THE EVIDENCE-BASED MANAGEMENT OF LEARNING: DIAGNOSIS AND DEVELOPMENT OF CONCEPTUAL THINKING WITH MEANING EQUIVALENCE REUSABLE LEARNING OBJECTS (MERLO)

Masha Etkind[1], Ron S. Kenett[2] and Uri Shafrir[3]
[1]Department of Architectural Science, Ryerson University, Canada
[2]KPA Ltd., Raanana, Israel and University of Torino, Italy
[3]Ontario Institute for Studies in Education (OISE); and Knowledge Media Design Institute (KMDI), University of Toronto, Canada
metkind@ryerson.ca

This paper discusses the management of learning for achieving deep understanding of concepts and techniques for the analysis of conceptual situations. Our goal is to to teach not only "how" but also "why". Conceptual understanding is a critical component of evidence-based management. The paper presents Meaning Equivalence Reusable Learning Objects (MERLO), a novel methodology for evaluating and enhancing deep comprehension of the essence of multi-dimensional, complex conceptual situations, often embedded in mixed data-sets. We take a broad view of conceptual understanding and use training of applied statisticians, architecture students, and Continuous Medical Education programs as examples. The next sections provide background to the evidence based management of learning; an introduction to MERLO; a concept mapping of information quality (InfoQ); and discussion and directions for future research.

BACKGROUND

Management decisions are usually based on two major inputs: Available evidence and decision making criteria. Managers' decisions are ostensibly based on objective evaluation of the evidence presented to them. What does 'objective' mean? Often 'objective' means circumstantial evidence: A combination of reliable personal/economical/engineering and market details relevant to the situation. Is there anything missing? There may be no missing data, as 'evidence' is often interpreted to include 'all relevant facts'; which leaves a substantive issue with respect to the nature of the managerial decision making criteria. This issue arises from the statistical - non-deterministic – nature of the processes driving modern markets: Supply uncertainties; variable material and production costs; customer psychology and behavior; etc. In other words: No formulas are available for making 'evidence based', data-driven decisions; rather, interpretation of the data plays an important, often critical, role in making management decisions. The issue becomes critical in view of managers' often lack of understanding and appropriate interpretation of patterns of variations in the data, versus memorizing particular facts. In order to make rational 'evidence based' decisions, managers' decisions must be based not only on 'evidence' but, critically, also on conceptual understanding of the issues involved in the interpretation of the evidence.

Listening to conversations among managers and practitioners in different professional fields, reveal a common trend to clarify statements by repeatedly re-formulating the issue under discussion from alternative points-of-view that share equivalence-of-meaning. These spontaneously and flexibly formulated statements are often encoded in alternative representations of complex issues in different sign systems: a conversation that began with a casual spoken dialog may evolve to include explicit references to numerical data; diagrams; and quotations of written statements in technical reports; that illustrate different aspects of the conceptual situation under discussion. We use the term 'conceptual thinking' to describe ways of thinking that explore patterns of equivalence-of-meaning and associations among ideas, relations, and underlying issues.

Managers who are conceptual thinkers engage in creative discovery of hidden, but potentially viable, relations among concepts, thus testing and extending such patterns of associations that may not be obvious or easily identified. Conceptual thinking also requires the ability, knowledge, and experience to generate novel ideas through alternative representations of

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

shared meaning of complex situations, and to create 'code words' – unique lexical labels of concepts - and procedures for their practical use, nurturing, and further development.

An example from the medical field: Thousand of articles are published weekly with potential impact on how physicians treat patients. In this context:1. The medical profession requires effective Continuing Medical Education (CME) to keep proper professional standards and provide ongoing upgrades to practitioners and clinicians. 2. However, medical professionals are often reluctant to undergo formal testing based on knowledge retention. 3. In addition, the trend is to introduce new Roles and Responsibilities to health care professionals. For example, in Israel, qualified nurses will be allowed to dispense certain drug products and write prescriptions. This requires the development of effective training and education systems. 4. Students in medical schools are exposed to new didactic approaches and methodologies, including competence in conducting regular semantic searches of comprehensive and up-to-date digital knowledge repository of bio-medical science.

Here is another example from the field of architecture: (i) The basis of the process of informing design is conceptual understanding of precedents that constitute case studies for the learning of architecture. (ii) An important expectation in design studio education is that all studio members share and exchange ideas in a creative problem solving process; and that critical questioning and common discussion generates progression that is mutually impacting and jointly enriching. It is in the course of a group discussion that concepts and design ideas are clarified and formulated. (iii) Recognition of meaning equivalence turns into an analytical tool that allows one to understand and compare architectural precedents; facilitates an ability to distil an initial concept's original intent from a final expression; and recognize an original idea behind the complexity of material manifestation of functionality of architectural form. (iv) One gains ability to extract the meaning from complex functionality of built form and space, and recognizes the presence of meaning equivalence in a great variety of architectural expressions.

Another example is in the teaching of applied statistics where the generation, from data, of information useful to decision makers, scientists or engineers is a key objective. In a later section we will expand on how concept mapping, guided by Concept Parsing Algorithms (CPA–see: Shafrir & Etkind, 2009) can be used to assess InfoQ, an approach to assess Information Quality proposed by Kenett and Shmueli (2009).

EVIDENCE-BASED MANAGEMENT OF LEARNING WITH MERLO

Educating managers to develop their conceptual thinking skills require tools for diagnostics and development of conceptual thinking. We describe a novel approach to evidence-based management of learning that provides such tools, and demonstrate its potential in evaluative implementations in several disciplines.

Meaning Equivalence Reusable Learning Object (MERLO) is a multi-dimensional database that allows the sorting and mapping of important concepts through exemplary target statements of particular conceptual situations, and relevant statements of shared meaning. Each node of MERLO is an item family, anchored by a target statement that describes a conceptual situation and encodes different features of an important concept; and also include other statements that may – or may not – share equivalence-of-meaning with the target (Figure 1). Collectively, these item families encode the complete conceptual mapping that covers the full content of a course (a particular content area within a discipline). Figure 1 is a template for constructing an item family anchored in a single target statement.

Statements in the four quadrants of the template - Q1; Q2; Q3; and Q4 - are thematically sorted by their relation to the target statement that anchors the particular node (item family); they are classified by two sorting criteria: *surface similarity* to the target, and *equivalence-of-meaning* with the target. For example, if the statements contain text in natural language, then by 'surface similarity' we mean same/similar words appearing in the same/similar order as in the target statement; and by 'meaning equivalence' we mean that a majority in a community that shares a sublanguage (Cabre, 1998; Kittredge, 1983) with a controlled vocabulary (e.g., statistics) would likely agree that the meaning of the statement being sorted is equivalent to the meaning of the target statement.

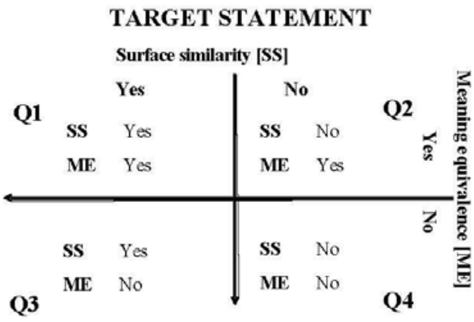

Figure 1. Template for constructing an item-family in MERLO

MERLO Pedagogy and Tests

MERLO pedagogy guides sequential teaching/learning episodes in a course by focusing learners' attention on meaning. The format of MERLO items allows the instructor to assess deep comprehension of conceptual content by eliciting responses that signal learners' ability to recognize and produce multiple representations that share equivalence-of-meaning. A typical MERLO item contains 5 unmarked statements: an unmarked target statement plus four additional (unmarked) statements from quadrants Q2; Q3; and Q4. Our experience shows that inclusion of statements from quadrant Q1 makes the item too easy, because it gives away the shared meaning due to the valence match between surface similarity and meaning equivalence, a strong indicator of shared meaning between a Q1 and the target statement. Task instructions for MERLO test are: At least two out of these five statements–but possibly more than two–share equivalence-of-meaning.1. Mark all statements–but only those–that share equivalence-of-meaning.2. Write down briefly the concept that guided you in making these decisions.

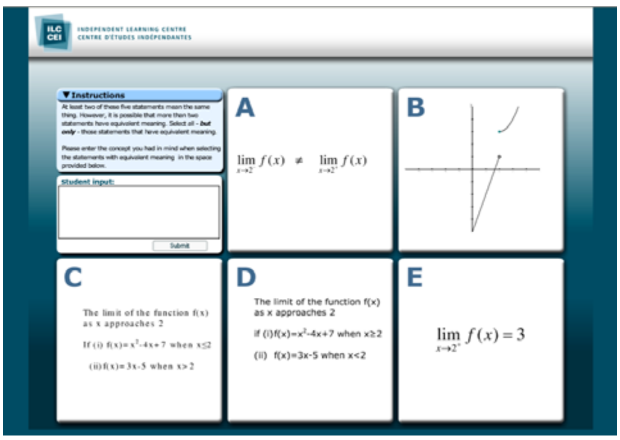

Figure 2. Example of MERLO item (mathematics/functions)

For example, the MERLO item in Figure 2 (mathematics/functions) contains 5 representations that include text, equations, tables, and diagrams; at least two of these representations share equivalence-of-meaning. Thus, the learner is first asked to carry out a recognition task in situations where the particular target statement is not marked, namely, features of the concept to be compared are not made explicit. In order to perform this task, a learner needs to begin by decoding and recognizing the meaning of each statement in the set. This decoding process is carried out, typically, by analyzing concepts that define the 'meaning' of each statement. Successful analysis of all the statements in a given 5-statement set (item) requires deep understanding of the conceptual content of the specific domain. MERLO item format requires both rule inference and rule application in a similar way to the solution of analogical reasoning items. Once the learner marked those statements that–in her opinion–share equivalence-of-meaning, she formulates and briefly describes the concept/idea/criteria she had in mind when making these decisions.

MERLO diagnostic scoring algorithms

Learner's response to a MERLO item combines multiple-choice/multiple-response (*recognition*); and short answer (*production*). Subsequently, there are two main scores for each MERLO item: *recognition score*; and *production score*. Specific comprehension deficits can be traced as depressed *recognition scores on quadrants Q2 and Q3*, due to the mismatch between the valence of surface similarity and meaning equivalence (Figure 1). *Production score* of MERLO test items is based on the clarity of learner's description of the conceptual situation anchoring the item, and the explicit inclusion in that description of lexical labels of relevant and important concepts and relations.

Evaluative implementations of MERLO documented enhanced learning outcomes (Etkind & Shafrir, 2010; Shafrir & Etkind, 2006). Such implementations were carried out since 2000 at different instructional situations and in different disciplines, including: Russian Academy of Sciences (mathematics; physics; and biology); University of Toronto (teacher training; developmental psychology); Ryerson University (architecture; project management); George Brown College (English as a second language; learning disabilities); Independent Learning Centre of TV-Ontario (mathematics); and Ontario Mining and Manufacturing Centre of Excellence (risk management).

Classroom implementation of MERLO pedagogy includes interactive MERLO quizzes, as well as inclusion of MERLO items as part of mid-term tests and final exams. *MERLO interactive quiz* is an in-class procedure that provides learners with opportunities to discuss a PowerPoint display of a MERLO item in small groups, and send their individual responses to the instructor's computer via mobile text messaging, or by using a clicker (CRS–Classroom Response System). Such a quiz takes 20-30 minutes, and includes the following 4 steps: Small group discussion; individual response; feedback on production response and class discussion; feedback on recognition response and class discussion.

Following several evaluative implementations (Shafrir & Etkind, 2006), we define *good vs. poor* conceptual thinkers as those students who score high (vs. low) on *both recognition and production on MERLO items*. In order to clearly identify good vs. poor conceptual thinkers, we convert MERLO raw scores for recognition and for production to Z-scores (standard scores with mean = 0.0, and standard deviation = 1.0), and define good and poor conceptual thinkers by performing a double median split of their MERLO recognition and production Z-scores in an exam in a core course in the discipline, as follows: *Good conceptual thinkers* are defined as those who score high (above the median) on MERLO items in *recognition* of different representations that share equivalence-of-meaning of at least 2 (out of 5) representations; as well as in *production* of a written, brief description of the concept (or idea), in the learner's mind, that led the learner to make those recognition decisions. In contrast, *poor conceptual thinkers* are those students who score low (at or below the median) on *both recognition and production*.

Results of a recent implementation of MERLO pedagogy at the Faculty of Engineering, Architecture and Science in Ryerson University in Toronto (Etkind & Shafrir, 2010), lend support to the following conclusions: 1. *MERLO interactive quizzes enhance conceptual thinking; therefore, the initial difference between good and poor conceptual thinkers in both recognition and production scores increases as they experience more MERLO interactive quizzes 2. Good vs. poor conceptual thinkers score high (low) on deep comprehension of the conceptual content of the course, as measured by their marks on an essay written as part of the final exam.3. Good vs. poor conceptual thinkers score high (low) on deep comprehension of the content of other courses.4. Interactive MERLO pedagogy, when implemented continuously in subsequent semesters as a regular part of the instructional methodology, replicates the above pattern of results.5. Conceptual thinking is learnable.* These results support our underlying rationale that inclusion of *MERLO interactive quizzes,* and emphasis on conceptual thinking in class activities, enhance learning outcomes and conceptual comprehension.

In our latest study, 30 students were classified as *'poor conceptual thinkers'* in May of 2009 as a result of their low (at or below median) Z-scores in a double-media-split on MERLO recognition and production; twenty five (out of the 30) students participated in a subsequent 2nd year course 'ASC 306: Ideas, technology, and precedents II'. Comparative analysis of the Z-scores on recognition and production of each of these 25 students reveals that most of them indeed improved their Z-scores on recognition *and/or* production in the final exam of this course. Specifically, these improvements were significant enough to take 14 of these 25 students out of the category of *'poor conceptual thinkers'*. The levels of conceptual thinking skills these 25 students in December 2009–who were classified as *'poor conceptual thinkers'* in May 2009–were as follows: 4 students improved *both their recognition and production Z-scores* to become *'good conceptual thinkers';* 6 students significantly improved only their recognition Z-score (but not their production Z-score); as a result, these students were *no longer classified as 'poor conceptual thinkers';* 4 students significantly improved only their production Z-score (but not their recognition Z-score); as a result, these students were *no longer classified as 'poor conceptual thinkers';* finally, 11 students did not improve significantly neither their recognition nor their production Z-scores; therefore, *these students remained 'poor conceptual thinkers'*

Concept maps of Information Quality (InfoQ)

Statistics and data mining are disciplines focused on extracting knowledge from data. They provide a toolkit for answering questions of interest, for predicting new observations, and for summarizing data efficiently. In both fields, observable data is used to derive knowledge. We focus here on using data in the context of statistical modelling or data mining, and on assessing its utility. The general term *empirical analysis* refers here to both statistical analysis and data mining. In particular, we consider the application of a technology (empirical analysis) to a resource (data) for a given purpose. In this sense, we follow Hand's (2008) definition of statistics as a technology: "*The technology of extracting meaning from data*". Kenett and Thyregod (2006) present the statistical consulting cycle to solve real problems as a five steps cycle: 1) Problem elicitation, 2) data collection, 3) Data analysis, 4) Findings formulation and 5) Findings presentation. Academia typically focuses only on step 3), the data analysis step. The application of statistics requires the full cycle. Practical Statistical Efficiency (PSE) and Information Quality (InfoQ), introduced below, addresses the full life cycle. InfoQ can be assessed ex-ante and PSE is assesses ex-post (Kenett, 2007). Practical Statistical Efficiency can be assessed after a specific research project has been completed. The PSE formula of Kenett et al. (2003) accounts for eight components and is computed as: $PSE = V\{D\} \times V\{M\} \times V\{P\} \times V\{PS\} \times P\{S\} \times P\{I\} \times T\{I\} \times E\{R\}$
Where: $V\{D\}$ *= value of the data actually collected.* $V\{M\}$ *= value of the statistical method employed.* $V\{P\}$ *= value of the problem to be solved.* $V\{PS\}$ *= value of the problem actually solved.* $P\{S\}$ *= probability level that the problem actually gets solved.* $P\{I\}$ *= probability level the solution is actually implemented.* $T\{I\}$ *= time the solution stays implemented.* $E\{R\}$ *= expected number of replications.* These components can be assessed qualitatively, using expert opinions, or quantitatively if the relevant data exists. Kenett and Shmueli (2009) propose a general concept of Information Quality (InfoQ) to assess how well, in a specific context, meaning has been extracted from data. "InfoQ is the quality of the information derived from data relative to the analysis goals and objectives. In other words, it is determined by the manner in which a dataset addresses a specific (scientific or practical) goal using empirical analysis" (Kenett & Shmueli, 2009). The eight dimensions of InfoQ identified in Kenett and Shmueli (2009) are: *1) Data granularity. 2) Data structure. 3) Data integration. 4) Temporal relevance. 5) Sampling bias.6) Chronology of data and goal. 7) Concept operationalization. 8) Communication and data visualization.*

We now proceed to discuss how concept mapping can be applied to InfoQ. A *Concept* is an organizing principle behind a collection of facts in context in a domain of knowledge; an invariant. InfoQ is a concept. *Concept Parsing Algorithms* (CPA – see: Shafrir & Etkind, 2009) is a procedure for identifying individual features in each of the three sets of building blocks that define the meaning of an *emergent super-ordinate concept*, uniquely indentified by its lexical label C'; they are: Set [Ci] of co-occurring sub-ordinate concepts, each uniquely indentified by its own lexical label (must contain at least 2 concepts); set [Rj] of relations between co-occurring concepts, and between co-occurring concepts and the emergent super-ordinate concept (may be an empty set); and set [Lk] of linguistic elements, descriptors that must obey syntactic, morphological, and grammatical rules of the language (may not be an empty set).

InfoQ is a lexical label of super ordinate concept [C']. From the InfoQ dimensions mapped in Kenett and Shmueli (2009) we can derive eight Co-occurring subordinate concepts [C1, C2,…,C8]: *C1= Data granularity; C2= Data structure; C3= Data integration; C4= Temporal Relevance; C5= Sampling Bias; C6= Chronology of Data and Goal; C7= Communication and Data Visualization; C8= Concept Operationalization.* The questions one asks to assess the InfoQ dimensions provide examples of the linguistic elements [L1, L2,……,Lk]. As a result one can determine the meaning of InfoQ. Based on such concept mapping, concept statements and target statements are derived. These form the basis of a MERLO assessment, as described above.

DISCUSSION AND CONCLUSIONS

The need for higher thinking skills has recently been recognized as a cornerstone of effective learning–'*understanding facts and ideas in the context of a conceptual framework*' (Bransford, 2000). *MERLO pedagogy* response to this need is direct and explicit. It is based on offering learners frequent, in-class, *interactive formative MERLO quizzes* with informal opportunities to discuss and recognize conceptual situations that share *equivalence-of-meaning*, as well as of producing written descriptions of these conceptual situations through representations in multiple sign systems (e.g., text; images; diagrams; equations). In subsequent *summative assessments* (e.g., mid-term, final) that include MERLO items, learners demonstrate mastery of learning and conceptual thinking skills. *MERLO Pedagogy* provides diagnostic tools that allow the instructor to manage the learning process by: (i) identifying *good vs. poor conceptual thinkers; (ii)* identifying in individual learners *'soft conceptual spots'* and provide specific hints for remediation of poor conceptual thinking; (iii) facilitate content learning and enhance learning outcomes; (iv) encourage and facilitate the development of good conceptual thinking. *MERLO pedagogy provides the essential tools for evidence-based management of learning.* Outcomes of MERLO assessments include detailed descriptions of the real-time evolution of an individual learner's process of understanding. Inclusion of MERLO assessment outcomes in individual learners' e-Portfolios provides an authentic and convincing evidence of mastery of learning.

REFERENCES

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.

Cabre, M. T. (1998). *Terminology: Theory, Methods, and Applications*. Amsterdam: Benjamins.

Etkind, M., & Shafrir, U. (2010). Development of Conceptual Thinking in Architecture Education: Interactive Pedagogy with MERLO. Manuscript in preparation.

Hand, D. J. (2008), *Statistics: A Very Short Introduction*, Oxford University Press.

Kenett, R. S., Coleman, S. and Stewardson, D. (2003), Statistical Efficiency: The Practical Perspective, *Quality and Reliability Engineering International*, *19*, 265-272.

Kenett, R. S. and Thyregod, P. (2006), Aspects of statistical consulting not taught by academia, *Statistica Neerlandica*, *60*(3), 396-412.

Kenett, R. S. (2007), Practical Statistical Efficiency. In F. Ruggeri, R. S. Kenett & F. Faltin (editors in chief), *Encyclopedia of Statistics in Quality and Reliability*, Wiley.

Kenett, R. S. and Shmueli, G. (2009), On Information Quality, http://ssrn.com/abstract=1464444, Submitted.

Kittredge, R. I. (1983). Semantic Processing of Texts in Restricted Sublanguages. In N. J. Cercone (Ed.), *Computational Linguistics*, pp. 45-58.

Shafrir, U., & Etkind, M. (2006). E-Learning for Depth in the Semantic Web. *British Journal for Educational Technology*, *37*(3), 425-444.

Shafrir, U., & Etkind, M. (2009). Concept Science: Content and Structure of Labeled Patterns in Human Experience (version 30.0). Manuscript in preparation.