learning assessment is providing the evidence that researchers, educators and policy makers need to improve 21st century science learning. We therefore envisage a lively debate with members of the audience led by our discussant Cindy Hmelo-Silver.

Assessment and inquiry; issues and opportunities

Ton de Jong & Pascal Wilhelm, Faculty of Behavioral Sciences, University of Twente, the Netherlands
Email a.j.m.dejong@utwente.nl, p.wilhelm@utwente.nl

Inquiry learning is an educational approach that involves a process of exploration, asking questions and making discoveries in the search for new understandings (National Science Foundation, 2000). In a typical (computer based) inquiry learning task, learners conduct experiments to test hypotheses about the relationships between variables in a particular knowledge domain (de Jong, 2006). Inquiry learning tasks vary in the constraints they pose to learners. Tasks may vary from open-ended, self-paced tasks in which learners follow their own particular inquiry paths, generating their own questions and hypotheses to tasks in which research questions and hypotheses are defined by an instructor. Although any particular study takes a stance somewhere along this continuum, there are still many routes possible for learners during the learning process and what is learned may differ between students. As a result a variety of types of learning outcomes are possible, ranging from different types of knowledge to specific skills. Assessing these can be done after the learning process outside the learning environment but also on-line during the process on the basis of the learners' interaction with the inquiry environment and the products (e.g., hypotheses, models) produced. In case of collaborative learning chat data can be included in this analysis. A specific challenge with on-line assessment is that there is no single "norm" behavior to which the learners' actions can be compared. This presentation sets out to structure the challenges and potential solutions for the assessment of inquiry processes and outcomes.

Since inquiry learning is an educational approach, domain knowledge is the first most obvious concept addressed. Posttests measuring different types of knowledge (e.g. content, structural and conceptual knowledge) and transfer tests have been applied. There is nothing specific to inquiry learning about these types of tests. However, the concept of intuitive knowledge is primarily seen only in inquiry learning and tests have been developed for this (Swaak & de Jong, 1996). On-line representations of domain knowledge include learner-generated models, concept maps, or research reports that are produced while learning. Automatic assessment of these products that represent domain knowledge is now being developed (see e.g., Bravo, van Joolingen, & de Jong, 2009).

Another concept pertains to the assessment of specific inquiry skills. Again a division can be made with a measurement outside the learning environment and one in which on-line interactions are the basis for the assessment. Outside the learning environment (e.g., as a post-test) inquiry skills have been measured with the use of paper- and- pencil tasks. The concept of critical thinking skills shares many characteristics with inquiry, e.g. the Watson-Glaser Critical Thinking Appraisal® test includes scales that call upon data interpretation and drawing conclusions. The Test of Science Processing (Tannenbaum, 1971) and the Test of Integrated Science Processes (Padilla, Okey, & Dillashaw, 1983) were developed to assess science skills (e.g., variable identification, hypothesis formation, operationalization, experimentation and data interpretation). Another way of assessing inquiry skills is using a task that includes all aspect of inquiry, but is domain-independent, thereby controlling for the effect of prior knowledge. Evidence on the validity of this method, however is still lacking.

Other concepts that have been related to inquiry learning are assessments of epistemological beliefs (Kuhn, Cheney, & Weinstock, 2000) or tests that call upon knowledge about the workings of science (Nature of Science, see Chen, 2006). Several motivational concepts, such as attitudes and self-efficacy towards science have been measured using questionnaires.

Computer technology enables extensive logging of actions performed in digital learning environments and data mining techniques are currently used to extract patterns indicative of specific learning behaviours. Inquiry skills are often induced from the inquiry cycle. These skills pertain to the formulation of hypotheses, systematic experimentation (e.g., usage of the CVS) and data interpretation, although other labels have been used. Various other skills, for example metacognitive skills also have been object of research. In fact, inquiry learning relies heavily on regulative processes. Learning process data may include specific activities of learners (e.g., values assigned to input variables), chatlogs of collaborating learners, and even neurobiological measures (van Leeuwen, van der Meij, & de Jong, submitted).

The characteristics of the different assessment and measurement techniques are as manifold as the concepts measured. They involve criterion measures (e.g., a model score calculated on the basis of the actual model in the task, and descriptive measures (e.g., measures indicative of transformative or regulative processes), individual and group measures (e.g., questionnaires measuring epistemological beliefs), and process data collected unobtrusively or explicitly (e.g., with prompts). Assessment is performed by teachers, researchers, sometimes peers and sometimes automatically.

The goals of assessment include grading, but also informing learners (online support) and creating a basis for pedagogical interventions. In both cases, the system may present hints to learners to optimize learning (Veermans, de Jong, & van Joolingen, 2000). System based assessment of online activities may even be focused on collaborative activities. For example, monitoring online communication using chat may provide for real time information of the contributions of the different collaborators to the learning process or provide hints about what is best to communicate about (Anjewierden, Chen., Wichmann, & van Borkulo, submitted). Of course many validity issues have to be solved using these system-based assessment techniques for these purposes, but progress is being made towards automatic online support in inquiry learning environments.

Many concepts, measurement and assessment techniques are applied with regard to inquiry. The open-ended and self-directed nature of inquiry makes it hard to define hard criteria for grading and aptitude in inquiry, but both the assessment of learner behaviour and learning outcomes are indicative of emerging understandings and "good" inquiry skills. System-based assessment of inquiry learning to support learning is promising, but also hindered by the fact that there are various effective inquiry paths, which raises several validity issues (e.g., vary several variables at a time may be unwise in general, but functional in the orientation phase). The current presentation will give a structured overview of issues involved illustrated with examples from running projects (e.g., the SCY project) that show what problems are encountered and how solutions to these problems are implemented.

Participatory Assessment: Supporting Engagement, Understanding, and Achievement in Scientific Inquiry

Daniel T. Hickey, Michael K. Filsecker, Eun Ju Kwon. Indiana University, Bloomington, IN 47404, USA,
Email dthickey@indiana.edu; mfilseck@indiana.edu; ejkwon@indiana.edu

Tensions over educational assessment and measurement are central to ongoing debates about inquiry-oriented science education. This presentation sheds new light on this issue by (1) reviewing widely-appreciated tensions over assessment of inquiry-oriented vs. more expository science instruction, (2) revisiting these tensions using newer situative views of measurement and assessment, (3) introducing a participatory assessment model that addresses these tensions, and (4) showing how this model was used to foster communal engagement, individual understanding, and aggregated achievement in three design studies of leading technology-based inquiry curricula.

The debate over assessing inquiry reflects the conflict between different views of what it means to "know" and therefore what counts as authentic evidence of that knowledge. Hickey & Zuiker (2005) examined how *associationist*, *rationalist*, and *situative* views of cognition support different assumptions about knowledge of inquiry and what those assumptions mean for evidence. The associationist perspective characterizes knowledge as numerous specific associations regarding behavior (i.e., stimulus-response) and/or cognition (e.g., if-then). Hence, they support more direct instructional methods that efficiently teach those associations, and then use conventional recognition/recall tests to reliably measure how much individuals have learned. Antithetically, the rationalist perspective on cognition characterizes knowledge as a smaller number of higher-order conceptual schemas that vary from one person to the next. This supports constructivist inquiry-oriented instruction and the use of more open-ended problem solving and performance-oriented assessments of learning. These assessments are more subjective and less reliable than conventional tests; to some this makes them less "scientific" as well. Schwartz, Lindgren, & Lewis (2009) argue that constructivist pedagogies are often evaluated through non-constructivist means. They point out measures of "efficiency at remembering, executing skills, and solving similar problems" are "*something of a mismatch to larger constructivist goals"* (p. 35). Our presentation will examine this tension, and summarize advances in constructivist assessments (e.g., Schwartz and Bransford, 1998).

We then explore these tensions using newer situative perspectives on cognition. In their examination of the broader debate over constructivism, Gresalfi and Lester (2009) suggest that "learning is about more than a change in memory but about a change in ability to interact with resources in the environment" (p. 265). As Greeno and Gresalfi (2008) pointed out, this assumption casts doubt on the validity of the entire enterprise of assessing and measuring individual proficiency. Such highly contextualized (i.e., "situated") characterizations of proficiency ultimately require more interpretive evidence which can account for the broader technological and social context of knowledgeable activity. But these methods do not yield the evidence of individual proficiency that other stakeholders demand. We will examine this issue and summarize the burgeoning literature on situative assessment (e.g., Gee, 2003; Moss et al., 2005) and discursive approaches to assessment and formative feedback (Hickey & Anderson, 2007).

The presentation will conclude by describing a comprehensive approach to assessment that addresses these tensions. Participatory assessment uses design-based refinement of informal discursive assessment and feedback to align inquiry curricula to constructivist assessments of individual understanding; once sufficiently large gains in understanding are obtained, achievement gains are formally measured using external achievement