On the Process and Outcomes of Inquiry Learning: Changing Approaches to Assessment

Shaaron Ainsworth, Learning Sciences Research Institute, University of Nottingham, Nottingham, NG7 2RD, UK
Email: Shaaron.Ainsworth@nottingham.ac.uk (Organizer)
Ton de Jong, Faculty of Behavioral Sciences, University of Twente, 7500 AE Enschede
Email: A.J.M.deJong@utwente.nl (Organizer)
Cindy Hmelo-Silver, Graduate School of Education, Rutgers University, New Brunswick, NJ 08901-1183.
Email cindy.hmelo-silver@gse.rutgers.edu (Discussant)

Abstract: Inquiry learning is an educational approach that involves a process of exploration, asking questions and making discoveries in the search for new understandings. Researchers however are divided about the value of the approach. In the symposium, it is argued that one of the reasons for this controversy is the way that inquiry learning is assessed. Consequently, we aim to present papers which reflect on the challenge of assessing inquiry learning by describing the prevailing approaches to assessment and how technological and theoretical advancement is changing these approaches. The aim is not just to describe these approaches but reflect upon the opportunities that are created and difficulties that must be overcome as we pursue the goal of assessing the processes and outcomes of inquiry learning.

Introduction

Inquiry-based learning involves learners asking questions about the natural or material world, collecting data to answer those questions, making discoveries and testing those discoveries rigorously (e.g., de Jong, 2006). It is an idea with a long history (Dewey, 1916; Bruner, 1961) and many researchers and educators argue for the benefits of an inquiry approach to science (e.g. Dunbar, 2000; Duschl, 2008; Linn 2006).Yet, this approach remains controversial with debates still ranging about its effectiveness (e.g., Klahr & Nigram 2004; Kirschner, Sweller, & Clark, 2002; Hmelo-Silver, Duncan, & Chinn, 2007). One complicating factor in attempting to resolve this debate is that researchers are divided about how best to assess inquiry learning: should we focus on the process on inquiry or its outcome and if outcomes, what is it that inquiry learning help students develop? Consequently researchers working in inquiry learning have produced in depth analysis of the processes of inquiry learning: they have shown how it develops over time, how different learners' participate in inquiry learning and the ways that teachers or technology can scaffold inquiry learning (e.g., Kuhn, & Pease, 2008; Quintana et al 2004; de Jong & van Joolingen, 1998). The outcomes of inquiry learning have been seen in terms of domain knowledge at different levels and in different modes, inquiry skills, nature of science and scientists; attitudes to science and science self-efficacy (e.g., Hickey et al 2003, Linn & Hsui, 2000; Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002; Fraser, 1981; Ketelhut; 2007; Linn 2006). With such a variety of approaches and their implicit value systems, it is perhaps not surprising that this inquiry debate still rages.

The papers presented in this symposium reflect on the challenges of assessing inquiry learning. The paper by de Jong and Wilhelm aims to provide a sucient overview of the range of methods and concepts that researchers have used to assess inquiry learning. It summarizes the traditional approaches and points forwards to how new technological approaches avilable to researchers are increasing the sophistication by which we can measure the process of inquiry learning. As a complement, Hickey, Filsecker and Kwon focus on how theoretical advances in our understanding of inquiry learning change our approach to assessment. As cognition is seen as increasingly situated, the nature of the evidence required to understand learning by inquiry changes. This provides difficulties for researchers then asked to show an individual's proficiency in inquiry learning. Hickey et al describe an approach to inquiry assessment, participatory assessment, which is designed to address this challenge. Ainsworth et al describe the problem of engaging learners' with assessment. They argue that traditional outcome tests of inquiry learning can under-represent learners' understanding of the inquiry process by requiring completion of pen and paper tests that do little to motivate learners. They present an approach – Inquiry Comics – which presents a narrative of a character's investigation of a meaningful question and ask students to respond to the character's decisions. Finally, Clarke-Midura, Mayrath, and Dede also tackle the problem of student engagement with assessment but with a decidedly more high tech solution – that of immersive virtual environments. They reflect on the opportunities that such an approach brings but also helpfully share the problems they face in developing an innovative form of assessment.

The issue of how we understand the processes and outcomes of inquiry learning is one that has no simple answer. The purpose of bring together the papers in this symposium is to reflect upon whether inquiry 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 tests. Results from three inquiry learning projects that employed this model will be summarized, including *GenScope* (Hickey et al., 2003; 2006), three NASA multimedia curricula (Taasoobshirazi et al., 2006; Anderson et al., 2007), and the Taiga ecological science curriculum in the *Quest Atlantis* videogame (Barab, et al., 2007; Hickey, et al., 2009). For all but one of the NASA curricula, gains in understanding and achievement were statistically significant and equal to or larger than the gains in comparison classrooms using expository curricula to teach the same content.

Engaging students with assessment: Inquiry cartoons.

Shaaron Ainsworth, Stamatina Anastopoulou; Mike Sharples, Charles Crook & Claire O'Malley, Learning Sciences Research Institute, University of Nottingham.
Email: Shaaron.Ainsworth, Stamatina.Anastopoulou; Mike.Sharples; Charles.Crook; Claire.O'Malley @nottingham.ac.uk

Personal Inquiry (PI) project aims to enable learners to explore questions to which they genuinely want to know the answer, carry out investigations that relate to their own needs and concerns, and analyse and interpret findings (e.g. Anastopoulou et al 2009; Scanlon et al, 2009). These inquiries are designed to link the school classroom and the children's community (such as homes, parks, leisure facilities). Learners are supported by teachers and by the PI Toolkit which runs on netbook computers with connected data probes and is designed to scaffold the process of inquiry learning through scripts (dynamic lesson plans). These are lofty aims, and of course, we face the challenge of evaluating the extent to which we achieve them. Consequently, we have evaluated many aspects of the processes and outcomes of inquiry science. The paper reflects upon one of these: the development of inquiry skills. The central challenge we faced was to design an inquiry process test that was: a) informative to researchers and teachers; b) a valuable learning experience in its own right, not 'just a test' and; c) one that engaged students – for personal learning we needed personal assessment.

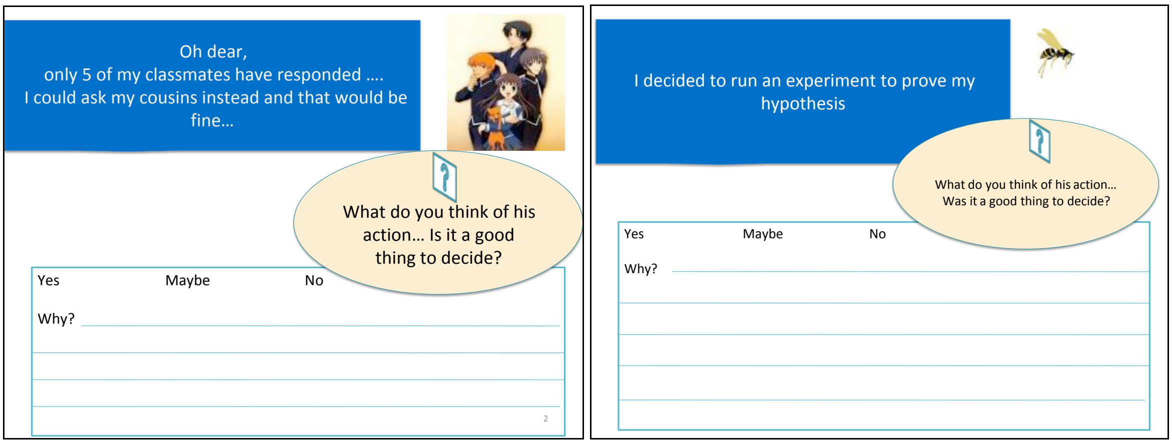

Figure 1. Two sample comic page (a) addresses what students understand about choosing appropriate samples; b) explores aspects of their understanding about hypothesis testing. Note these examples are taken from the middle of extended narratives.

The solution we have trialed is Inquiry Comics. These assessments are related to Concept Cartoons™ (e.g. Keogh & Naylor, 1999) which are cartoons of children discussing different (correct and incorrect) ideas about scientific phenomena in everyday settings. Concept Cartoons have a number of uses including promoting argumentation, stimulating children's own investigations, and formative assessments (e.g. Keogh & Naylor, 1999; Chin & Teou, 2009). Inquiry Comics share with Concept Cartoons™ the principles of minimal text, visual representation, familiar everyday settings, and correct and incorrect statements of scientific ideas. However, we use a comic form (a series of connected events presented in sequence to form a narrative) to present a character's inquiry process from the early stages of choosing a topic to investigate, through to selecting appropriate methods, collecting data, presenting it and drawing appropriate conclusions. At each stage in the process, the character makes decisions about aspects of scientific investigations that are known to be difficult for learners, such as judging veracity of source information, data collection, controlling variables, hypothesis testing, and drawing appropriate inferences from data (e.g. de Jong 2006; Kuhn, Pease, & Wirkala, 2009; Schauble, Glaser, Duschl, Schulze, & John, 1995). Sometimes those decisions are appropriate and sometimes less so (see Figure 1 for a decision considered to be less appropriate).

There are a number of challenges that we have faced in creating Inquiry Comics. It is, of course, difficult to create informative situations with minimal text. We also need to trade off the benefits of a full inquiry cycle and the resulting potential to probe understanding of the inquiry process at all stages with the length of the comic book which would result from such a complete activity. We needed dialogue that was easy to read, lacked jargon and that kept learners' attention. This was a particular necessity for the current project as we used Inquiry Comics as summative assessment so learners read them silently and did not discuss their meaning with peers or teachers (unlike the typical use of Concept Cartoons™). The comic form required a consistent narrative so any decision made by the character needed to continue through the comic. This obviously posed little problem for appropriate decisions but when the character was portrayed as making a decision that would threaten the veracity of the whole investigation (such a mischoosing a sample), this needed to be resolved within the comic. On these occasions, after the learners had been invited to write their interpretation of the investigator character's choice, a teacher or friend character would intervene to nudge the investigation back on course. Again this needed to be done with a light touch and in minimal text. Finally, given our requirements to use Inquiry Comics as pre and post-tests, we needed to create parallel version of each comics. Consequently, we needed similar topics to investigate and which contained the same methods. It is not clear in our first attempts with this technique the extent to which we have achieved this objective.

However, there were a large number of benefits to using Inquiry Comics. Students' performance on the tests revealed a depth of understanding that was not visible through other assessments (e.g., we had previously trialed asking students to design their own investigations). They were motivated and engaged with the comics, again far more so that with our other written tests. Analysis of their responses has revealed when learners are responding superficially (e.g. '*not a fair test'*) rather than more deeply (why something might not be a fair-test). This might be difficult to achieve in a multiple choice style of assessment. Their post-test responses also shed light on problems they had faced during their own investigations (e.g. their sensitivity to the problems of accurate on-going data collection).

The use of Inquiry Comics seems a promising addition to the battery of approaches to assess inquiry learning. We created comic books but they could also be adapted and used within other forms of Technology Enhanced Learning. They can be used summativly as we have done but also could be used for formative assessment and, of course, like Concept Cartoons™ for many other teaching and learning processes.

Measuring Inquiry: New Methods, Promises & Challenges

Jody Clarke-Midura, Michael Mayrath, Chris Dede, Harvard University, Cambridge, MA, USA,
Email: jody_clarke@mail.harvard.edu, mayratmi@gse.harvard.edu, chris_dede@harvard.edu

Contemporary views of science education regard scientific inquiry and the ability to reason scientifically as the essential core of science education (American Association for the Advancement of Science (AAAS), 1993; Chinn & Malhotra, 2002; NRC, 1996; Krajcik et al, 1998; Songer et al, 2003). According to White and colleagues, scientific inquiry is an active process comprised of four primary components: theorizing, questioning and hypothesizing, investigating, analyzing and synthesizing (White & Frederiksen, 1998; White, Frederiksen & Collins, in preparation). Measuring these inquiry processes as well as the products that result from the processes has long been a challenge for educators and researchers (Marx et al, 2004); however, advances in technology and measurement are creating new possibilities for assessing both process and product (Pellegrino, Chudowsky & Glaser, 2001; Behrens, 2009). There are three themes that this symposium is addressing: what inquiry is and is not, the best way to teach inquiry, and the best way to measure inquiry. We have chosen a widely accepted definition of what inquiry is by White et al, described above, and are focusing our work on the latter, how to best measure inquiry.

One such possibility for measuring inquiry comes in the form of immersive virtual assessments (IVAs). IVAs are three dimensional (3-D) environments, either single or multi-user, where digitized participants engage in virtual activities and experiences. Each participant takes on the identity of an avatar, a virtual persona that can move around the 3-D environment. IVAs allow us to create and measure authentic, situated performances that are characteristic of how students learn inquiry (NRC, 2000). These immersive environments have advanced capabilities for student experimentation and data analysis in a virtual setting, such as working with large data sets, GIS map visualizations, and simulated models of phenomena unobservable to the naked eye. Further, these environments enable the automated and invisible collection of very rich and detailed event-logs on individual learners in real-time, during the very act of learning (Pellegrino et al, 2001). Such event-logs provide time-stamped records of the details of students' actions while they interact with the IVA. Our prior work on using immersive technologies for learning environments lead us to believe that assessments delivered via this technology will be more motivating and engaging (Clarke, 2006; Clarke & Dede, 2009). We hypothesize that some students will be less likely to "freeze up" when taking the assessment and will try harder on these assessments than on paper-and-pencil and multiple choice tests.

In order to measure inquiry process and products in situ, we are using the Evidence Centered Design framework (Mislevy, Steinberg, & Almond, 2003) to develop IVAs for measuring inquiry at the middle school level, grades 6-8. These assessments are intended to be part of a standardized component of an accountability program. Our work on developing assessments for measuring inquiry is guided by the knowledge, skills, and abilities that underlie White and colleagues four components of scientific inquiry, e.g., theorizing, questioning and hypothesizing, investigating, analyzing and synthesizing. We are designing tasks that allow us to observe students gathering appropriate data, interpreting data, drawing conclusions, and providing evidence. The design of our tasks allows us to capture evidence of student learning outcomes and processes that contribute to an ongoing student model of proficiency in inquiry.

While IVAs are promising in their potential for studying student performances and learning trajectories, they also come with a cost. The field does not have a common model or algorithm for modeling the complexity of student learning and behaviors (Behrens, 2009). In our design process, we have had to work through the following challenges and issues such as: *how do we ensure task dependency without disrupting the flow of the inquiry process? What model best fits our data for proficiency? How do we score process separately from content? How do we measure a learning progression?* These and other issues will be addressed in more depth in the presentation.

If we succeed in addressing the challenges and issues on how to measure inquiry, we believe we will aid in the discussion of what inquiry is and also cast light on how to best teach inquiry. Only through deep understanding of how to measure and model inquiry will we better understand the best methods for teaching it (i.e. direct instruction vs problem-based learning debate).

References

American Association for the Advancement of Science. (1993). Benchmarks for science literacy: A Project 2061 report. New York: Oxford University Press.

Anastopoulou, S., Sharples, M., Wright, M., Martin, H., Ainsworth, S., Benford, S., Crook, C., Greenhalgh, C., & O'Malley, C. (2008) Learning 21st Century Science in Context with Mobile Technologies. In J. Traxler, B. Riordan & C. Dennett (eds.), Proceedings of the mLearn 2008 Conference: The bridge from text to context, Wolverhampton, UK: University of Wolverhampton, pp. 12-19.

Anderson, K. T., Zuiker, S., & Hickey. (2007). Classroom discourse as a tool to enhance formative assessment and practise in science. *International Journal of Science Education*, 1721-1744.

Barab, S., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E. J., et al. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education*, 91(5), 750-782.

Behrens, J.T. (2009). Response to Assessment of Student Learning in Science Simulations and Games. A NAS-commissioned response paper.

Bravo, C., van Joolingen, W. R., & de Jong, T. (2009). Using co-lab to build system dynamics models: Students' actions and on-line tutorial advice. *Computers & Education, 53*, 243-251.

Chen, S. (2006). Development of an instrument to assess views on nature of science and attitudes toward teaching science. *Science Education, 90*, 803-819.

Chinn, C. A., & Malhotra, B. A. (2001). Epistemologically authentic scientific reasoning. In K. Crowley, C. D. Schunn, & T. Okada (Eds.), Designing for science: Implications from everyday, classroom, and professional settings, (pp. 351-392). Mahwah, NJ: Erlbaum.

Chin, C., & Teou, L. Y. (2009). Using Concept Cartoons in Formative Assessment: Scaffolding students' argumentation. *International Journal of Science Education*, 31, 1307-1332.

Clarke, J. (2006). Making Learning Meaningful: An exploratory study of using Multi-User Virtual Environments for teaching inquiry in middle school. *Qualifying Paper presented to the Committee of Degrees at Harvard Graduate School of Education*, Cambridge, MA.

Clarke, J., & Dede, C. (2009). Design for Scalability: A Case Study of the River City Curriculum. *Journal of Science Education and Technology*, 18(4), 353-365.

de Jong, T. (2006). Technological advances in inquiry learning. Science, 312(5773), 532-533.

Dewey, J. D. (1916). Democracy in education. New York: Macmillan

Duffy, T. D., & Jonassen, D. H. (1991). Constructivism: New implications for instructional technology? Educational Technology, 31(5), 7–12.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. Review of research in education, 32, 268–291.

Fraser, B. (1981). TOSRA: Test of science related attitudes. *Australian Council for Educational Research*.

Gee, J. P. (2003). Opportunity to learn: a language-based perspective on assessment. Assessment in Education: Principles, Policy & Practice, 10(1), 27-46.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In Berliner, D. C. & Calfee, R. C. (Eds.), Handbook of educational psychology (pp. 15-46). New York: Macmillan Library Reference USA, Prentice Hall International.

Greeno, J. G., & Gresalfi, M. S. (2008). Opportunities to learn in practice and identity. In Moss, P, Pullin, D. C, Gee, J. P, Haertel, E. H., & Young, L. J (Eds.), Assessment, equity, and opportunity to learn (pp. 170–199). Cambridge, MA: Cambridge University Press.

Gresalfi, M. S., & Lester, F. (2009). What's worth knowing in mathematics? In S. Tobias & T. D. Duffy (Eds.), *Constructivist instruction: success or failure?* Routledge.
Hickey, D. T., & Anderson, K. T. (2007). Situative approaches to student assessment: Contextualizing evidence to support practice. P. Moss, Ed.) *Yearbook of the National Society for the Study of Education: Evidence and Decision Making*, *106*(1), 264–287.
Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology*, *18*(2), 187-208.
Hickey, D. T., Kindfield, A. C. H., Horwitz, P., & Christie, M. A. T. (2003). Integrating curriculum, instruction, assessment, and evaluation in a technology-supported genetics learning environment. *American Educational Research Journal*, *40*(2), 495.
Hickey, D. T., & Zuiker, S. J. (2005). Engaged participation: A sociocultural model of motivation with implications for educational assessment. *Educational Assessment*, *10*(3), 277-305.
Hickey, D. T., Zuiker, S. J., Taasoobshirazi, G., Schafer, N. J., & Michael, M. A. (2006). Balancing varied assessment functions to attain systemic validity: Three is the magic number. *Studies in Educational Evaluation*, *32*(3), 180-201.
Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107.
Keogh, B., & Naylor, S. (1999). Concept cartoons, teaching and learning in science: an evaluation. *International Journal of Science Education, 21*(4), 431-446.
Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *Journal of Science Education and Technology, 16*, 99-111.
Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.
Krajcik, J.S., Blumenfeld, P., Marx, R.W., Bass, K.M., Fredricks, J., & Soloway, E. (1998). Middle school students' initial attempts at inquiry in project-based science classrooms. Journal of the Learning Sciences. 7(3&4), 313-350.
Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, *15*, 309-328.
Kuhn, D., & Pease, M. (2008). What Needs to Develop in the Development of Inquiry Skills? *Cognition and Instruction, 26*(4), 512-559.
Kuhn, D., Pease, M., & Wirkala, C. (2009). Coordinating the effects of multiple variables: A skill fundamental to scientific thinking. *Journal of Experimental Child Psychology, 103*, 268-284.
Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching, 39*(6).
Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. *The Cambridge Handbook of the Learning Sciences*, 243-264.
Linn, M. C., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*: Lawrence Erlbaum.
Marx, R.W., Blumenfeld, P.C., Krajcik, J.S., Fishman, B., Solloway, E., Geier, R., & Tal, R.T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. Journal of Research in Science Teaching, 41, 1063–1080.
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3–62.
Moss, P. A., Pullin, D., Gee, J. P., & Haertel, E. H. (2005). The idea of testing: Psychometric and sociocultural perspectives. *Measurement: Interdisciplinary Research and Perspectives*, *3*(2), 63-83.
National Research Council. (1996). National Science Education Standards: observe, interact, change, learn. Washington, D.C.: National Academy Press.
National Research Council. (2000). How People Learn: Brain, Mind, Experience, and School. Washington, DC: National Academies Press.
National Science Foundation (2000). An introduction to inquiry *Foundations. Inquiry: Thoughts, views and strategies for the k-5 classroom.* (Vol. 2, pp. 1-5).
Padilla, M. J., Okey, J. R., & Dillashaw, F. G. (1983). The relationship between science process skill and formal thinking abilities. *Journal of Research in Science Teaching, 20*, 239-246.
Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know: The science and design of educational assessment. Washington, DC: National Academy Press.
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Scaffolding: A Special Issue of the Journal of the Learning Sciences, 13*(3), 337-386.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536.

Scanlon, E. Littleton, K., Gaved, M., Kerawalla, L., Mulholland, P., Collins, T., Conole, G., Jones, A., Clough, G., Blake, C. and Twiner, A. (2009) Support for evidence-based inquiry learning: teachers, tools and phases of inquiry. In Proceedings of the 13th Biennial Conference of the European Association for Research on Learning and Instruction (EARLI), Aug 25-29 2009, Amsterdam.

Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences, 4*(2), 131-166.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–522.

Schwartz, D. L., Lindgren, R., & Lewis, S. (2009). Constructivism in an age of non-constructivist assessments. In S. Tobias & T. D. Duffy (Eds.), *Constructivist Instruction: Success or Failure?* Routledge.

Songer, N. B., Lee, H. -S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. Science Education, 87, 490-516.

Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: The development of the what-if test. *Studies in Educational Evaluation, 22*, 341-362.

Taasoobshirazi, G., Zuiker, S. J., Anderson, K. T., & Hickey, D. T. (2006). Enhancing inquiry, understanding, and achievement in an astronomy multimedia learning environment. *Journal of Science Education and Technology*, *15*(5), 383-395.

Tannenbaum, R. (1971). The development of the test of science processes. *Journal of Research in Science Teaching, 8*, 123-136.

Tobias, S., & Duffy, T. D. (Eds.). (2009). *Constructivist theory applied to instruction: Success Or failure?* Routledge.

van Leeuwen, T., van der Meij, J., & de Jong, T. (submitted). Event-related potentials as a window on external representations.

Veermans, K. H., de Jong, T., & van Joolingen, W. R. (2000). Promoting self directed learning in simulation based discovery learning environments through intelligent support. *Interactive Learning Environments, 8*, 229-255.

White, B., & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. Cognition and Instruction, 16(1), 3-118.

White, B., Collins, A., Frederiksen, J. (in preparation). The Nature of Scientific Meta-Knowledge.

Wiggins, G., & McTigue, J. (2000). *Understanding by Design, standards-based instruction and assessment*. Association for Supervision and Curriculum Development, Alexandria, VA.