use.	D/C

C3. Insufficient Training	D/C
C3-1. Teachers were not trained how to integrate the new media including hardware and software into the curriculum.	D/C
C3-2. The teachers required training on how to effectively integrate student use of the tablet PCs into their instruction.	D/C
C3-3. The students also required training on how to effectively make use of the tablet PCs in support of learning goals and assignments.	D/C
C3-4. One factor that probably has a significant effect on instructional practice is training that aims at pedagogical uses of the technology for teachers and learning uses of the technology for students.	D/C
C3-5. A second factor that has an effect on instructional practice is time of training. The training should be extensive and include opportunities for tryouts with feedback.	D/C

C4. Design Issue	D/C
C4-1.There was likely insufficient study of how instructional and learning practices in the classroom were being conducted already *without* the technology.	D/C
C4-2. One big issue is defining what a successful integration or change in instructional practice actually is.	D/C
C4-3. According to Ely (1990), successful technology implementations must begin with an identified instructional need.	D/C
C4-4. While everyone in the situation may have felt that they knew this already, the assumptions inherent in a design situation need to be articulated and checked if they are not to distort the design space.	D/C
C4-5. A primary generator (Lawson, 1990) or central narrative of the change are not evident in the case as presented, nor is it clear that there were efforts to establish such a narrative consciously and follow and adjust it through the course of the project.	D/C

C5. Belief and Attitude	D/C
C5-1. Teacher beliefs also play a big role in adopting new practices and changing their instructional practice. They may not believe that students learn with laptops, and thus do not use them in their instruction. They may believe that they do not have enough time to integrate the laptops.	D/C
C5-2. Similarly, if teachers or students believe that a new technology is not likely to have an impact on outcomes of interest, then they are not likely to embrace a new technology.	D/C
C5-3. Stakeholders need to view technology as the solution (or partial solution) to that need and believe it is possible to successfully integrate technology to fix the problem.	D/C

C6. Professional Development	D/C

C6-1. Professional development (PD) for teachers is a major determinant as to whether or not teachers will employ technology in classrooms.	D/C
C6-2. Teachers didn’t have enough professional development using the technology in classroom teaching and learning, on ways to integrate use into their teaching, and best practices with regard to effective educational use.	D/C
C6-3. Teacher professional development that discusses not just technical know-how but also pedagogy could help teachers realize how to do things differently that takes full advantage of the affordances of the tablets.	D/C
C6-4. Professional development is needed for teachers to become skilled or knowledgeable about a topic that will in turn increase the knowledge and achievement of their students.	D/C
C7. High-Stakes Tests	D/C
C7-1. A big barrier today is state-mandated testing.	D/C
C7-2. Many teachers see technology as taking away from preparing for those tests.	D/C
C7-3. Teachers were more interested in ensuring that the students do pass high-stakes year-end tests and, thus, reverted back to the traditional method of instruction.	D/C
C8. Mentoring	D/C
C8-1. The only support teachers had during implementation was technical support; they lacked a mentor who could assist them as instructional issues arose throughout the year.	D/C
C8-2. Mentoring on additional and advanced uses of the technology in the classroom is critical for teachers to increase their skills and maintain their motivation in utilizing the technology.	D/C
C9. Empowerment & Engagement	D/C
C9-1. Teachers were not involved in the decision to implement the new media; thus, they did not fully “buy into” the plan.	D/C
C9-2. The intervention seems to have been “applied to” this community rather than involving them from the beginning as collaborators in its design and modification.	D/C
C9-3. The most salient component of the intervention from the standpoint of the participants was therefore likely to be the literal technology.	D/C
C10. Community of Practice	D/C
C10-1. There was no attempt to get the four teachers to meet on a regular basis throughout the year to discuss instructional issues, nor was there any attempt to link these teachers with teachers at other schools who were also attempting to integrate these types of tools (tablet PCs or similar devices) into their instructional practices.	D/C
C10-2. In other words, there was no attempt to establish a community of practice.	D/C
C11. Students	D/C

C11-1. The grade level involved might be an additional consideration – 7th grade involves students who are about 12 years old.	D/C
C11-2. In principle, they should be able to understand the advantages of a new technology in terms of learning and performance, but they may have expected other uses of the laptop such as games in addition to educational uses.	D/C

Please RANK the key issues in order of their contribution to the poor project results from 1 (most influential) to 11 (least influential).

Key Issues	**Rank Order**
C1. Resisting Change	
C2. Lack of Supportive Environment	
C3. Insufficient Training	
C4. Design Issue	
C5. Belief and Attitude	
C6. Professional Development	
C7. High-Stakes Tests	
C8. Mentoring	
C9. Empowerment & Engagement	
C10. Community of Practice	
C11. Students	

PLEASE SEE THE NEXT PAGE

Panel Responses to Question #2

Q2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

Twelve key strategies that could have been employed to mitigate the project have been identified and specific descriptions associated with each strategy have been distilled from the first round responses. Please read each strategy and the associated descriptions.

1. If you disagree with or have comments on it, please mark on “D (Disagree)” or “C (Comment)” and put your opinions on it using “Review>New Comment” function of MS-WORD.

2. Please find out TWO or THREE of the most important concepts (terms) associated with each key strategy. Use the “Text Highlight” function of MS-WORD to indicate those key concepts as was done in the previous section.

	Disagree/ **C**omment
S1. Empowerment & Engagement	D/C
S1-1. A more appropriate approach to this intervention would have been the selection of some form of participatory design as the basis for process decisions.	D/C
S1-2. The administration should have gotten at least some of the teachers (preferably the opinion-leaders) involved in the initial planning of this project so that the teachers had a voice in the decisions that were made.	D/C
S1-3. All efforts that make up a school-based technology initiative should focus on empowering teachers to create innovative practices (i.e., activities and exercise) that can help all learners succeed.	D/C
S1-4. If you want teachers to totally transform their curriculum, give them some extra time to prepare their lessons, and compensate for additional time outside of school they need to go to professional development.	D/C
S2. Professional Development (PD)	D/C
S2-1. Borrowing from a systems-based model for technology integration, one big component of successful initiatives like this is professional development.	D/C
S2-2. The PD workshop should also have focused on the changing instructional role that the teachers were likely to play (i.e., serving as guides on the side, not sages on the stage).	D/C
S2-3. A professional development plan for teachers could have been put into place to ensure that teachers had proper technological, pedagogical, and content knowledge to make effective use of the tablet PCs.	D/C
S2-4. The PD plan should be implemented well in advance of introducing the tablet PCs in the classroom, and it should include an experiential training approach with substantial formative feedback.	D/C
S2-5. The goal of modeling how to use the technology for instruction in the PD program would be to have teachers learn effective strategies as form of best practices on integrating the technologies into the classroom.	D/C
S2-6. The teachers in turn would model the use of the technologies to the students, who would then learn	D/C

effective uses of the tablets PCs.

S3. Training	D/C
S3-1. It is stated that all were trained on how to use the tablet PCs.	D/C
S3-2. After initial training on how to use the tablet PCs, teachers need to engage in extensive technology and curriculum integration training in order to effectively use the innovation in their classrooms.	D/C
S3-3. This training could have been done in a variety of different ways. For example, the teachers could have been provided with an interactive hands-on summer workshop on how to integrate the new media (hardware and software) into the curriculum.	D/C
S3-4. They should then engage students in a similar training regimen that includes not only technology training on the use of the tablet PCs, but also how to make use of the computers in support of learning activities and exercises.	D/C
S4. Need Assessment	D/C
S4-1. It is not illustrated in the case whether a needs assessment was ever done to determine the purpose of providing the tablet PCs to the teachers and students.	D/C
S4-2. Successful classroom innovations require that the desired change be clearly understood among stakeholders in terms of its need, purposes and processes, complexity, and quality/practicality.	D/C
S4-3. The observation process of current practices would also provide opportunities for the teachers and students to input directly to the process, increasing their connection to the change before it begins.	D/C
S4-4. Prior to introducing the tablet PCs to teachers and students, it would be a good idea to determine attitudes toward technology innovation and the tablet PC as well as beliefs about whether use of the tablet PC is likely to improve learning and instruction.	D/C
S5. Community of Practice	D/C
S5-1. The school might have established an online community of practice involving the four teachers at the school as well as teachers at other schools who were also attempting to integrate these types of tools (tablet PCs or similar devices) into their instructional practices.	D/C
S5-2. It would be imperative to set up an online support program for teachers on using the technologies.	D/C
S5-3. This could include daily/weekly updates on effective strategies on integrating the technology in the classroom, how to overcome integration obstacles, how to use the tablets for instruction and assessment of individual students and the entire classroom group, etc.	D/C
S5-4. In addition, the schools might have set up weekly meetings for the teachers to discuss instructional issues with each other and with a mentor.	D/C
S6. Mentoring	D/C
S6-1. The school might have provided the teachers with an on-site mentor who could have helped them solve instructional issues that came up during the year.	D/C
S6-2. The mentor focuses on basic skills and individualizes the instruction so that every teacher's basic skills increase in some way.	D/C
S7. Curricula Alignment	D/C

S7-1. One thing is to make sure the curricular materials that are loaded on the tablets are not like traditional workbooks and textbooks. Rather, have them be things like cognitive flexibility hypertexts, problem solving scaffolds, and so forth.	D/C
S7-2. Teachers should be given time enough to redesign their instruction given the new technologies.	D/C
S7-3. Teachers use technology to house a digital library of activities for teachers to share by grade level with the mentor helping them create materials.	D/C
S8. Leadership Support	D/C
S8-1. It is imperative to make school leader support visible and consistent.	D/C
S8-2. The stakeholders feel they are to supporting and sustaining meaningful educational change.	D/C
S8-3. Rewards/incentives given to teachers can help them expend the additional effort to fully integrate the new technology into their classrooms.	D/C
S9. Evaluation & Pilot Test	D/C
S9-1. Test new ideas first, then improve and replicate them based on the test.	D/C
S9-2. Because the tablet PCs are being deployed in 7h grade math, language arts, social studies and science classes, an evaluation plan should be developed to see if differences in terms of use and impact emerge for subject area sub-groups.	D/C
S9-3. A pilot test should be made with a few teachers to revise training and evaluation plans, and then all of the teachers should be involved.	D/C
S10. Support for Parent	D/C
S10-1. Offer Support for parents.	D/C
S11. Tech Specialists	D/C
S11-1. Provide one or more tech specialists to deal with technical issues (someone other than the mentor).	D/C
S12. Project Period	D/C
S12-1. Extend project period- one year is too short to see meaningful change.	D/C

Please RANK the key strategies in order of your preference from 1 (most liked) to 11 (least liked).

Key Strategies	Rank Order
S1. Empowerment & Engagement	
S2. Professional Development (PD)	
S3. Training	
S4. Need Assessment	
S5. Community of Practice	
S6. Mentoring	
S7. Curricula Alignment	
S8. Leadership Support	
S9. Evaluation & Pilot Test	
S10. Support for Parent	
S11. Tech Specialists	
S12. Project Period	

APPENDIX F

THE ROUND #3 OF THE DELPHI STUDY

Summary of the Delphi Study (Round #2)

Aimed at Developing a Consensus Expert Model for Thinking about the Problem Situation Below

An expert response to each of the two questions was written based on the panel responses gathered through round #1 and #2 of this Delphi survey. For each response, key issues and terms (concepts) are presented next. You are asked to add, omit, or change them if needed. This should require no more than 15 minutes. Just key word additions or deletions are required now. Please send me your response by May 2, 2011. Thanks again for your support.

CASE STUDY

Assume that you have been involved in evaluating a media implementation project in an urban inner middle school. At the beginning of the school year all of the students assigned to four subject area teachers (math, language arts, social studies and science) in the seventh grade at the middle school were given tablet PCs (laptop computers also equipped with a stylus/pen and a touchscreen that can be written upon) and were also given wireless internet access at home and in school for an entire year.

The students took the tablet PCs home every evening and brought them into classes every day. The teachers were also provided with tablet PCs 24/7 (24 hours a day, every day of the week) for the entire year. The teachers and students were trained on how to use the tablet PCs. Moreover, all of the curriculum materials (textbooks, workbooks, student study guides, teacher curriculum guides, some activities, tests, etc.) were installed on the tablet PCs or were accessible through the tablet PCs.

Your job as one of the evaluators for the project was to examine how this innovation (providing teachers and students with tablet PCs 24/7) changed the way instruction was presented in the classrooms of the four teachers. Results indicated that the innovation had very little effect on the manner in which instruction took place in the teachers' classrooms.

Q1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

Q2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

Question #1

Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.

Key Issue	**Key Terms (Concepts)**
Professional Development	Professional Development, Best Practice, Training, Mentor, Mentoring, Pedagogy, Affordance
Design Issue	Instructional Need, Instructional Practice, Integration, Change, Assumption, Design Space, Intervention
Lack of Supportive Environment	Environment, Support, Culture, Performance
Empowerment	Empowerment, Leadership, Inceptive, Decision, Motivation, Collaborator
Belief and Attitude	Belief, Attitude

Is this a fair summary? What would you add, omit, or change?

Question #2

Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

Key Strategies	Key Terms (Concepts)
Professional Development	Professional Development, Training, Formative Feedback, Strategy, Best Practice, Goal, TPACK, Mentor, Community of Practice
Needs Assessment	Innovation, Need Assessment, Change, Purpose, Process, Current Practice
Leadership Support/ Empowerment	School Leadership, leader, educational Change, Initial Planning, Voice, Reward, Incentive
Curriculum Alignment	Curricular Material, Time, Design
Evaluation & Pilot Test	Pilot Test, Evaluation

Is this a fair summary? What would you add, omit, or change?

A Sample Expert Response to the Question 1

Technology implementations usually begin with an identified instructional need. There was likely insufficient study of how instructional practices in the classroom were being conducted already without the technology. One big issue is defining what a successful integration or change in instructional practice actually is. While everyone in the situation may have felt that they knew this already, the assumptions inherent in a design situation need to be articulated and checked if they are not to distort the design space.

Teachers didn't have enough professional development using the technology in classroom teaching and learning, on ways to integrate use into their teaching, and best practices with regard to effective educational use. Teacher professional development that discusses not just technical know-how but also pedagogy could help teachers realize how to do things differently that takes full advantage of the affordances of the tablets. Training as a professional development effort should be extensive including teacher belief and attitude. Teacher beliefs play a role in adopting new practices and changing their instructional practice. They may not believe that students learn with laptops, and thus do not use them in their instruction.

The only support teachers had during implementation was technical support; they lacked a mentor who could assist them as instructional issues arose throughout the year. Mentoring on additional and advanced uses of the technology in the classroom is critical for teachers to increase their skills and maintain their motivation in utilizing the technology. In addition, mentors could help teachers to maintain the belief that these efforts will have positive results.

There are concerns that the environment does not support change. An ongoing supportive environment where teachers initially learn how to use the technology, how to use the technology with their content, and how to continue to develop their expertise in the technology and incorporating it to the classroom is critical. Environments could involve a culture that does not support the performance. For example, the lack of incentives to make effective use of a new technology could also contribute to lack of use. The intervention seems to have been applied to this community rather than involving them from the beginning as collaborators in its design and modification. Teachers were not involved in the decision to implement the new media; thus, they did not fully "buy into" the plan.

A Sample Expert Response to the Question 2

Successful classroom innovations require a needs assessment prior to introducing the tablet PCs. A needs assessment makes the desired change be clearly understood among stakeholders in terms of its need, purposes and processes, and practicality. The observation process of current practices would provide opportunities for the teachers and students to input directly to the process, increasing their connection to the change before it begins.

The professional development (PD) plan should be implemented well in advance of introducing the tablet PCs in the classroom, and PD plan should include an experiential training approach with substantial formative feedback. The PD program would require teachers to learn effective strategies as forms of best practice on integrating the technologies into the classroom as they incorporate those technologies and appreciate their affordances. A professional development plan for teachers could be put into place to ensure that teachers establish their own goals for change and, with respect to those goals, have proper technological, pedagogical, and content knowledge (TPACK) to make effective use of the tablet PCs.

It is imperative to have school leadership support visible and consistent. School leaders should support and sustain meaningful educational change. Leaders need to show how they value as teachers. The administration should get at least some of the teachers (preferably the opinion-leaders) involved in the initial planning of this project so that the teachers have a voice in the decisions that are made. Rewards/incentives given to teachers can help them expend the additional effort to fully integrate the new technology into their classrooms.

One thing to make sure is that the curricular materials that are loaded on the tablets are not like traditional workbooks and textbooks. Teachers should be given time enough to redesign their instruction given the new technologies and be compensated for additional time outside of school they need to go to professional development.

A pilot test should be made with a few teachers to revise training and evaluation plans, and then all of the teachers should be involved. Because the tablet PCs are being deployed in 7h grade math, language arts, social studies and science classes, an evaluation plan should be developed to see if differences in terms of use and impact emerge for subject area sub-groups. Finally providing one or more tech specialists to deal with technical issues will help teachers to test, prepare, and implement a new technology.

APPENDIX G

A STUDENT RESPONSE TO THE PROBLEM-SOLVING TASK

An Example of Student Response to the Problem-Solving Task

I think that one of the big reasons that the introduction of tablet PCs had very little effect on the instructional practices is that the teacher did not seem to use the tools in any way different than she was teaching before.
You can easily transfer paper tests, worksheets, and text onto a technology piece, it is the way you use them and format them that makes a difference.
While the teacher probably was successful in meeting standards, the technology could lend itself better if it was used in a collaborative way.
There are so many tools for this online, and so many ways to integrate them into lessons.
It may not even have been necessary for each student to have had a laptop to make a difference.
Another factor that could have limited the positive effect of the technology was the time that was spent simply teaching the children how to use the tool.
Children are fast learners and most of the time are better at learning as they do something.
Perhaps it would be better to teach them to use the tool as they perform activities on them.
It also may have helped if the four teachers collaborated on activities for the students, and incorporated activities that had to do with all subject areas together.
Performing the same activities in every class could get boring, and will lower the motivation of the children.
I believe that it is most important for the children to use these tools to learn in a way that they had never done before.
These tools should be incorporated into all areas of their lives.
Perhaps the tools are only being used at school for learning, or the homework they are sent home with is the same format that they had before they were given the tool.
It is simpler for the teacher to continue in the same teaching patterns they were using before the tools were implemented, but these tools should completely change the way the children are taught.
So in conclusion, a lack of innovative ideas, low motivation to search for new methods of teaching, and repetitive and archaic ways of using the tool all combine to produce the results seen.

APPENDIX H

LIST OF CONCEPTS (NOUNS)[7]

[7] The list of concepts (nouns) was created involving all concepts gathered from the participants' responses. Here is presented an example captured from the original spread sheet.

LIST OF CONCEPTS

	A	B	C	D	E	F	G
1		Primary Noun	Synonym 1	Synonym 2	Synonym 3	Synonym 4	Synonym 5
74	73	champion					
75	74	chance					
76	75	change	educational_change				
77	76	changing_environment					
78	77	changing_time					
79	78	character					
80	79	child	children	youth	kid		
81	80	choice					
82	81	circle					
83	82	circumstance	social_environment				
84	83	class_organization					
85	84	class_time					
86	85	class_website					
87	86	classroom	class	classroom_setting	room		
88	87	classroom_activity					
89	88	classroom_management_po	classroom_management				
90	89	cog					
91	90	collaboration	collaborative_way	collaborative_work	coordination		
92	91	collaboration_tool	interactive_tool				
93	92	collaborator					
94	93	collection					
95	94	commitment					
96	95	common_sense					
97	96	communication					

APPENDIX I

DISTILLED SEMANTIC RELATIONS TOGETHER WITH CONCEPTS[8]

[8] A list of semantic relations together with concepts was distilled from each participant's response. Here is presented an example captured from the original spread sheet.

DISTILLED SEMANTIC RELATIONS TOGETHER WITH CONCEPTS FROM A STUDENT’S RESPONSE

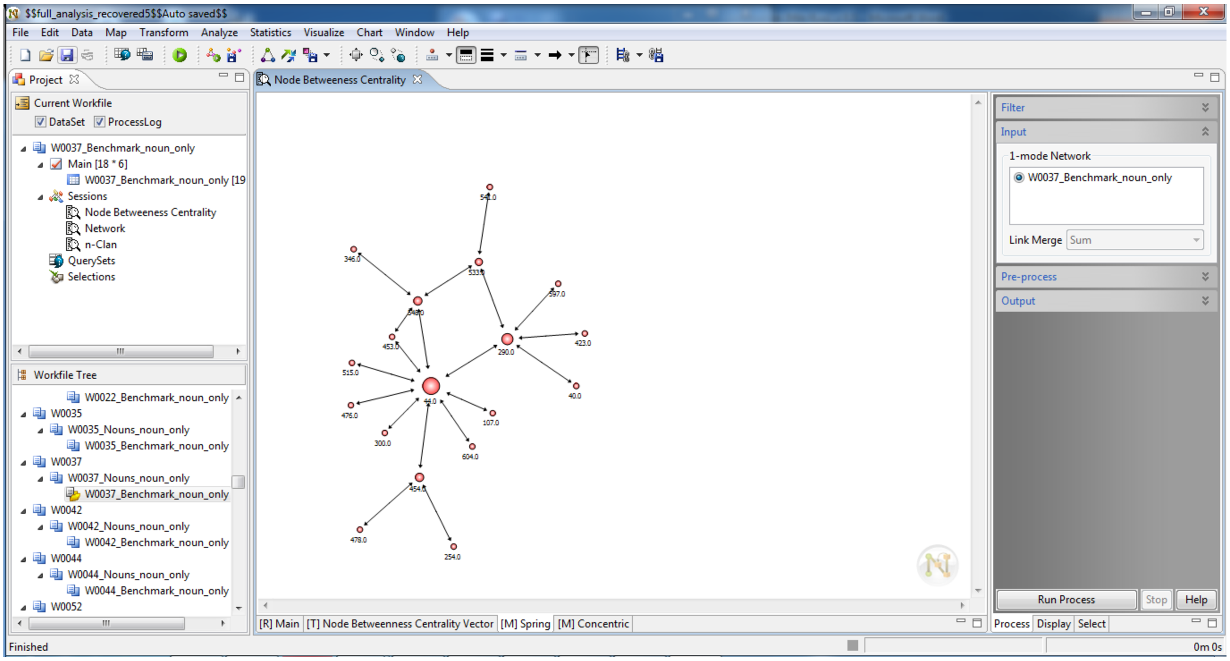
	A	B	C	D	E
1	**Source_Name**	**Target_Name**	**Linking Words**	**Source**	**Target**
2	story	teaching_method	for why-	507	542
3	teacher	technology	be not used to use	533	548
4	children	classroom	in	79	86
5	children	computer	use	79	107
6	children	cellphone	use	79	69
7	children	tablet	use	79	531
8	children	video_game	use	79	230
9	children	technology	use	79	548
10	technology	classroom	in	548	86
11	technological_difference	student	between	546	44
12	technological_difference	teacher	and	546	533
13	reason	technology	for	446	548
14	technology	teaching_method	not be implemented i	548	542
15	whiteboard	math_problem	write down	71	339
16	whiteboard	way	is	71	604
17	novel	circle	in	371	81
18	circle	way	is	81	604
19	way	language_art	to learn	604	299
20	science_lab	experience	with	470	199
21	science_lab	children	teach	470	79
22	technology	resource	be	548	453
23	resource	student	for	453	44
24	basic	technology	have to be for	49	548
25	teacher	technology	decide not to impleme	533	548

APPENDIX J

CONCEP MAP ANALYSIS TOOL: NETMINER[9]

[9] A sample screen was attached here.

CONCEP MAP ANALYSIS TOOL: NETMINER

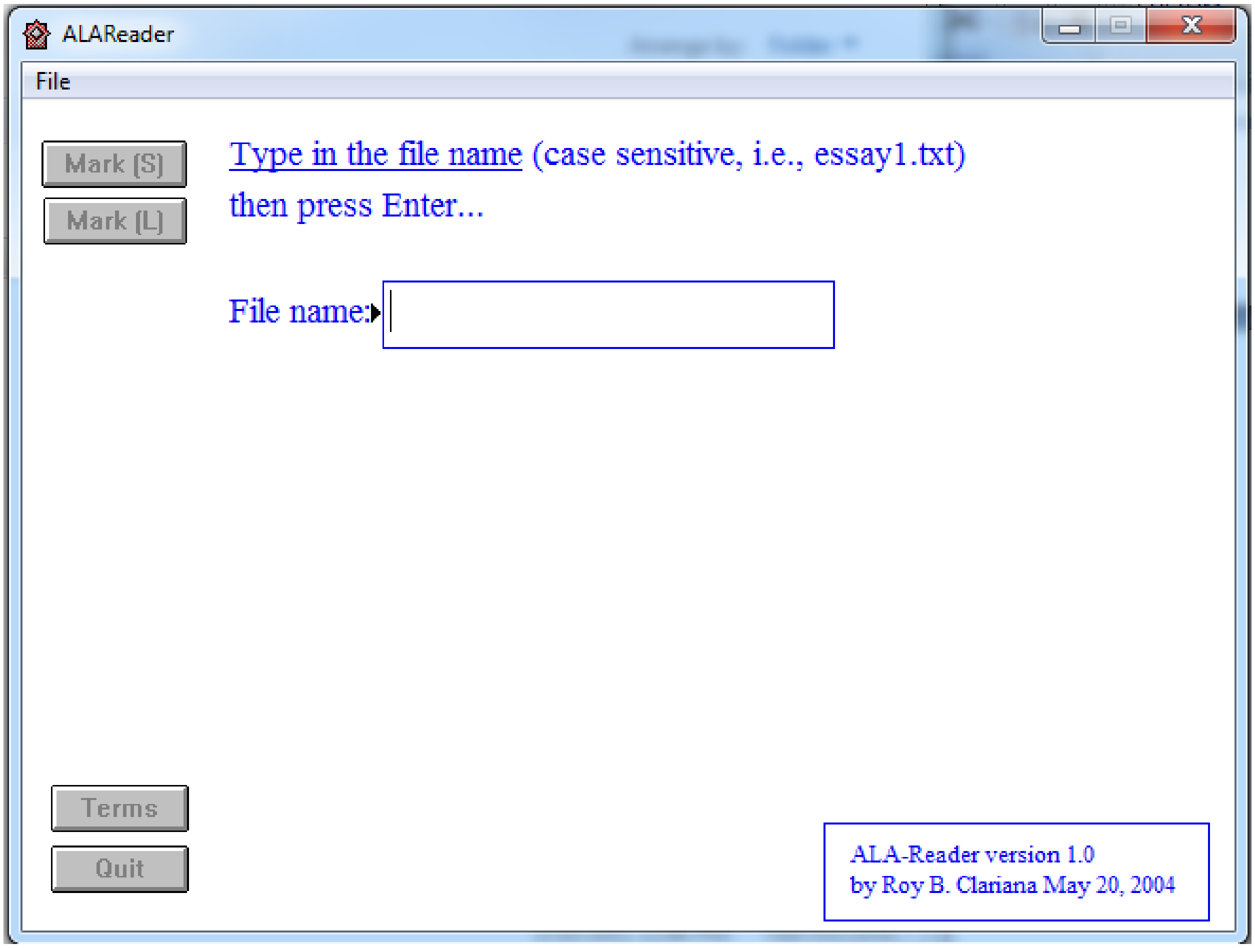


APPENDIX K

CONCEP MAP ANALYSIS TOOL: ALA-Reader[10]

[10] A sample screen was attached here.

CONCEP MAP ANALYSIS TOOL: ALA-Reader

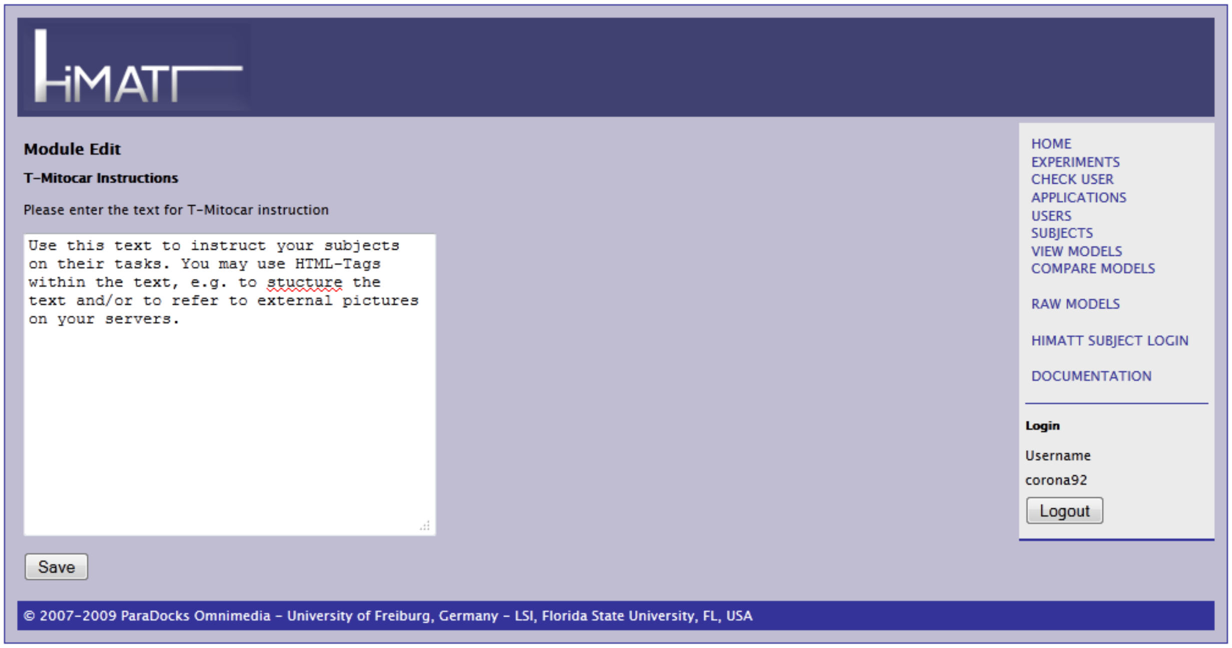


APPENDIX L

CONCEP MAP ANALYSIS TOOL: T-MITOCAR[11]

[11] A sample screen was captured from the website.

CONCEP MAP ANALYSIS TOOL: T-MITOCAR



APPENDIX M

3S SIMILARITY ANALYZER[12]

[12] A batch file to run the program was presented together with a student's output and log files.

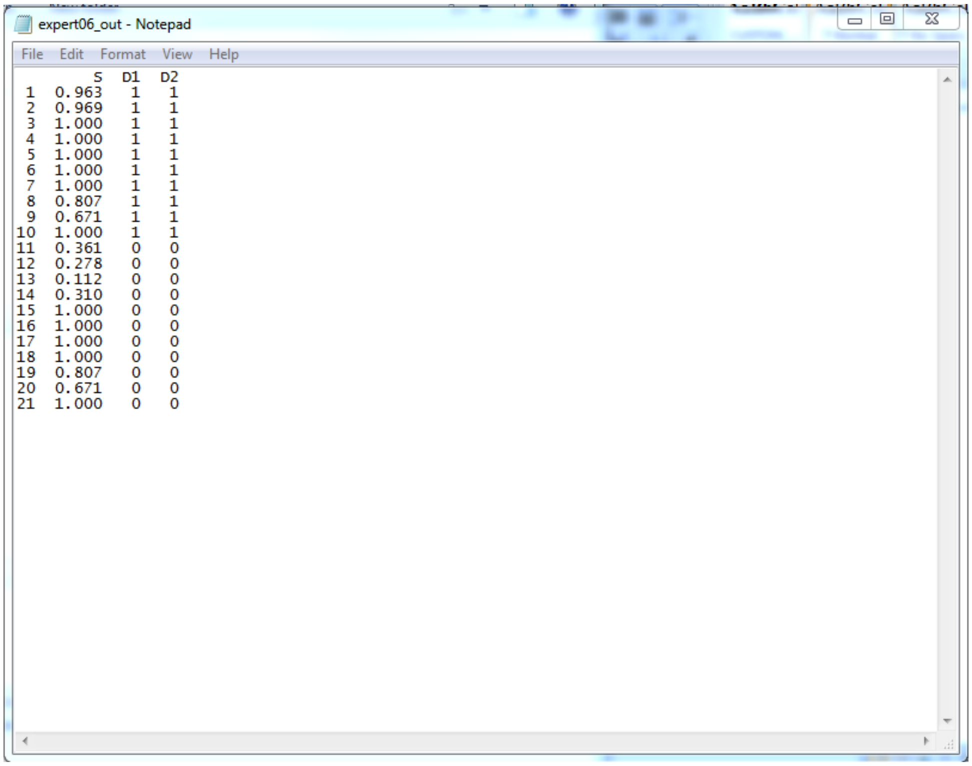
A BATCH FILE

run all_b7 c5 - Notepad

File Edit Format View Help

```
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s reference.txt >> reference_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert01.txt >> expert01_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert02.txt >> expert02_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert03.txt >> expert03_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert04.txt >> expert04_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert05.txt >> expert05_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert06.txt >> expert06_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s expert07.txt >> expert07_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0001.txt >> w0001_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0002.txt >> w0002_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0003.txt >> w0003_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0004.txt >> w0004_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0005.txt >> w0005_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0006.txt >> w0006_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0007.txt >> w0007_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0008.txt >> w0008_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0009.txt >> w0009_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0010.txt >> w0010_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0011.txt >> w0011_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0012.txt >> w0012_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0013.txt >> w0013_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0014.txt >> w0014_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0015.txt >> w0015_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0016.txt >> w0016_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0017.txt >> w0017_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0018.txt >> w0018_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0019.txt >> w0019_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0020.txt >> w0020_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0021.txt >> w0021_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0022.txt >> w0022_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0023.txt >> w0023_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0024.txt >> w0024_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0025.txt >> w0025_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0026.txt >> w0026_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0027.txt >> w0027_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0028.txt >> w0028_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0029.txt >> w0029_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0030.txt >> w0030_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0031.txt >> w0031_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0032.txt >> w0032_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0033.txt >> w0033_log.txt
SDSAnalyzer.exe -a 0.3 -b 0.7 -c 0.5 -r reference.txt -s w0034.txt >> w0034_log.txt
```

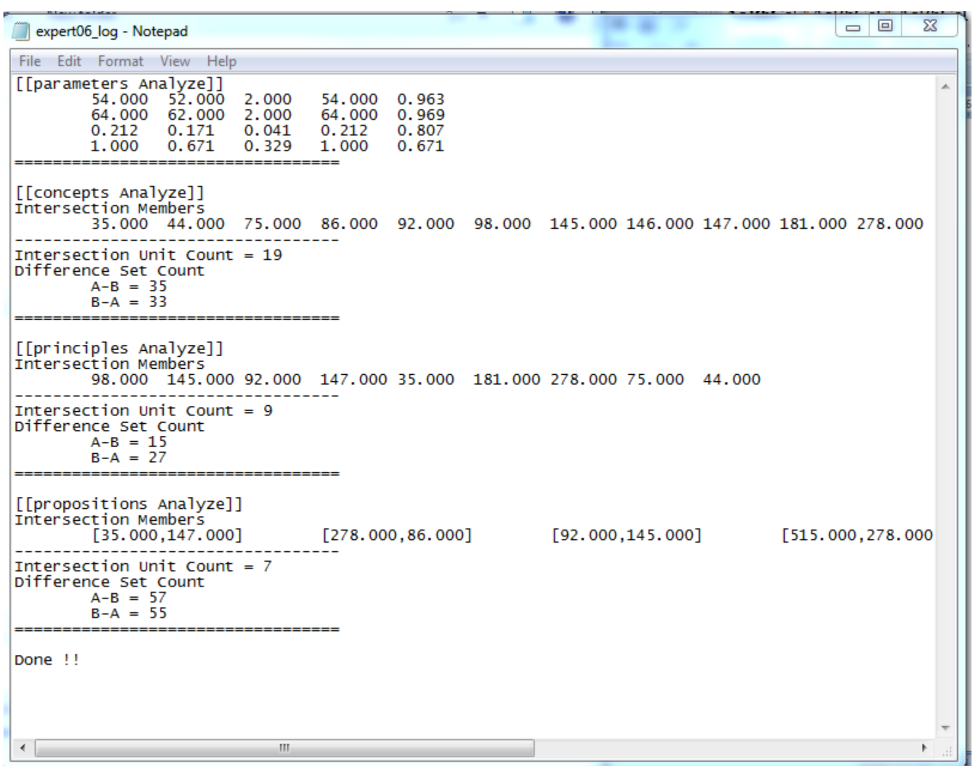
A STUDENT OUTPUT FILE


expert06_out - Notepad

File Edit Format View Help

```
         S   D1  D2
 1  0.963   1   1
 2  0.969   1   1
 3  1.000   1   1
 4  1.000   1   1
 5  1.000   1   1
 6  1.000   1   1
 7  1.000   1   1
 8  0.807   1   1
 9  0.671   1   1
10  1.000   1   1
11  0.361   0   0
12  0.278   0   0
13  0.112   0   0
14  0.310   0   0
15  1.000   0   0
16  1.000   0   0
17  1.000   0   0
18  1.000   0   0
19  0.807   0   0
20  0.671   0   0
21  1.000   0   0
```

A STUDENT LOG FILE


expert06_log - Notepad

File Edit Format View Help

```
[[parameters Analyze]]
        54.000  52.000  2.000   54.000  0.963
        64.000  62.000  2.000   64.000  0.969
        0.212   0.171   0.041   0.212   0.807
        1.000   0.671   0.329   1.000   0.671
==================================

[[concepts Analyze]]
Intersection Members
        35.000  44.000  75.000  86.000  92.000  98.000  145.000 146.000 147.000 181.000 278.000
----------------------------------
Intersection Unit Count = 19
Difference Set Count
        A-B = 35
        B-A = 33
==================================

[[principles Analyze]]
Intersection Members
        98.000  145.000 92.000  147.000 35.000  181.000 278.000 75.000  44.000
----------------------------------
Intersection Unit Count = 9
Difference Set Count
        A-B = 15
        B-A = 27
==================================

[[propositions Analyze]]
Intersection Members
        [35.000,147.000]        [278.000,86.000]        [92.000,145.000]        [515.000,278.000
----------------------------------
Intersection Unit Count = 7
Difference Set Count
        A-B = 57
        B-A = 55
==================================

Done !!
```

APPENDIX N

M-PLUS SYNTAX USED FOR THE CFA ANALYSIS

```
Mplus VERSION 6
MUTHEN & MUTHEN
08/26/2011   2:28 PM

INPUT INSTRUCTIONS

  TITLE:            First analysis of the initial model with non-normal Data
in MLM

  DATA:             FILE are cfadata.dat;

  VARIABLE:         NAMES ARE V1-V21;
                    USEVARIABLES ARE V2-V6 V10 V11-V14;

  ANALYSIS:         type=general;
                    estimator=mlm;

  MODEL:            SUR   BY V2* V3 V4 V10;
                    STR   BY V3* V4 V5 V6 V10;
                    SEM   BY V11* V12-V14;


                    SUR@1;
                    STR@1;
                    SEM@1;

                    V3 with V2;
                    V4 with V3;
                    V10 with V2;

  OUTPUT:  standardized sampstat modindices residual tech1 tech4;
```

APPENDIX O

M-PLUS SYNTAX USED FOR THE LCDM ANALYSIS

```
TITLE: !section that appears in header of output file
Parameterization of Stages of Learning Progress
Test Model 3 _ eliminating 3-way interaction
DINA applied to 2,3,6 item, three attribute data set.

DATA:!location of free format data file (in syntax file folder);
    FILE IS cut5.dat;

VARIABLE:
    NAMES = x1-x10; !list of variables in input file
    USEVARIABLE = x1-x7 x10; !use only variables x1 through x7
    CATEGORICAL = x1-x7 x10; !variables x1 through x7 are categorical (binary)
    CLASSES = c(8); !8 possible attribute patterns for 3 attribute model;

ANALYSIS:
    TYPE=MIXTURE;  !estimates latent classes;
    STARTS=0; !turn off multiple random start feature (disabled anyway);

MODEL:

%OVERALL%
[C#1] (m1); !latent variable mean for attribute pattern [0,0,0];
[C#2] (m2); !latent variable mean for attribute pattern [0,0,1];
[C#3] (m3); !latent variable mean for attribute pattern [0,1,0];
[C#4] (m4); !latent variable mean for attribute pattern [0,1,1];
[C#5] (m5); !latent variable mean for attribute pattern [1,0,0];
[C#6] (m6); !latent variable mean for attribute pattern [1,0,1];
[C#7] (m7); !latent variable mean for attribute pattern [1,1,0];

%c#1% !for attribute pattern [0,0,0];
[x1$1] (t1_1); !threshold for item 1 LCDM kernel 1
[x2$1] (t2_1); !threshold for item 2 LCDM kernel 1
[x3$1] (t3_1); !threshold for item 3 LCDM kernel 1
[x4$1] (t4_1); !threshold for item 4 LCDM kernel 1
[x5$1] (t5_1); !threshold for item 5 LCDM kernel 1
[x6$1] (t6_1); !threshold for item 6 LCDM kernel 1
[x7$1] (t7_1); !threshold for item 7 LCDM kernel 1
![x8$1] (t8_1); !threshold for item 8 LCDM kernel 1
![x9$1] (t9_1); !threshold for item 9 LCDM kernel 1
[x10$1] (t10_1); !threshold for item 10 LCDM kernel 1

%c#2% !for attribute pattern [0,0,1];
[x1$1] (t1_1); !threshold for item 1 LCDM kernel 1
[x2$1] (t2_1); !threshold for item 2 LCDM kernel 1
[x3$1] (t3_1); !threshold for item 3 LCDM kernel 1
[x4$1] (t4_1); !threshold for item 4 LCDM kernel 1
[x5$1] (t5_1); !threshold for item 5 LCDM kernel 1
[x6$1] (t6_1); !threshold for item 6 LCDM kernel 1
[x7$1] (t7_2); !threshold for item 7 LCDM kernel 2
![x8$1] (t8_2); !threshold for item 8 LCDM kernel 2
![x9$1] (t9_2); !threshold for item 9 LCDM kernel 2
[x10$1] (t10_2); !threshold for item 10 LCDM kernel 2

%c#3% !for attribute pattern [0,1,0];
[x1$1] (t1_1); !threshold for item 1 LCDM kernel 1
[x2$1] (t2_2); !threshold for item 2 LCDM kernel 2
[x3$1] (t3_2); !threshold for item 3 LCDM kernel 2
```

```
[x4$1] (t4_2); !threshold for item 4 LCDM kernel 2
[x5$1] (t5_2); !threshold for item 5 LCDM kernel 2
[x6$1] (t6_2); !threshold for item 6 LCDM kernel 2
[x7$1] (t7_1); !threshold for item 7 LCDM kernel 1
![x8$1] (t8_1); !threshold for item 8 LCDM kernel 1
![x9$1] (t9_1); !threshold for item 9 LCDM kernel 1
[x10$1] (t10_1); !threshold for item 10 LCDM kernel 1

%c#4% !for attribute pattern [0,1,1];
[x1$1] (t1_1); !threshold for item 1 LCDM kernel 1
[x2$1] (t2_2); !threshold for item 2 LCDM kernel 2
[x3$1] (t3_2); !threshold for item 3 LCDM kernel 2
[x4$1] (t4_2); !threshold for item 4 LCDM kernel 2
[x5$1] (t5_2); !threshold for item 5 LCDM kernel 2
[x6$1] (t6_2); !threshold for item 6 LCDM kernel 2
[x7$1] (t7_2); !threshold for item 7 LCDM kernel 2
![x8$1] (t8_2); !threshold for item 8 LCDM kernel 2
![x9$1] (t9_2); !threshold for item 9 LCDM kernel 2
[x10$1] (t10_2); !threshold for item 10 LCDM kernel 2

%c#5% !for attribute pattern [1,0,0];
[x1$1] (t1_2); !threshold for item 1 LCDM kernel 2
[x2$1] (t2_3); !threshold for item 2 LCDM kernel 3
[x3$1] (t3_3); !threshold for item 3 LCDM kernel 3
[x4$1] (t4_1); !threshold for item 4 LCDM kernel 1
[x5$1] (t5_1); !threshold for item 5 LCDM kernel 1
[x6$1] (t6_3); !threshold for item 6 LCDM kernel 3
[x7$1] (t7_1); !threshold for item 7 LCDM kernel 1
![x8$1] (t8_1); !threshold for item 8 LCDM kernel 1
![x9$1] (t9_1); !threshold for item 9 LCDM kernel 1
[x10$1] (t10_1); !threshold for item 10 LCDM kernel 1

%c#6% !for attribute pattern [1,0,1];
[x1$1] (t1_2); !threshold for item 1 LCDM kernel 2
[x2$1] (t2_3); !threshold for item 2 LCDM kernel 3
[x3$1] (t3_3); !threshold for item 3 LCDM kernel 3
[x4$1] (t4_1); !threshold for item 4 LCDM kernel 1
[x5$1] (t5_1); !threshold for item 5 LCDM kernel 1
[x6$1] (t6_3); !threshold for item 6 LCDM kernel 3
[x7$1] (t7_2); !threshold for item 7 LCDM kernel 2
![x8$1] (t8_2); !threshold for item 8 LCDM kernel 2
![x9$1] (t9_2); !threshold for item 9 LCDM kernel 2
[x10$1] (t10_2); !threshold for item 10 LCDM kernel 2

%c#7% !for attribute pattern [1,1,0];
[x1$1] (t1_2); !threshold for item 1 LCDM kernel 2
[x2$1] (t2_4); !threshold for item 2 LCDM kernel 4
[x3$1] (t3_4); !threshold for item 3 LCDM kernel 4
[x4$1] (t4_2); !threshold for item 4 LCDM kernel 2
[x5$1] (t5_2); !threshold for item 5 LCDM kernel 2
[x6$1] (t6_4); !threshold for item 6 LCDM kernel 4
[x7$1] (t7_1); !threshold for item 7 LCDM kernel 1
![x8$1] (t8_1); !threshold for item 8 LCDM kernel 1
![x9$1] (t9_1); !threshold for item 9 LCDM kernel 1
[x10$1] (t10_1); !threshold for item 10 LCDM kernel 1

%c#8% !for attribute pattern [1,1,1];
```

```
[x1$1] (t1_2); !threshold for item 1 LCDM kernel 2
[x2$1] (t2_4); !threshold for item 2 LCDM kernel 4
[x3$1] (t3_4); !threshold for item 3 LCDM kernel 4
[x4$1] (t4_2); !threshold for item 4 LCDM kernel 2
[x5$1] (t5_2); !threshold for item 5 LCDM kernel 2
[x6$1] (t6_4); !threshold for item 6 LCDM kernel 4
[x7$1] (t7_2); !threshold for item 7 LCDM kernel 2
![x8$1] (t8_2); !threshold for item 8 LCDM kernel 2
![x9$1] (t9_2); !threshold for item 9 LCDM kernel 2
[x10$1] (t10_2); !threshold for item 10 LCDM kernel 2


MODEL CONSTRAINT: !used to define LCDM parameters and constraints
!NOTE: Mplus uses P(X=0) rather than P(X=1) so terms must be multiplied by -1

!STRUCTURAL MODEL:
NEW(g_0 g_11 g_12 g_13 g_212 g_213 g_223);
m1=-(g_11+g_12+g_13+g_212+g_213+g_223);
m2=g_13-(g_11+g_12+g_13+g_212+g_213+g_223);
m3=g_12-(g_11+g_12+g_13+g_212+g_213+g_223);
m4=g_12+g_13+g_223-(g_11+g_12+g_13+g_212+g_213+g_223);
m5=g_11-(g_11+g_12+g_13+g_212+g_213+g_223);
m6=g_11+g_13+g_213-(g_11+g_12+g_13+g_212+g_213+g_223);
m7=g_11+g_12+g_212-(g_11+g_12+g_13+g_212+g_213+g_223);
g_0=-(g_11+g_12+g_13+g_212+g_213+g_223);

!ITEM 1:
!Q-matrix Entry [1 0 0]
!One attibute measured: 1 intercept; 1 main effect
NEW(b1_0 b1_11);    !define LCDM parameters present for item 1
t1_1=-(b1_0);       !set equal to intercept by LCDM kernel
t1_2=-(b1_0+b1_11); !set equal to intercept plus main effect for attribute 1
b1_11>0;            !make sure main effect is positive

!ITEM 2:
!Q-matrix Entry [1 1 0]
NEW(b2_0 b2_e);    !define LCDM parameters present for item 2 b2_e is common
effect
t2_1=-(b2_0);
t2_2=-(b2_0);
t2_3=-(b2_0);
t2_4=-(b2_0+b2_e);
b2_e>0;            !make sure main effect is positive

!ITEM 3:
!Q-matrix Entry [1 1 0]
NEW(b3_0 b3_e);    !define LCDM parameters present for item 3
t3_1=-(b3_0);
t3_2=-(b3_0);
t3_3=-(b3_0);
t3_4=-(b3_0+b3_e);
b3_e>0;            !make sure main effect is positive

!ITEM 4:
!Q-matrix Entry [0 1 0]
!two attibutes measured: 1 intercept; 1 main effects
NEW(b4_0 b4_12);       !define LCDM parameters present for item 4
```

```
t4_1=-(b4_0);
t4_2=-(b4_0+b4_12);
b4_12>0;                            !the order constraints necessary for the
main effects

!ITEM 5:
!Q-matrix Entry [0 1 0]
!two attibutes measured: 1 intercept; 1 main effects
NEW(b5_0 b5_12);       !define LCDM parameters present for item 5
t5_1=-(b5_0);
t5_2=-(b5_0+b5_12);
b5_12>0;                            !the order constraints necessary for the
main effects

!ITEM 6:
!Q-matrix Entry [1 1 0]
NEW(b6_0 b6_e);    !define LCDM parameters present for item 6
t6_1=-(b6_0);
t6_2=-(b6_0);
t6_3=-(b6_0);
t6_4=-(b6_0+b6_e);
b6_e>0;            !make sure main effect is positive

!ITEM 7:
!Q-matrix Entry [0 0 1]
!two attibutes measured: 1 intercept
NEW(b7_0 b7_13);   !define LCDM parameters present for item 7
t7_1=-(b7_0);
t7_2=-(b7_0+b7_13);
b7_13>0;

!ITEM 8:
!Q-matrix Entry [0 0 1]
!two attibutes measured: 1 intercept
!NEW(b8_0 b8_13);   !define LCDM parameters present for item 8
!t8_1=-(b8_0);
!t8_2=-(b8_0+b8_13);
!b8_13>0;

!ITEM 9:
!Q-matrix Entry [0 0 1]
!two attibutes measured: 1 intercept
!NEW(b9_0 b9_13);   !define LCDM parameters present for item 9
!t9_1=-(b9_0);
!t9_2=-(b9_0+b9_13);
!b9_13>0;

!ITEM 10:
!Q-matrix Entry [0 0 1]
!two attibutes measured: 1 intercept
NEW(b10_0 b10_13);   !define LCDM parameters present for item 10
t10_1=-(b10_0);
t10_2=-(b10_0+b10_13);
b10_13>0;

OUTPUT:
    TECH10; !request additional model fit statistics be reported
```

```
SAVEDATA:
    FORMAT IS f10.5;              !format for output file
    FILE IS respondent_m3_eli 3way.dat;
    SAVE = CPROBABILITIES;
```