Title:
The relationship between students’ inquiry skills for experimenting and their skills at sense making in science microworlds.

Authors and Affiliations:

Janice Gobert
Worcester Polytechnic Institute

Juelaila Raziuddin
Worcester Polytechnic Institute

Kenneth Koedinger
Carnegie Mellon University
Abstract Body

Limit 4 pages single-spaced.

Background / Context:
Description of prior research and its intellectual context.

In reform policies, science inquiry is described as a set of process skills (NSES, 1996). Here we categorize these as skills including processes by which students design and conduct experiments (i.e., generate hypotheses and collect data), and engage in sense-making of data collected (i.e., interpreting data, warranting claims, and communicating findings). Students’ difficulties at many inquiry skills have been well-documented.

Regarding hypothesizing, they avoid stating hypotheses that could be rejected (Van Joolingen & de Jong, 1993; Klayman & Ha, 1987; Klahr, Fay & Dunbar, 1993), and show a bias toward investigating features believed to be causal as opposed to non-causal (Kuhn, Schauble, Garcia-Mila, 1992).

Regarding designing and conducting experiments, they have been shown to run experiments that try to achieve an outcome (i.e., make something burn as quickly as possible) or design experiments that are enjoyable to execute or watch (White, 1993), as opposed to testing a hypothesis (Schauble, Klopfer & Raghavan, 1991a; Njoo & de Jong, 1993a).

Regarding interpreting data, engage in confirmation bias, i.e., they won’t discard a hypothesis based on negative results (Klayman & Ha, 1987; Klahr & Dunbar, 1988). They change ideas about causality many times (Kuhn, Schauble & Garcia-Mila, 1992), don’t relate outcomes of experiments to theories being tested (Schunn & Anderson, 1999), and reject theories without disconfirming evidence (Klahr & Dunbar, 1988). They have difficulty linking data back to hypotheses (Chinn & Brewer, 1993; Klahr & Dunbar, 1988; Kuhn et al, 1995).

Regarding warranting claims, students can be overly reliant on theoretical arguments/explanations as opposed to evidence (Kuhn, Katz & Dean, 2004; Ahn et al., 1995; Brem & Rips, 2000; Schunn & Anderson, 1999); they struggle to provide appropriate evidence for their claims (McNeill & Krajcik, 2007), and they analyze data so as to protect prior beliefs, which can lead to faulty causal attribution (Kuhn et al., 1995; Keselman, 2003).

Lastly, regarding communicating findings, students have difficulty articulating and defending claims (Sadler, 2004). They tend to focus on what they did as opposed to what they found out, do not link data and conclusions, and do not relate results to their own knowledge/questions (Krajcik, et al., 1998). They also struggle to provide reasoning to describe why evidence supports claims (McNeill & Krajcik, 2007).

In prior work, we developed production rules (Koedinger, Suthers, & Forbus, 1999; Schunn & Anderson, 1999) to conduct automated performance assessment of some inquiry skills (Gobert & Koedinger, 2011). Production rules auto-scored when students made discoveries versus when they engaged in confirmation bias. Discovery refers to instances in which students, who originally made a scientifically inaccurate hypothesis, ‘discover’ a scientific phenomenon, as indicated by appropriate experimental trials and correct interpretation of these data. Confirmation bias students, on the other hand, were those who originally made a scientifically inaccurate hypothesis and held on to their false hypothesis, as represented in their data interpretation, even though their trials generated scientifically accurate data.
**Purpose / Objective / Research Question / Focus of Study:**
*Description of the focus of the research.*

We seek to more deeply understand the relationship between students’ experimentation during inquiry and their sense-making of these data with the goal of better understanding the processes that lead to both discovery as well as possible barriers that prevent it, which in turn, may lead to confirmation bias.

**Setting:**
*Description of the research location.*
Our data were collected in a rural town in Central Massachusetts.

**Population/Participants/Subjects:**
*Description of the participants in the study: who, how many, key features, or characteristics.*

145 eighth grade students, ranging in age from 12-14 years, participated in this study. Students belonged to one of six class sections and had one of two science teachers. Approximately 25% of the students are on free- or assisted-lunch and approximately 51% are “Below proficient” on the MCAS science test.

**Intervention / Program / Practice:**
*Description of the intervention, program, or practice, including details of administration and duration.*

Our learning environment Inq-ITS (www.inq-its.org; NSF-DRL# 0733286; NSF-DRL# 1008649; U.S. Dept of Ed.# R305A090170; R305A120778) assesses and scaffolds middle school students’ scientific inquiry skills, namely, hypothesizing, designing and conducting experiments, interpreting data, warranting claims with evidence, and communicating findings. Within this environment, a state change microworld was used to engage students in inquiry with the goal of assessing their inquiry skills.

**Research Design:**
*Description of the research design.*

All students were in the same condition for this study.

**Data Collection and Analysis:**
*Description of the methods for collecting and analyzing data.*

First, a pre-test for inquiry skills was administered. This was a short battery of multiple-choice items (n=12) was used to get a baseline measure of their inquiry skills including hypothesizing, independent and dependent variables, the control of variables strategy, and data interpretation. Secondly, a domain pre-test was administered. This was a short battery of multiple-choice items for content knowledge (n=7) was used to get a baseline measure content knowledge of this domain. Next, four Phase Change activities were used in our learning environment (described above). Our data for the present study is drawn from these tasks.

1) *Try to find out how the size of the container (trial 1), amount of substance (trial 2), level of heat (trial 3), and cover status (trial 4) affects each of the dependent variables: the melting*
point of the ice, the time it takes the ice to completely melt, the boiling point of the water, and, the time it takes for the water to completely boil.

2) After each trial, students were asked to generate an interpretation of their data and select trials that supported their claim.

3) The *communicate your findings task*: “Pretend you are explaining your conclusions about the effects of cover status on each of the dependent variables to a friend who did not do the experiments. Discuss how you conducted the experiments and how you came to your conclusions. Be as specific as possible.”

Data were collected and automatically displayed in a table by our system, Inq-ITS ([www.inq-its.org](http://www.inq-its.org); Gobert et al, 2012).

**Findings / Results:**

*Description of the main findings with specific details.*

Our model tracer, applied to students’ log data, coded: whether students’ initial hypotheses were scientifically accurate, whether their experimental trials were relevant to their hypotheses, whether their trials demonstrated controlled for variables strategy (Chen & Klahr, 1999), and whether their final interpretation was either supported or unsupported by their trials. Their open responses (4 tasks), reflecting communicating findings skills (NRC, 2011) were coded by hand to score the accuracy and level of details relevant to the task. These data were also used to examine students’ reasoning and to check whether students returned to their original, incorrect hypothesis (i.e., another demonstration of confirmation bias) or whether their discoveries were reflected in their explanations when they communicated their findings as a summative activity during inquiry.

Of the instances identified by our model tracer as *reflecting scientific discoveries* made by the students, 73% of the open responses also reflected this discovery. For example, students’ explanations on the task in which they made discoveries were scientifically accurate, detailed and thorough in describing variables, observations, and the effects of each level of independent variable on the dependent variable. Their scores for level of details and accuracy were high (max of 2) and consistently high thereafter, in subsequent tasks. An example of a scientific discovery open response was “I found that as the amount of the substance decreases, so does the time it takes for the ice cube to melt and the water to boil. I noticed that though it took less time to melt and boil, the temperatures at which the ice melted and the water boiled remained the same”. In the other 27% of instances, although the model tracer identified that the student had collected the appropriate data and had entered a “scientifically accurate” interpretation, their open response explanation did not reflect an accurate understanding of the data they had collected. For instance, the explanations reflected their observations during experimentation but did not explain the effects of the independent variable on the dependent variable. In general, these explanations were also shorter and incomplete and/or inaccurate. An example of such an open response is “it will change the conclusion depending on the grams of the substance”.

Of the instances that were identified by our model tracer as *reflecting confirmation biases*, 80% of these also reflected this confirmation bias in their explanation(s). Although the sample of students was too small for statistical analyses, these students, in general, scored lower on level of detail and understanding; for example, “The level of heat changes the boiling point by more heat melts the ice faster and less ice takes more time to melt ice”. The other 20% reflected a
scientifically accurate explanation of the phenomena during the last phase of inquiry, namely communicating findings, suggesting that they may have learned about the phenomenon upon reflecting about the results of their inquiry, even though their interpretation, entered into our interpretation widget in an earlier phase of inquiry, was scientifically inaccurate; as such, these instances may reflect “productive failure” (Kapur, 2009). An example is “in the experiment i found out that the size of the container does not affect the time it takes the ice to melt but the size of the ice does determine the time it takes to melt”. These cases are particularly interesting because they potentially represent instances of deep learning through reflection; and if so, prompts might be used to try to promote this meta-level of understanding.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

We are continuing to analyze these data with the goal of better understanding the processes that lead to both discovery as well as possible barriers that prevent it, which in turn, may lead to confirmation bias. In order to do this, we are examining students’ prior skills on our inquiry pre-test, their prior content knowledge on our content pre-test, as well as their trajectory of inquiry performance across the four state change tasks on the inquiry skills of interest. Specifically, an important issue we are addressing is whether there are barriers/drivers to sense-making that go beyond content knowledge and inquiry skills; this could be meta-cognitive or epistemological in nature. Taken together these data are important for understanding inquiry processes more deeply, specifically the relationship between designing and conducting experiments and sense-making processes, i.e., data interpretation, warranting claims, and communicating findings. These data are also important for performance assessment and real time scaffolding of inquiry, both of which are goals of the Inq-ITS project.
Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.


