Assessing Progress of Learning in Complex Domains

J. Michael Spector
The University of Georgia, USA

Abstract

The specific problem addressed in this paper is the lack of assessment and feedback tools appropriate to support learning and instruction in complex and dynamic task domains. In such domains, representative problems can be identified but they often lack single solutions, and there is a great deal of variability between and among problems. As a consequence, in order to determine progress of learning and the efficacy of interventions, one must consider not only the solutions developed by learners but what they were thinking about when developing those solutions. Previously, think-aloud protocol analysis was used for the latter (Ericsson, 2001), but this methodology is researcher intensive and does not scale up for use in instructional settings. What are needed are tools that can be used in a variety of instructional settings, including classrooms, online settings, and informal learning environments. A review of what has been done in the past will be presented along with the most promising technologies. A specific approach which involves capturing a person’s conceptualization of a complex problem-solving situation and comparing that with an expert representation will be presented as a promising instructional and research solution to the general problem of assessing learning in complex domains.

Keywords: assessing progress of learning, complex domain, concept mapping, instructional feedback, similarity metrics
1. Introduction

Assessment is a core aspect of instruction. There are two general forms of assessment: formative and summative. Formative assessment can inform a learner how well he or she is doing based on recognized goals, objectives, standards, prior and current performance. Formative assessment can also inform an instructor how well a lesson or unit or instructional sequence is working with a particular group of learners. On the other hand, summative assessment can inform a learner how well he or she did based on expected outcomes and performance measures. Summative assessment can also inform an instructor and other instructional planners how well a certain course did in terms of achieving its objectives. The focus in this paper is on formative assessment. The primary concern here is about the means and methods (a) for providing learners with timely and informative feedback with regard to their developing knowledge and expertise in complex domains, and (b) for providing instructors and instructional planners with timely and informative feedback with regard to the effectiveness of learning activities and the needs of learners.

Formative assessment is critical to the success of individual learners as well as the success of lessons and courses. The expectation is that learners will progressively acquire knowledge and skills. Formative assessment focuses on the dynamic development of knowledge and skills – the progress of learning. Learning is defined herein as involving stable and persisting changes in what a person knows and can do. That is to say that learning involves changes in a person’s abilities, attitudes, beliefs, behavior, habits, knowledge, or skills. There are several overlooked aspects of this relatively standard definition of learning. First, there is the fundamental notion of change. To make a claim that learning has occurred, we need to know (have measured or have reliable indicators of) both a before and after state of affairs (abilities, attitudes, beliefs, etc.). Second, in order to measure the change in a progressive or desirable manner, we need to have an established target such as a standard or expert reference model. Third, we should have some basis for believing that the observed change(s) will persist for some time.

In addition, it is beneficial when doing an assessment to understand the kind of thing(s) being learned. A formative assessment that might be appropriate for detecting progress of learning historical facts (a relatively simple task) might involve the recall of facts and the identification of key dates and events. Such an assessment would not likely be appropriate for detecting progress of learning strategies to promote the peaceful resolution of disputes between countries (a very complex task).

This leads us to the concept of simple and complex tasks and task domains. It is clearly the case that ‘simple’ and ‘complex’ in this context are somewhat vague terms involving some aspect of subjectivity (what is simple for one learner may seem quite complex for another). For the sake of this discussion, a simple domain (or lesson or task) is vaguely and simply defined as one involving (a) primarily declarative knowledge (facts and concepts), and (b) limited procedural knowledge in the form of applying basic declarative knowledge in the performance of straightforward procedures and activities; the procedural knowledge in these cases do not involve
much variation from one problem to the next, do not require significant analysis to
determine what to do next, and do not undergo significant change during the problem
solving process. On the other hand, complex domains involve significant procedural
knowledge (knowing how to perform a variety of tasks) as well as causal or
contextual knowledge (knowing why things have happened or are likely to happen in
different situations).

Van Merriënboer (1997) distinguished recurrent from non-recurrent tasks. Recurrent
tasks are performed about the same way regardless of changes in the surrounding
circumstances. Non-recurrent tasks involve analysis of the situation and adjustments
to the solution approach based on the analysis. Simple domains are those in which
recurrent tasks predominate; complex domains involve a significant number of non-
recurrent tasks.

Sterman (1994) distinguished complex, dynamic problems from simpler problems.
Complex, dynamic problems are characterized as those with many interrelated and
interacting factors, with relationships among factors that may change over time, and
with relationships that may be non-linear and involve delayed effects. As it happens,
many things are complex and dynamic in this way, including the human body, an
economic system, an environmental plan, and many more.

Yet another way to differentiate simple from complex domains is with regard to the
type of frequently encountered well-structured and ill-structured problems (Newell &
Simon, 1972; Simon, 1973). A complex domain is one in which there are significant
ill-structured problems to be solved. An ill-structured problem is one in which there
may be one or more unknown input factors, or the goal may be vague, or it is not clear
how to transform the given situation into the desired situation. Complex problems
involve some uncertainty with regard to what to do to achieve desired outcomes
and/or how to assess those outcomes. Rittel and Webber (1973) used the term ‘wicked
problem’ to refer to a problem that lacked a well-defined outcome state. Jonassen
(1997, 2000, 2004) has developed a taxonomy of problems that includes several types
and degrees of ill-structured problems (e.g., policy analysis, design problems, and
dilemmas).

The boundaries between simple and complex domains may not be perfectly clear or
well-defined, but there are clearly important differences between algorithmic,
recurrent problems involving only static factors and dynamic, non-recurrent problems
involving interrelated and unknown factors. A critical aspect of complex problems is
that they typically lack a standard solution and in many cases lend themselves to
multiple solution approaches. This fact introduces a significant challenge for
assessment which will be addressed in the next section. The point of this brief review
of the key terms involved in this paper is to locate the emphasis herein on formative
assessment in the context of providing timely and informative feedback to learners
with regard to the progressive development of their abilities to solve complex and
challenging problems successfully.
2. Assessing Complex Learning

As already suggested, examples of complex, ill-structured problems are found in many domains, including engineering design and the bio-medical sciences. Engineering design examples include: (a) improving the heat shielding on the space shuttle to allow safer re-entries with greater payloads; (b) developing new traffic flows across Chesapeake Bay to reduce congestion, open new areas along the coast and maintain naval security; and (c) designing Web-accessible databases that provide security, redundancy and ease of use. Examples in the bio-medical domain include: (a) controlling the spread of the human immunodeficiency virus, (b) responding effectively to critical care emergencies involving multiple victims of a terrorist attack, and, (c) diagnosing and treating patients with histories of chronic pain and allergies.

In most of these examples, the final goal state is not completely defined, although in some cases the goal would become more well-defined in an actual situation (e.g., with an actual patient). In some cases, conditions and constraints are not completely identified, although in actual cases (e.g., a critical care emergency) conditions and constraints would become more completely defined as the emergency evolves. What effect specific interventions might have on transforming actual conditions into desired goal states is also unclear in many of these cases (e.g., a particular treatment regimen with a specific patient). With these and other examples, there is some uncertainty with regard to the effects of particular interventions or solutions. While many actual problems may be less ill-structured and somewhat less complex than these examples, several key aspects create challenges for learners and those responsible for helping learners in developing competence and confidence in solving complex problems – a learning goal that can be called deep understanding:

- dynamics – the problem situation evolves due to changing circumstances and new technologies, creating the potential for alternative approaches and solutions;
- expertise – what is perceived to be complex or ill-structured by a person with limited training and experience may be perceived as less complex and more well-structured by a person with many years of professional experience;
- collaboration – complex, ill-structured problems in everyday settings are typically resolved by teams rather than by individuals; the collective and changing knowledge and experience of team members is relevant to team performance.

These aspects of complex, ill-structured problems require: (a) a method to assess learning and performance that recognizes that problem approaches and conceptualizations are significant since standard solutions for these problems do not exist; (b) metrics to determine levels of performance and expertise that can provide a basis for formative assessment; and, (c) support for collaborative teams engaged in complex and challenging problem-solving tasks. In the remainder of this paper, the focus is on the first two of these issues.

Complex, ill-structured problems are pervasive and especially prominent in such
socially significant areas as crisis management, economic planning, emergency medical care, environmental engineering, epidemiology, public health policy, scientific discovery, strategic management, and technology development. Resolving complex problems in these areas is of obvious and high importance to society. Sustained economic and social progress in the information age requires a highly trained workforce able to respond flexibly and intelligently to changing circumstances. How best to prepare individuals to be effective problem solvers in these areas is a challenge that requires reliable means to assess progress of learning and determine levels of experience and performance with regard to problems that lack single solutions and lend themselves to alternative problem-solving approaches. Yet the means and methods to assess individual progress of learning in these complex domains are very well developed.

3. A Model-based Approach

One solution approach involves an innovative, research-based assessment method that integrates a cognitive model of mental model development and the acquisition of expertise with a systems perspective of learning and performance (Spector & Koszalka, 2004). An early model of human cognitive architecture by Anderson (1983) emphasized connections among input and processing mechanisms but did not address how internal representations accumulate and become more sophisticated. Collins (1991) also did not address this aspect of internal representations of complex problems but the cognitive apprenticeship model of instruction assumes that such growth and develop occur. One way to represent a cognitive model of learning, thinking and knowing is to consider two important human abilities noted by Ludwig Wittgenstein, one from the *Tractatus Logico-Philosophicus* and a second from *Philosophical Investigations*. The first ability Wittgenstein (1922) noted concerned the fact that humans create internal representations of things we experience – ‘we picture facts to ourselves’ is closer to the language Wittgenstein actually used. Of course we also picture things that are not facts to ourselves. Fundamentally, this is the core concept of a naturalistic epistemology now referred to as constructivism. We construct internal representations in order to make sense of our experiences. Cognitive psychologists call these internal representations mental models. The closing remark in the *Tractatus* is that what we cannot speak about clearly we must pass over in silence (Wittgenstein, 1922). This conclusion proved unsatisfactory to Wittgenstein – we do in fact speak about many things that lack clear and distinct foundations. He went on to explore the many things we believe and understand in *Philosophical Investigations*. These beliefs are ultimately based on internal representations but they also involve further development of our thinking through what he called language games (Wittgenstein, 1953). Humans have the amazing abilities to create internal representations of things they experience and to talk about those representations with others.

To account for the development of knowledge, cognitive psychologists have developed similar ideas in the form of mental models (Johnson-Laird, 1983; Seel, 1991) and distributed cognition (Hutchins, 1995; Salomon, 1993). For the purpose of
this discussion, a mental model is an internal representation created just when needed to make sense of events, experiences, phenomena or unusual circumstances. Mental models are hypothetical entities that are not directly observable – not even your own. Fortunately, it is possible share external representations of mental models and combine these with other information as the basis of actions, decision, and problem solving, which can be considered as a form of distributed cognition.

Not all mental models are equally useful for everyone in promoting effective action, decision making, and problem solving (Kalyuga, 2007; Sweller, van Merriënboer, & Paas, 1998). Two general approaches address the problem of encouraging productive mental models; both involve creating and sharing external representations of mental models, and the two approaches may be combined. One approach is explicitly collaborative; the basic approach is to get those involved talking about their conceptualizations in the context of a problem situation with the expectation that useful thinking (mental models) will develop. The other approach is to make use of the external representations of an expert to guide individual (or group) thinking and problem solving (Shute et al., 2009; Spector, 2006). It is likely that a combination of these two approaches is likely to prove optimal in many cases, but this is as yet not established by empirical research (Stoyanova & Kommers, 2002. The remainder of this section focuses on a method to elicit external representations of mental models and compare them with a reference model, such as that elicited from an expert. The methodology was developed in a prior National Science Foundation (NSF) project, and has now been used in about a dozen studies.

The NSF project entitled “The DEEP Methodology for Assessing Learning in Complex Domains” (Spector & Koszalka, 2004), examined the feasibility of using annotated problem representations to determine relative level of expertise in three domains: biology, engineering and medicine. That study involved the selection of two representative problems for each complex task domain. Subjects (both expert and non-expert respondents) were provided with a problem scenario (typically about one page in length) and asked to indicate things that they thought would be relevant to a solution. Subjects were asked to document these items and provide a short description of each item along with a brief explanation why it is relevant. They were also asked to indicate and document other assumptions about the problem situation that they are making (initially and again at the end of the activity). Subjects were asked to develop a solution approach to the problem. Required parts of this representation included: (a) a list of key facts and factors influencing the problem situation; (b) documentation of each causal factor - what it is and how it influences the problem; (c) a graphical depiction of the problem situation (see Figure 3); (d) annotations on the graphical depiction (descriptions of each item - linked to factors already described when possible); (e) a solution approach based on the representations already provided; and, (f) an indication of other possible solution approaches.
Findings suggest that the method can be used to predict performance and relative level of expertise in some complex domains. Expert conceptualizations in each of the domains examined were noticeably different from non-expert responses, although there were differences among expert responses and much variation in non-expert responses. Differences occurred at all three levels of analysis (surface, structure, semantic). In general, experts tended to identify more inter-linked factors and generally said more about each factor and link (although there were some exceptions to this pattern). In most cases, experts tended to identify more causal relationships as opposed to other types of relationships, although this was not the case with medical diagnosticians. Debriefing with expert medical respondents and the domain specialist indicated that the experts were very familiar with standard diagnostic procedures and used that knowledge to quickly develop a diagram reflecting the diagnostic process whereas non-experts (medical school students) had knowledge of the human body and used that knowledge to reason through a sequence likely to result in a successful diagnosis. Table 1 reflects this difference in expert and non-expert medical reasoning with experts identifying significantly more process links and non-experts identifying significantly more correlation links.
<Table 1> Sample of summary data for a medical problem scenario

<table>
<thead>
<tr>
<th>Experts (5)</th>
<th>Non-experts (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of link</td>
<td>Links</td>
</tr>
<tr>
<td>cause/effect</td>
<td>73</td>
</tr>
<tr>
<td>Example</td>
<td>0</td>
</tr>
<tr>
<td>Correlation</td>
<td>8</td>
</tr>
<tr>
<td>Process</td>
<td>25</td>
</tr>
<tr>
<td>total links</td>
<td>100</td>
</tr>
</tbody>
</table>

Tables 2 and 3 show that experts and non-experts exhibited noticeable differences in the key nodes identified. Experts (Table 2) tended to identify the same critical nodes as indicated by close percentages associated with the four most cited nodes and the similarity in numbers of links associated with those critical nodes. Non-experts (Table 3) differed significantly from experts (and from each other) in what they believed to be critical to resolving the problem situation. None of the experts cited stress as a critical factor yet many non-experts did. Expert medical diagnosis was driven by evidence based on tests as the most critical factor and experts also mentioned follow-up visits and tests, while non-experts did not.

<Table 2> Sample nodes cited by medical experts (N = 5; Links = 128)

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Links</th>
<th>From / To</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Tests</td>
<td>19</td>
<td>8 / 11</td>
<td>17.93%</td>
</tr>
<tr>
<td>2 Differential diagnosis</td>
<td>16</td>
<td>10 / 6</td>
<td>15.09%</td>
</tr>
<tr>
<td>3 Diagnosis</td>
<td>14</td>
<td>8 / 6</td>
<td>13.21%</td>
</tr>
<tr>
<td>4 History</td>
<td>10</td>
<td>8 / 2</td>
<td>9.43%</td>
</tr>
<tr>
<td>4 Post-visit additional info</td>
<td>10</td>
<td>3 / 7</td>
<td>9.43%</td>
</tr>
<tr>
<td>4 Clinical knowledge</td>
<td>10</td>
<td>7 / 3</td>
<td>9.43%</td>
</tr>
</tbody>
</table>

<Table 3> Sample nodes cited by medical non-experts (N = 16; Links = 147)

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Links</th>
<th>From / To</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 History of present illness</td>
<td>67</td>
<td>31 / 36</td>
<td>21.07%</td>
</tr>
<tr>
<td>2 Hypothesis testing</td>
<td>55</td>
<td>17 / 38</td>
<td>17.30%</td>
</tr>
<tr>
<td>3 Chief complaint</td>
<td>49</td>
<td>14 / 35</td>
<td>15.41%</td>
</tr>
<tr>
<td>4 Patient’s background/ social history</td>
<td>45</td>
<td>27 / 18</td>
<td>14.15%</td>
</tr>
<tr>
<td>5 Stress</td>
<td>44</td>
<td>23 / 21</td>
<td>13.84%</td>
</tr>
</tbody>
</table>

In this one-year study, there was no opportunity to examine changes in problem representations after an instructional sequence or period of sustained deliberate practice. The goal was to determine if the annotated problem representation methodology was suitable for use in multiple domains, if it would show differences in expert and non-expert responses, and whether or not it could provide a basis for assessing relative level of expertise. These goals were achieved. The logical next steps are to investigate use in assessing changes in how individuals represent problems and to integrate the method into personalized feedback for problem solving activities. This
implies using the method before, during and after instructional sequences and periods of deliberate practice. Furthermore, the method can be used to assess team problem solving and has potential use in providing personalized and high level feedback to individuals and groups that is likely to improve metacognitive skills and support self-regulation.

The DEEP methodology has the potential to scale up for use in educational and performance settings involving many individuals whereas the more traditional think-aloud protocol analysis methodology is suitable only for examining a small number of individuals in order to investigate particular hypotheses with regard to learning and performance. The DEEP methodology has the additional advantage of being easy to learn and implement, which makes it potentially suitable for classroom and workplace use. Further refinements of the methodology and the associated tool, including extensions for use with small groups and problem-solving teams, are required. Moreover, investigations of more problems in more domains with persons at different levels of knowledge and skill are required in order to develop more precise and reliable assessment metrics. Finally, this tool is useful in revealing differences in types of problems and how they are perceived by problem solvers. Such knowledge is pertinent to understanding the variety of complex, ill-structured problems encountered in different domains. These issues form the core set of questions to be investigated in this follow-on study.

Variations of this methodology have been effectively demonstrated in simpler domains (Herl et al., 1999; Novak, 1998; Schvaneveldt, 1990; Taricani & Clariana, 2006). Those who have employed an analogous method for simpler learning tasks have relied on:

- simple quantitative measures for measures of similarity to expert responses (e.g., presence/absence of salient features and their location in a concept map), which do not provide insight into progressive development of expertise and improvement in higher-order reasoning, especially in complex, ill-structured problem-solving domains; and,

- qualitative analyses of responses, which are time-intensive and costly and, consequently, which are hardly ever used when a laboratory effort scales up to full-scale implementation and which are not useful for assessing large numbers of individuals or evaluating programs.

DEEP asks the learner to construct an annotated problem representation to determine how a learner or group of learners is thinking about that problem situation. Once the learner (or group of learners) constructs that diagram, it can be compared with one created by an expert practitioner and feedback provided to the learner. As these problem scenarios are accumulated over a sequence of instructional units or sustained period of deliberate practice, it is possible to determine if learner responses increasingly reflect the same kinds of considerations (factors, complexity of interactions among factors, and explanations of relationships) found in expert responses. More specifically, it is in principle possible to use this method to predict how problem representations will develop and change in individuals and groups,
which is consistent with Seel’s (2003) studies of progressive mental model development.

Since the initial development and validation of the DEEP annotated concept-map assessment methodology, two other tools were developed based on similar theoretical and empirical foundations, and all three tools are now integrated in a common Web-based environment called HIMATT (Highly Integrated Model-based Assessment Tools and Technologies; see Parnay-Dummer, Ifenthaler, & Spector, 2010; Shute et al., 2009). Collectively, HIMATT represents a research tool that is designed to elicit a problem conceptualization from a person either in the form of an annotated concept map (as in DEEP) or as open text which is then transformed automatically into an association network. Two concept maps or association networks can be compared using six similarity metrics within the system (Ifenthaler & Seel, 2005). The similarity metrics include three kinds of analysis: (a) an analysis based on the surface structures of the representations (e.g., node-link-node number and combinations), (b) matching structure analysis (e.g., number of same nodes and node-link-node combinations), and (d) deep structure analysis based on the semantic similarity of nodes and node-link-node combinations. One of the most consistent yet surprising results of the many studies conducted using these tools is that the connectedness of an association network or concept map is a strong predictor of relative level of expertise (Kim, 2008; Lee, 2008; McKweon, 2008; Smith, 2008; Spector & Koszalka, 2004). This is an especially promising as it suggests that a time-consuming protocol analysis is not always needed to determine relative levels of expertise.

While DEEP and HIMATT were developed as tools for assessing learning progress, there is significant potential for their use in supporting active learning, especially as feedback mechanisms, support for formative assessment, and the means to develop metacognitive reasoning. These possibilities are explored in the next section.

4. Concluding Remarks

There is much that remains to be done to further develop effective tools and technologies for assessing progress of learning in complex domains (Ifenthaler, Parnay-Dummer, & Seel, 2007; Ifenthaler, Parnay-Dummer, & Spector, 2008). It is obviously important to be able to determine if particular programs of instruction in domains such as engineering design and environmental planning are contributing to the development of expertise. Knowledge-based tests alone are not sufficient in this regard. Portfolios are helpful but they only represent an artifact or product developed by an individual and not how that individual thinks or might approach other projects and problems. Gaining an insight into problem solving thought processes remains an important practical task as well as an important research area. Other tools and approaches for assessing complex problem solving should be explored, and further studies should be conducted using HIMATT and other advanced tools.

Of particular importance is the need to develop tools and technologies to support instructors in providing timely and informative feedback on learning progress. All of
the tools mentioned in this paper were developed for research purposes and do not lend themselves easily for use by instructors in classroom and virtual learning environments. However, the potential for these tools to provide a learner with near real-time feedback comparing his or her problem conceptualization with that of an expert or another learner exists now and is of obvious value in promoting the development of productive mental models to solve complex problems. I strongly encourage such development and will gladly contribute to such efforts.

References

Lee, J. (2008). Effects of model-centered instruction and levels of learner expertise on


