J Sci Educ Technol
DOI 10.1007/s10956-016-9607-7

Drawing and Writing in Digital Science Notebooks: Sources of Formative Assessment Data

Angi Shelton[1] · Andrew Smith[2] · Eric Wiebe[3] · Courtney Behrle[3] · Ruth Sirkin[4] · James Lester[2]

Abstract Formative assessment strategies are used to direct instruction by establishing where learners' understanding is, how it is developing, informing teachers and students alike as to how they might get to their next set of goals of conceptual understanding. For the science classroom, one rich source of formative assessment data about scientific thinking and practice is in notebooks used during inquiry-oriented activities. In this study, the goal was to better understand how student knowledge was distributed between student drawings and writings about magnetism in notebooks, and how these findings might inform formative assessment strategies. Here, drawing and writing samples were extracted and evaluated from our digital science notebook, with embedded content and laboratories. Three drawings and five writing samples from 309 participants were analyzed using a common ten-dimensional rubric. Descriptive and inferential statistics revealed that fourth-grade student understanding of magnetism was distributed unevenly between writing and drawing. Case studies were then presented for two exemplar students. Based on the rubric we developed, students were able to articulate more of their knowledge through the drawing activities than through written word, but the combination of the two mediums provided the richest understanding of student conceptions and how they changed over the course of their investigations.

✉ Eric Wiebe
eric_wiebe@ncsu.edu

[1] Friday Institute, NC State University, Raleigh, NC, USA

[2] Computer Science, NC State University, Raleigh, NC, USA

[3] STEM Education, NC State University, Raleigh, NC, USA

[4] Psychology, NC State University, Raleigh, NC, USA

Keywords Science education · Elementary grades · Digital science notebooks · Drawing · Writing · Magnetism

Introduction

Assessment of student knowledge is central to the practice of teaching and learning in almost all educational settings. It provides a starting point for designing instruction based on what students already know, how their understanding is shaped and nurtured during instruction, and how successful a teacher has been at meeting learning objectives they have established (Bransford et al. 2000). In addition to more distal, summative assessments—usually in the form of end-of-unit tests—formative assessment is and should be woven into daily classroom instruction. Central to success of formative assessment techniques is both making accurate inferences about student knowledge based on information at hand and adjusting instruction appropriately (Bennett 2011). For that reason, teachers need appropriate, high-quality sources of information and analytic tools to rapidly and accurately ascertain student knowledge. Because formative assessment is woven into the fabric of the classroom, it should not be disruptive to learning goals by demanding too much time and resources from either the teacher's instructional activities or students' learning (Bell and Cowie 2001; Sadler 1989). In summary, a well-designed and implemented formative assessment strategy should be able to suggest how instruction is to be modified, as well as suggest impressionistically to the teacher what students know and can do (Bennett 2011), including how technology might facilitate learning goals (Falk 2012; Furtak 2012).

Broadly speaking, formative assessment strategies are used to direct the instructional processes by establishing where learners are going (e.g., by sharing learning expectations), where they are (e.g., through questioning), and how to get them where they should be (e.g., through scaffolding; Wiliam and Thompson 2008). Decisions to engage in assessment depend on a number of factors, including the current learning goal, curricular time, instructional resources at hand, and the state of the students (Harlen 2006). In addition, there are meta-factors, such as the likelihood of success and the prior experience of the teacher (Tomanek et al. 2008).

The rapid cyclical and ad hoc nature of formative assessment makes interpretation of data a tentative, evolving set of working hypotheses about students. Analytic tools need to go beyond simple multiple-choice questions on science content knowledge, especially given the rich content and practice that is possible in new computer-based science inquiry environments such as virtual chemistry laboratories and automated machine scoring of student work (Donnelly et al. 2014; Nehm et al. 2012; Sil et al. 2012). As Tempelaar et al. (2014) note, the temporal characteristic of formative data drives the types of data harvested, how it is analyzed, and how the teacher utilizes it. In addition, the tentative nature of formative data usually means accumulation and triangulation of evidence over time. Therefore, tools capable of integrating student activity over time, over activities, and across students are necessary. Finally, analytic tools must not only help identify teachable moments, but also support the teacher's formulation of the instructional intervention.

Science Notebooks as an Analytic Source

For the science classroom, one rich source of formative assessment data about student scientific thinking and practice is artifacts of student work in science notebooks during inquiry-oriented activities in the classroom (Aschbacher and Alonzo 2006; Ruiz-Primo et al. 2004). Well-structured classroom inquiry can generate formative data in the form of written argumentation (Berland and Hammer 2012) as students both make claims about the scientific principles underlying witnessed phenomena and provide evidence to support those claims (Sampson and Clark 2008). More specifically, claims are the conclusions students develop following theoretical instruction or laboratory exercises, while the evidence refers to any information, empirical data or accepted content knowledge, that functions as support of a claim (Berland and Reiser 2008).

Science notebooks have been used extensively in elementary and secondary grades as a mechanism to promote and reveal reflective thought through both drawing and written words (Campbell and Fulton 2003; CAPSI 2006; Martinez et al. 2012). For elementary student notebooks, these formative data come from both written text and drawings since understanding and the outcome of practice is distributed across both of these sources (Minogue et al. 2011). Not surprisingly, this research demonstrates that younger students are typically much better at illustrating their understanding through drawing than providing the same detail with the written word. Because both drawing shape and writing shape reveal underlying student mental models (Alesandrini 1981Schnotz and Bannert 2003), these learning goals situate students as being both a consumer and producer of text and graphics. Studies of graphics as a tool in modeling scientific phenomena (Acher et al. 2007; Schwarz et al. 2009) have relied on analysis of student graphic artifacts to reveal the process by which students developed understanding of scientific concepts and phenomena. While literature cited here has supported both the value of writing and drawing, individually, as instructional strategies for learning and sources of formative assessment, less is known as to how these sources work together to provide diagnostic data on student learning for teachers to utilize (Minogue et al. 2011; Shelton et al. 2015).

As computing technologies become more mainstream in all grade levels in science education, electronic science notebooks (ESN) will likely begin to supplant paper-based versions. ESNs have the potential of capturing student work in real time for use by analytic tools supporting formative assessment (Minogue et al. 2010; Lester et al. 2010). While the use of data to inform instruction is as old as the foundations of our educational system, the emergence of virtualized learning environments for students (Honey and Hilton 2011), combined with the strides made in business intelligence and similar analytic tools, has led to a new field of learning analytics as an intense focus of research and development in education (Siemens and Long 2011). Though initially focused on higher education learning environments supported by learning management system technologies, applications of similar digital platforms for K-12 education have also become more prevalent (Fulantelli et al. 2014). The first wave of learning analytic applications emerged primarily in the form of dashboards, with a more summative, distal orientation between student and instructor, and focused on providing insight into student progress toward more higher-level learning or performance outcomes (Tempelaar et al. 2014; Timms 2014). Less research and development has been focused on teacher-centric formative tools utilizing more proximal interactions and activities in settings such as classrooms (Sharples 2013).

Formative assessment is a powerful tool for guiding instruction, and science notebooks are a rich source of data for the science classroom. ESNs, in turn, have great potential for facilitating the harvesting and analysis of these formative data. However, little is known about how the written arguments and drawings created by students in an ESN reveal information about student knowledge, especially at the elementary grades where drawings are a particularly important source. Our goal is to investigate what conceptual knowledge about specific science topics is revealed in both writing and drawing, individually and collectively. Specifically, we ask:

RQ1: What level of representational conceptual understanding do students use in drawings and writing?
RQ2: How does conceptual understanding relate between drawings and writings?
RQ3: What is the relationship between the conceptual understanding documented in drawing and writing and the traditional summative knowledge test?

The goal is to use this information to not only move forward our understandings about formative assessment in elementary science education, but also to inform the development of technology tools, utilizing learning analytic methodologies built into digital science notebooks, and to assist teachers in their formative assessment activities.

Methods

We draw our data from a research project called the Leonardo Project, an intelligent cyber-learning system designed to support elementary students' scientific inquiry through investigation and reflection within a digital science notebook called the CyberPad (Lester et al. 2010). This application supports learning scientific ideas (e.g., electricity and magnetism) by providing embedded content and images along with a virtual pedagogical agent to help scaffold the experience. Students compose written arguments and illustrate their understanding of concepts within the digital science notebook, with complementary prompts for the writing and drawing to facilitate deeper learning. Drawing prompts are meant to connect the observed phenomena with more abstract conceptual representations (e.g., magnetic particles), while writing prompts are designed to prompt students to articulate why phenomena are happening, rather than simply reporting observations. Moreover, the writing was designed to set up the drawing and laboratory exercises and help link drawing with the concepts. Both the writing and drawing were designed as tools for self-explanation both to further learning and provide an accessible artifact for formative assessment of student knowledge (Ainsworth and Loizou 2003).

Materials

There are six instructional magnetism topics in the CyberPad, aligned with the FOSS™ kit-based curricular activities (FOSS Project 2013). However, these activities were modified to include new content coupled with writing and drawing activities that encouraged students to explore magnetism through a microscopic, particulate nature of matter lens (Harlow 2010; Minogue et al. 2010; Wiser and Smith 2008). The CyberPad topics included: magnetic attraction, magnetic properties, induced magnetism, magnetic interaction, magnetic fields, and force of attraction over a distance. Each instructional topic included multiple prompts, in both illustrative and written form, designed to elicit student understanding by encouraging students to create arguments (i.e., claims and supporting evidence) and images surrounding their predictions and observations of scientific phenomena and ideas. Topics 3 (induced magnetism) and 5 (magnetic fields) are the focus of this analysis.

Participants

In the 2013–2014 school year, 48 teachers instructing 69 classrooms of students implemented the six-topic magnetism unit with their classrooms. All teachers completed a professional development session to insure implementation fidelity. Of the 69 implementing classrooms, 20 representing 14 teachers were chosen as a subset because of the high degree of commonality of implementation of the study protocols (i.e., completed assessments, had a large percentage of consenting students, took between 10 and 14 days to implement the software, had only minor technological glitches). The majority of teachers were from sites within North Carolina and California, while four teachers with five classes were from sites in the Southeastern and Midwestern parts of the country. All 309 students analyzed in this study completed the six-topic inquiry-based instructional experience that included a combination of embedded curriculum, physical and virtual experiments, and reflections based on aforementioned laboratories.

Data Sources

Though participating students individually composed textual arguments in response to seventeen written prompts, only five were evaluated in depth for this paper (3 from Topic 3; 2 from Topic 5). Moreover, participants created eleven drawings, with three (2 from T3; 1 from T5) being the focus for this paper (see Table 1). This set of writing and drawing artifacts were chosen for their close, integrated set of concepts distributed between writing and drawing, and general representativeness to the other notebook activities. These writing and drawing prompts specifically allowed for paired modes of expression on the same set of conceptual understandings.

Table 1 Question name and prompt direction for students

Question name	Constructed response prompt
Focus Question 3 (FQ3)/ return to FQ3 (R2FQ3)	Notebook: What happens to the particles when an object is turned into a temporary magnet?
Drawing 3_1	Leo: Which type of particle can you find in each object? If a particle has poles, which way are the poles oriented?
Drawing 3_2	Notebook: Follow the same instructions as in *Draw It 1*. This time, place both the straw and paperclip close to the magnet
	Will this change anything about the particles you place in the magnifying circles?
	Leo: Which type of particle goes with each object? Which way do the particles with poles point?
Compare (occurs between FQ3 and R2FQ3)	Notebook: Look at the two drawings you made, I've placed them side-by-side for you. Use your own words to explain to me what your drawings show
	Leo: How do your drawings compare? What can you tell me about the two drawings?
FQ5/return to FQ5	Notebook: Can magnets work through materials like paper, cardboard, and metal foil to make magnetic objects feel a push or pull?
Drawing 5	Draw your observations. Don't forget to place the paperclip, three magnifiers, and at least three particles for each magnifying area onto the drawing space!

In Topic 3, a large donut magnet appears in their drawing space and they have access to the following graphic elements to annotate the drawing with one or more of the following elements: an arrow, a straw, a paperclip, a magnetic particle, a non-magnetic particle, and a magnifying glass (Fig. 1). Students are asked to show the particles when a paperclip and straw are both close and far away from the magnet. In Topic 5, students have to show what the particles look like when a paperclip is near a magnet but cardboard is placed in between the two objects (Fig. 2). In all drawings, students must place the specified objects onto the screen. They must also choose whether one can observe the particles (e.g., magnetic or non-magnetic) with or without the magnifier. The magnifier was a symbolic, semiotic tool that allowed students to indicate that the elements inside the magnifier circle represented very small (i.e., microscopic) elements not visible to the unaided human eye. As such, particles within the magnification circle represented atomic-scale elements. In addition to the constructed writing and drawing responses, students took a 20-question multiple-choice posttest to measure content comprehension covering all six topics, representing a traditional summative unit-based knowledge test. Prior field trials over 2 years were used to develop the test items using curricular review by content experts and reliability testing and refinement with over 500 students, resulting in a valid test aligned to the CyberPad content with acceptable internal reliability ($\alpha = .77$).

Procedures

A rubric was designed to evaluate student responses in both written and graphic form. Central to growth in scientific understanding is the ability to link observed scientific phenomena to underlying conceptual mechanisms (Schwarz et al. 2009). These mechanisms are usually invisible to the unaided eye and abstract in nature; however, they can be represented through drawn or written conceptual elements. For these investigations, the central overarching concept was the particulate nature of matter (Wiser and Smith 2008). The rubric was designed to identify and link (through space and time) the visible elements of the phenomena to the underlying conceptual mechanisms. Conceptual understanding of magnetism is spatial in nature, requiring appropriate depiction in drawing or words of the relative location and orientation of elements relative to each other (Sederberg and Bryan 2009). Four of the scores were about the usage of core "actors" from the magnetism investigation: paperclips, straws, magnifiers, and particles. These actors are linked to key language used in the CyberPad to describe the scientific concept and associated phenomena and to the graphic objects that the students work with (Figs. 1, 2). Scoring is based on how words and graphic objects are arranged to convey strong conceptual meaning. Three dimensions were related to the accurate depiction of the particulate nature of permanent magnets, objects that could be magnetized (e.g., paperclips), and non-magnetic objects (e.g., cardboard). The semiotic dimension indicated whether the nature of the drawings and written arguments was evaluated as iconic (i.e., only using words and images to represent concrete ideas) or symbolic, (i.e., using words or drawn elements to denote abstract concepts and moving beyond physical representations) or somewhere in between (Groisman et al. 1991; Han and Roth 2005; Lemke 1998). This score was placed on an ordinal scale to represent that the ideal answers would contain more abstract rather than concrete (macro scale) descriptions of the scientific phenomena (see

Fig. 1 Drawing space for Topic 3

Fig. 2 Drawing space for Topic 5

Table 2 Example scores for Student 867

Rubric category	Writing score	Drawing 3_2 score
Paperclip placement	0	1
Straw placement	0	1
Magnifier placement	0	1
Particles	1	1
Magnet particles	1	2
Paperclip particles	1	2
Straw particles	0	2
Static/dynamic	1	–
Iconic/symbolic	3	3
Quality of science	3	–

Table 2 for rubric criteria). The final two dimensions, the scientific and dynamic nature of the response, were assessed only for the written portion. The dynamic dimension scored whether students referenced a change over time ($0 =$ no mention; $1 =$ mention of some change; $2 =$ mention of change over time). Finally, a general score for scientific accuracy was assessed from 0 (no answer) to 4 (ideal answer).

Two raters coded the graphic and textual student artifacts in response to specific prompts in Leonardo. Inter-rater reliability was calculated via Cohen's kappa (κ) and a protocol for drawing and writing coding using a 3-classroom training set before coding the entire corpus. Coders initially coded a portion of the training set and discussed differences in order to refine the coding process and ambiguities in the rubrics. Coders then independently coded drawings for each question from three training classrooms and achieved an acceptable level of agreement ($\kappa = .88$) before coding the remainder of the corpus. The procedure was then repeated for the writing prompts, achieving an acceptable level of agreement ($\kappa = .76$), after which the remainder of the corpus was coded.

All written and drawn responses from the 309 consenting students were coded using the rubric. Analysis was performed on a per topic basis and only excluded students from the respective Topic (3 or 5) when they were missing any target prompt for that unit (see Table 1). The curricular design purposefully created pairings of drawing and writing activities around focused conceptual ideas. Since the goal was to compare drawn and written expression of these concepts, it was important to only use responses for which students provided both. However, in order to maximize the overall sample, it was not required that a student has a complete set of responses for both Topics to be included in the analysis. This allowed for a total $n = 87$ for Topic 3 (28 %) and $n = 133$ (43 %) for Topic 5. A lower count for Topic 3 was likely because it required an additional written response and drawing, increasing the probability of a missed response.

Table 2 provides for an example scoring for the responses based on the return to Focus Question 3 and Drawing 3_2 prompts. Higher scores would be indicative of richer, deeper descriptions of all of the key “actors” in the phenomena, their states at the macro- and microlevels, and relationship to the key scientific concepts. In this example, Student 867 wrote for Focus Question 3: “At the begining [sic] when the “iron” was not close to the magnet the particles were all wibbly [sic] and wobbly and then when i [sic] moved the “iron” closer to the magnet the poles started to align.” This student did not earn points for paperclip, straw, the magnifier or non-magnetic particles, because none of them were mentioned. Though they did not specifically mention the paperclip particles being magnetic, they did reference their iron content and how the particles began to align, earning them one point. This response is correct and discusses the content at an abstract level.

For Drawing 3_2, Student 867 earned almost all the points possible. Only one point was missed because they did not show non-magnetic particles in the paperclip (i.e., paperclips contain both magnetic and non-magnetic material). The student placed the objects close to the magnet, per the instructions, and included multiple, properly oriented particles inside the magnifiers (see Fig. 3). The static and quality of science dimensions were not evaluated in the drawings.

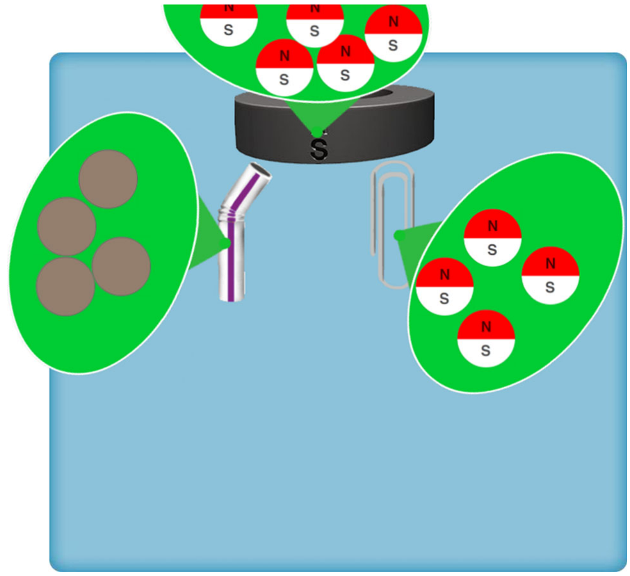


Fig. 3 Student 867, Drawing 3_2

Data Analysis

Analysis of the three research questions was addressed in two parts. First, quantitative analysis was conducted to answer each of the three questions. This analysis included both the results and brief discussion of their implications. Next, case studies of two representative students are presented through thick description to examine holistically each of their writing and drawing artifacts (Yin 2009). This mixed methods approach provides for both quantitative trends emerging from the student population as a whole and an qualitative unpacking of a representative sample to better understand the meaning of these quantitative results (Creswell et al. 2003; Tashakkori and Teddlie 2003).

The statistics were organized based on the research questions. Specifically, we looked at the mean values for drawings, written arguments, and the level of the representation in semiotic terms (i.e., from iconic to symbolic). To determine how student conceptions associate across drawings and writing, Pearson’s correlations were run between the continuous data drawing and writing scores per topic area. In order to examine the utility of using drawing and/or writing to reveal conceptual understanding, a multiple regression was run. The regression used the score of the summative posttest questions as the outcome variable and the following as predictor variables: summed rubric score for all writings, summed rubric score for all drawings, summed rubric score for the scientific accuracy of all writings samples, and semiotic scores for drawings and writings. Case studies of two representative students’ drawings and writings that exemplified the majority of participants in terms of modal rubric scores were then analyzed in order to provide complementary analysis into the range of student responses (Yin 2009). That is, these cases highlight how student understanding is revealed in drawing and writing. Of particular interest in the case study will be to look at students who represented differing levels of understanding about magnetism, as reflected in their written and drawn artifacts used in formative assessment.

Results and Discussion

RQ1: What Level of Representational Conceptual Understanding Do Students Use in Drawing and Writing?

Rubric score descriptive statistics, along with percentage scores based on means, are shown in Table 3. In all categories except the compare question, students scored approximately twice as high on their drawings as their writings for the same topic area, which is especially telling since the drawings had less possible points. Even though the students had completed the laboratories and drawing activities before creating an answer to the return to focus question, the written answers for return to FQ3 and FQ5 had lower mean scores. Two of the three drawings had some of the highest (most abstract) semiotic mean scores, although the highest mean was for students' attempt at Focus Question 3. The compare written prompt for FQ3 and both the return to FQ3 and FQ5 had markedly low (concrete) semiotic scores even though these prompts were designed to be reflective responses after writing and drawing about the phenomena. In summary, drawings revealed higher overall scores compared to the writing, indicating deeper, richer depictions of the abstract actors and their role in the scientific phenomena. However, these depictions were not always more abstract in nature than the writing, even though drawing elements were provided for students to do so. The compare written prompt seemed to engender a fairly rich comparison; however, this was also at a fairly concrete level. Perhaps most surprising were the return to focus questions, where students had the opportunity to summarize their findings using scientific concepts, but they failed to provide rich, descriptive, or abstract ideas in writing. In addition to the differences seen between writing and drawing scores, collectively the scores for FQ3 trended higher than FQ5. This could reflect the fact that FQ3, earlier in the unit, provided more directive hard scaffolds integrated into the prompts. FQ5 both had fewer hard scaffolds provided and was considered a conceptually more difficult topic.

RQ2: How Does Conceptual Understanding Relate Between Drawings and Writings?

Given that the curricular materials were specifically designed to allow students to explore the same scientific concepts in two modalities, the association between student drawing and writing was assessed in part using a Pearson correlation analysis between drawing and writing rubric scores. For Topic 3 ($n = 87$), there were four significant positive correlations between the drawing and the writing (Table 4). Drawing 3_1 was significantly correlated with the initial writing prompt FQ3 ($r = .2175$, $p = 0.05$) and the compare writing prompt, where students were asked to compare their drawings ($r = .3101$, $p = .005$). For Drawing 3_2, there was also two significant positive correlations to writing prompts; the compare writing prompt ($r = .3547$, $p = .001$) and the return to focus question prompt ($r = .2391$, $p = .03$). For all significant correlations, increases in the drawing scores were associated with increases in the writing scores. These findings are reflective

Table 4 Correlations of writing and drawing samples for Topic 3 (T3)

	FQ3 writing	Compare writing	R2FQ3 writing
Drawing 3_1	.2175*	.3101**	.1889
Drawing 3_2	.1228	.3547**	.2391*

* $\leq$.05, ** $\leq$.01

Table 3 Means and standard deviations for overall prompt and semiotic scores

Writing/drawing prompt	N	Overall mean	SD	%	Semiotic mean	SD
FQ3	87	2.99	1.42	23	2.29	0.823
Drawing 3_1	87	6.74	3.46	52	1.43	0.497
Drawing 3_2	87	6.85	3.06	53	2.15	0.833
Compare	87	4.40	2.997	34	0.89	0.80
Return to FQ3	87	2.04	1.36	17	0.768	0.89
FQ5	133	1.57	1.22	13	1.15	0.54
Drawing FQ5	133	5.17	3.40	43	2.01	0.86
Return to FQ5	133	1.26	1.62	11	0.86	0.95

of a number of curricular strategies. Note that the compare writing prompt was highly significant with both drawings, which was appropriate given students were specifically asked to discuss their drawings and how they are similar and different. For FQ3, the prompts specifically asked students to both depict (in drawing) and discuss (in writing) the particulate-level phenomena; perhaps strengthening the connection between drawn and written responses.

Pearson's correlations were also calculated to evaluate the relationship between the writing and drawing variables for Topic 5 ($n = 133$). Drawing 5 was negatively correlated with the return to focus question writing prompt ($r = -0.207$, $p = .02$; Table 5), meaning increases in the drawing score were correlated with a decrease in the score of the return to focus question. The negative correlation depicts a strong divergence not seen in FQ3. As noted earlier, FQ5 was both conceptually more difficult and provided with fewer hard scaffolds supporting the drawing and writing around a unifying scientific concept. It seems that the drawing task, by virtue of its modality or implicit level of scaffolding, allowed students to better express their ideas compared to the writing task.

RQ3: What is the Relationship Between the Conceptual Understanding Documented in Drawing and Writing and the Traditional Summative Knowledge Test?

A multiple regression was used to explore the relationship between conceptual knowledge demonstrated on the constructed response drawing and writing items for the two examined topics and a multiple-choice posttest representing all of the topics of the unit (Table 6). More broadly, this analysis looked at how predictive the formative assessments of student artifacts were of the more distal summative posttest assessment. As designed, the posttest and formative drawing and writing activities all covered the same key conceptual ideas in magnetism. In addition, the posttest items utilized graphic elements utilized in the drawing activities and wording used in background scientific text and prompts. However, the posttest used the standard forced (multiple) choice format typical of summative tests. Predictors used for the multiple regression model paralleled the design of the rubric. There were predictors for the overall depiction/description of the scientific concepts for each modality: drawing score (i.e., a sum of Drawing 3_1, Drawing 3_2, and Drawing 5) and writing score; the level of abstraction in drawing and writing: drawing semiotic score and writing semiotic; and finally the level of scientific accuracy in writing: Writing Scientific Accuracy score. A single model was run with variable entry guided by prior analyses of descriptive statistics and correlations (Table 6). Collectively, all of these predictors explained 42 % of the variance ($R^2 = .4178$, $F(5,48) = 6.88$, $p < .0001$). However, the drawing score was the only single variable found to predict the overall posttest scores significantly ($b = .295$, $p = .005$). Writing score was the next highest contributor, but this was not significant. Thus, the conceptual content (i.e., sophistication) of a student drawing is predictive of a general conceptual understanding of magnetism as represented by the multiple-choice questions on the posttest (see Appendix 1 for example questions from the posttest). However, the student writing scores were less predictive of student conceptual understanding on the summative assessment. Scientific accuracy and the level of abstraction (semiotics scores), while useful locally in formative assessment of student work, were not predictive of posttest scores.

Table 5 Correlations of writing and drawing samples for the third topic area

	FQ5 writing	R2FQ5 writing
Drawing 5	−.015	−.207*

* ≤.05

Table 6 Multiple regression model for variables predicting posttest scores

Predictor	b	$CI_{95\ \%}$ for b		β	sr^2	SE
		Lower	Upper			
Drawing (D)	.295	.095	.496	.802**	.107	.10
Writing (W)	.155	−.034	.344	.212	.033	.09
Sci accuracy (W)	−.137	−.52	.245	−.136	.006	.19
Semiotics (W)	−.044	−.708	.619	−.026	.000	.33
Semiotics (D)	−.515	−1.353	.322	−.332	.018	.42

** $p < .01$

Case Studies

The quantitative analysis indicated that the level of documented conceptual knowledge (as indicated by the rubric score) was not evenly distributed between drawings and writing. In addition, within each of these two mediums, both the level of documented conceptual knowledge and degree of abstraction (i.e., semiotic score) varied from prompt to prompt. Correlational tests seem to indicate that the strongest relationship between drawing and writing scores happened when students were specifically asked to write about their drawings. Finally, drawing scores seemed to be most predictive of how students would perform on the summative posttest of multiple-choice questions. These findings can be further explored by an in-depth analysis of select student work. To illustrate a more thorough analysis of student work, two case studies are shown and discussed at a per question level. A review of student drawing and writing scores, and their corresponding semiotic scores showed that students broadly fell into three groups. The highest performing group of students demonstrated the ability to articulate the observed phenomena of scientific investigations and articulate the connections of these phenomena to underlying scientific conceptual ideas using an appropriate level of abstraction. The middle performing group were also able to articulate the observed phenomena, but struggled to link these to the underlying abstract concepts. The weakest performing group was not able to demonstrate an understanding of either the concrete observed phenomena or the underlying abstract concepts. The two cases describe typical students from the highest (Student #855) and middle (Student #217) groups, as they provided the richest comparative descriptions of how knowledge of magnetism distributed between drawn and written artifacts of student work. Each case will be presented referencing the questions from Table 1, in chronological appearance during the magnetism module.

Case Study for Student #217

Focus Question 3 (What happens to the particles when an object is turned into a temporary magnet?). The student responded: "When an object is turned into a temporary magnet, its particles sometimes dont [sic] work." The interpretation of this answer is that temporary magnets have particles that sometimes are not attracted to magnets. In the context of this module, this student understands the temporary nature of induced magnetism and that the particles themselves do not change, but it is unclear whether they understand that the reason the temporary magnets only sometimes work is the alignment of particles due to induced magnetism.

Drawing 3_1 (Fig. 4a) While the student followed the directions in that they placed all of the key actors—a straw, paperclip, magnifier, and particles—onto the drawing space, they did not understand the critical relationships of the elements. Having the particles in the magnifying bubble and the magnifier touching an object is central to representing graphically the abstract notion of visible objects consisting of many (different) invisible particles. The particles, in turn, need to be grouped and oriented in ways that represent the magnetic qualities of the objects. Essentially, this student is wielding all the objects in their drawing as discrete, concrete icons rather than leveraging their symbolic meaning.

Drawing 3_2 (Fig. 4b) In this drawing, the student again does not appropriately utilize the magnifier tool. The student indicates that the paperclip is magnetic and the straw is non-magnetic by putting the respective particles touching each object, but has not used the magnifier to imbue the symbolic nature of these particles. That is, the polar, magnetic particle seemed to be utilized as a simple label for the magnetized paperclip and the non-magnetic (gray) particle for the straw that cannot be magnetized. Missing was the use of the magnifier to indicate the microscopic, abstract nature of these particles, or multiple polar particles needed to represent the key concept of alignment. The association of particles to objects without utilizing the magnifier was a common theme in drawings for students in this group. A more subtle but important point made in the background material is that paperclips contain a mixture of particle types and should have some non-magnetic particles since they are made of steel. Finally, an arrow is included, but it is not clear whether it is indicating the attractive movement of the paperclip, the straw, or both. A more

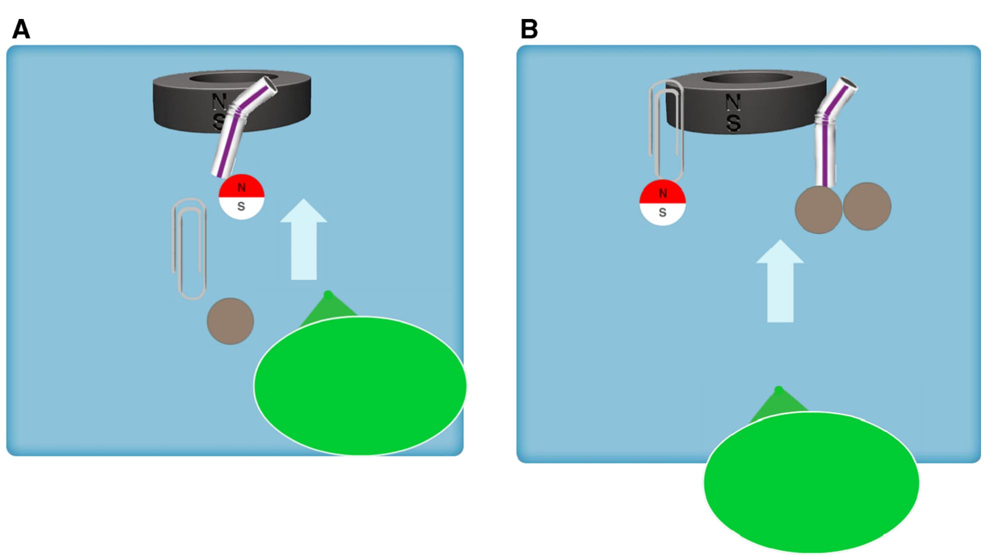


Fig. 4 **a** Case 217, Drawing 3_1. **b** Case 217, Drawing 3_2

accurate drawing would only have the paperclip attached with an arrow associated with it. However, this drawing is an improvement over their first drawing in terms of the amount of conceptual knowledge about magnetism being represented.

Compare Student 217 stated: “It show that the metal has magnetism and the straw doesn’t” This written response parallels the level of conceptual knowledge revealed in Drawing 3_2. However, it is missing many key ideas about the particulate nature of magnetism. The response effectively represents an observable, concrete understanding of the witnessed phenomena.

Return to FQ3 Student’s answer to the return to Focus Question 3: “When an object is turned into a temporary magnet, its particles sometimes dont [sic] work.” This student chose not to change their original FQ3 answer at all. If we had considered this answer without the context of any other learning, it would seem that the student did not advance at all. However, Drawing 3_2 and the compare prompt provide us with some insight into the student’s knowledge about materials, like steel and plastic, containing different kinds of particles, of which only some of which respond to magnets. In that regard, it represents movement forward in terms of wielding abstract terms such as “particles.”

Focus Question 5 (Can magnets work through materials like paper, cardboard and metal foil to make magnetic objects feel a push or pull?) Student 217 answered the fifth focus question (FQ5) saying, “No but i think it can work threw [sic] metal.” The student thought that magnets could only attract other objects through metallic objects. The student’s understanding of electricity and conductivity may have influenced this answer since in a circuit materials must be conductors in order to facilitate the movement of electrical energy. In contrast, magnets have a field of attraction, which does not require contact through contiguous metallic material.

Drawing 5 As with the drawings in Topic 3, Student 217 is still not able to utilize the abstraction capabilities of the magnifier (see Fig. 5). They have put a magnetic particle over the N/S etched into the magnet and close to the paperclip, perhaps indicating the magnetic qualities of these objects and/or the magnetic field. That is, they are using the magnetic particle icon to “label” the magnet and possibly the paperclip, much as they did in Drawing 3_2. This is representative more of the observable magnetic qualities of the concrete objects than the abstract ideas of magnetism. Also, the non-magnetic particle does not seem to be associated with any object; it is not touching either the cardboard or the paperclip, so it is impossible to determine what the student is trying to portray. Similarly, the magnifier was put in the drawing, but not utilized, paralleling the lack of understanding of this semiotic tool seen in the FQ3 drawings. Interestingly, the paperclip is being attracted through the cardboard, just like the student should have observed, but contrary to what the student initially articulated in the initial focus question prompt.

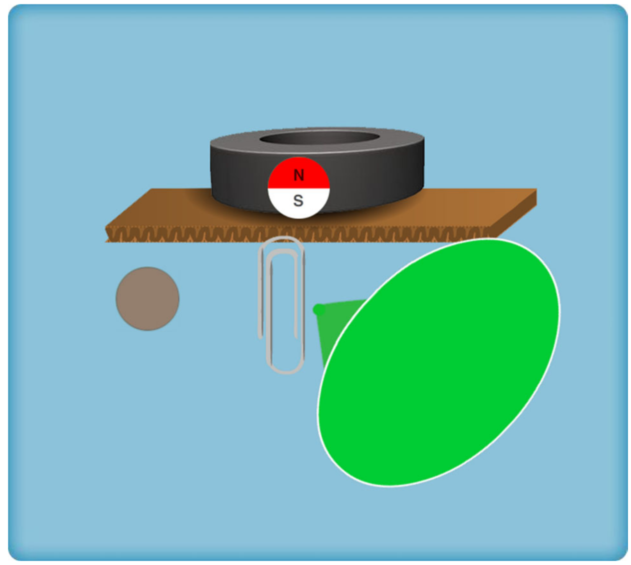


Fig. 5 Case 217 Drawing T5

Return to FQ5 Student 217 stated, “Yes it can work threw [sic] like paper.” In line with what was represented in the drawing, the student has now shifted their stance as to whether magnetic materials (the paperclip) can be attracted through non-magnetic materials (cardboard). What is lacking is evidence that they understand the underlying, abstract mechanism of magnetic fields. Neither this written answer nor the drawing demonstrates how particles can be used to represent the underlying abstract concept of magnetic fields working at a distance to induce magnetism in magnetic materials.

Case Study #855

Focus Question 3 Student 855 answered the third focus question by saying, “When an object is turned into a temporary magnet, its particles turn into temporary mini-magnets.” “Mini-magnets” is the term presented within the CyberPad’s content to refer to microscopic particles that have magnetic qualities. Thus, this answer provides a somewhat incomplete understanding of the textual content knowledge provided to the student; articulating the notion of the particulate nature of matter, but seeming to imply that perhaps particles change types. However, this is the expected level of understanding at the beginning of the topic.

Drawing 3_1 This drawing represents close to the ideal target drawing (as articulated by the curriculum developers) for this activity. The straw and paperclip are placed away

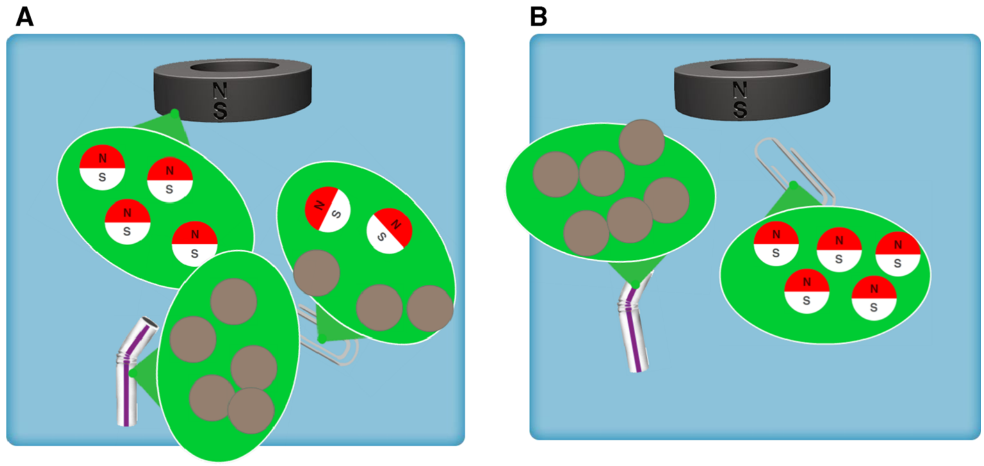


Fig. 6 **a** Case 855's Drawing 3_1. **b** Case 855's Drawing 3_2

from the magnet, as they were instructed to do (see Fig. 6a). In addition, it has a magnifier over each object with an adequate number of particles for each magnified area needed to represent possible particle alignment. This drawing correctly illustrates that the straw only has non-magnetic particles and that the magnet only has aligned, magnetic particles. The paperclip, in turn, has both magnetic and non-magnetic particles. Moreover, the magnetic particles in the paperclip are not aligned with the magnet, since the paperclip has not yet been magnetized. In summary, the student has used the semiotic tools of magnifier and particles to communicate an understanding of the particulate nature of mater and how different objects may be composed of different types of particles. Only in the next drawing will we know for sure whether the alignment (or lack thereof) of particles in the paperclip was purposeful.

Drawing 3_2 This drawing shows that the straw is still non-magnetic. However, the particles in the paperclip are now all oriented, indicating the magnetizing influence of the donut magnet (see Fig. 6b). Thus, the student is showing that as the paperclip gets closer to the magnet, it becomes a temporary magnet. However, unlike Student #855's first drawing, the paperclip no longer has non-magnetic particles. This potentially represents a misconception that the non-magnetic particles have transformed into magnetic particles during the process of magnetization. This conclusion is supported by the student's explanation of temporary magnetism in the initial focus question written response.

Compare When asked to compare their drawings, Student 855 said, "When the paperclip and the straw were away from the magnet the paperclip had some mini magnet particles and some normal particals [sic] and the straw had all normal particals. When they were close the paperclip had all mini magnets and the straw was still the same as when it was away from the magnet." In this response, the student not only accurately describes the surface level observed state of the objects in the investigation, but most importantly, describes the underlying abstract conceptual explanation of what was observed. Interestingly, the thoroughness of the written response seems to confirm that the student believes that the non-magnetic particles in a paperclip transform into magnetic particles when the paperclip moves closer to the permanent magnet.

Return to FQ3 After completing the entire topic, the student returns to FQ3 and answers the question again. Student 855 says, "When a paperclip is near a magnet the particals change into mini magnets." This answer is not very different from their initial answer, and without considering the images, it would appear there has been almost no growth in understanding. However, the drawings seem to indicate a strong understanding of the particulate nature of temporary magnets and non-magnetic materials. Yet, there is both drawn and written confirmation that a full understanding of material properties of temporary magnets is still lacking at some levels.

FQ5 Before completing the laboratories, reading the content, and constructing the drawings, Student 855 answered FQ5 by writing, "I think the magnet could pull the paperclip through the foil and the paper but not cardboard." This student is perhaps alluding to the fact that magnetic fields did not have a very far range and could not go through materials as thick as cardboard. This response to the initial focus question seems to put Student 855 farther along than student 217 in their understanding of magnetic fields.

Drawing 5 Following a physical laboratory, the student constructed this drawing, demonstrating that the magnet could, in fact, attract the paperclip through cardboard (see Fig. 7). Furthermore, the student shows that both the magnet and paperclip have aligned magnetic particles, while the cardboard has no magnetic particles. The student again has produced a drawing that is close to the ideal target drawing, utilizing the semiotic tools to depict the state of the particles

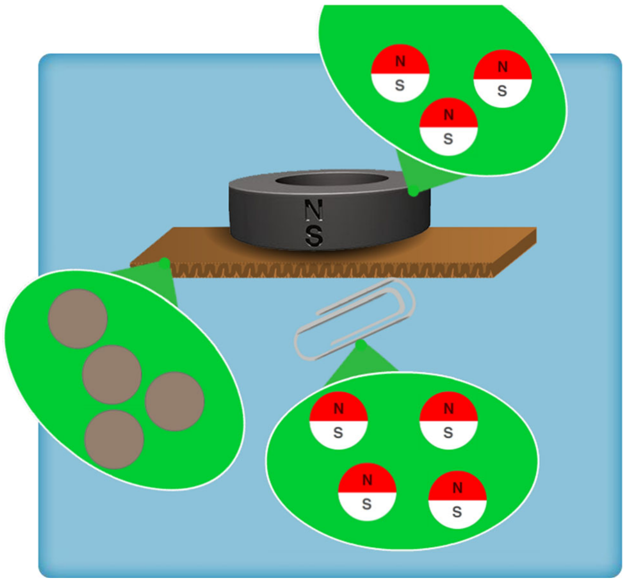


Fig. 7 Case 855's Drawing 5

for each of the materials in their correct relation both intra- and inter-object. From this drawing, it can be inferred that the student understands that the magnet's magnetic field must be what is attracting the paperclip because the cardboard is not magnetic, and that the magnetic field is able to work through the cardboard. However, the student seems to maintain the misconception that if an object is attracted to a magnet, any of its non-magnetic particles are now transformed into magnetic objects.

Return to FQ5 Following the drawing and laboratories, the student answered the return to Focus Question 5. This answer does not really stand alone without the context of the original response to the question, but it is understandable. Student 855 said, "Yes all of them do." That is, that cardboard, paper, and aluminum foil can all be placed in between a paperclip and magnet, and the attraction will still occur between the two objects.

Conclusion

Our analysis of student digital science notebook work revealed that fourth-grade student understanding of magnetism was distributed unevenly between writing and drawing, but the combination of the two mediums provided the richest understanding of student conceptions as they engaged in investigations concerning magnetism. Based on the rubric we developed, we were able to demonstrate that students were able to articulate more of their knowledge through the drawing activities than through written word. However, it was also clear that the design of hard scaffold prompts relative to the conceptual understanding being elicited was not equal—that is, students were clearly more successful on some writing and drawing prompts than others. This unevenness was also revealed in the correlations between drawing and writing around the same conceptual ideas, where the prompt specifically asking students to compare drawings showed the highest correlations in scores between drawing and writing. In contrast, Topic 5, where this more advanced concept was not scaffolded as well, students seemed to really struggle with providing strong written responses, showing a negative correlation between drawing and writing. The utility of these student artifacts as a tool for formative assessment was further strengthened by the results of the multiple regression model. While the drawing and writing scores were collectively a strong predictor of summative knowledge scores on the entire unit, it was the drawing scores which providing the largest predictive component.

The case studies of two students—one still working on developing a deeper conceptual understanding and the other more successful—provided powerful examples of how the drawing activities revealed student conceptual understanding. In particular were students' ability to engage in the deeper expression of the more abstract ideas around the particulate nature of matter and the mechanisms of magnetism. While student writing often provided confirmatory evidence toward the level of abstraction students were able to engage in, the drawings often provided more powerful evidence of the level of understanding of students. The regression analysis provides supporting evidence that, while collectively the drawing and writing artifacts provide a richer picture of student conceptual understanding, the drawings were a more substantive contributor.

A full explanation as to why drawings seemed to be a richer source of formative data is beyond the scope of this study; however, it is worth speculating along a couple of lines. First, this could be because students at this grade level simply have a harder time expressing themselves in the written form due to vocabulary acquisition or general writing strength. While students' writing was scaffolded with well-structured prompts and sentence starters, the level of written expression expected of students was still out of reach for many. On the other hand, is could be that the scaffolds provided for the drawing activity were simply more accessible to students. The drawing provided a more direct, homologous representation of what was observed in the physical investigation, while also providing rich semiotic tools for representing the abstract concepts underlying the observed phenomena. Though it might be a concern that the scaffolds provided by the drawing environment created an unearned shortcut to a higher rubric score, our case studies seem to support the counter argument that strong conceptual knowledge was still needed to guide drawing construction. That is, there was a rather precise visual grammar that needed to be utilized to thoroughly express the more abstract concepts.

The findings of this study show the near-term benefit of teachers using both drawn and written artifacts in notebooks as formative assessment data sources, as neither medium provides a full picture of student conceptual understanding (Minogue et al. 2011). These findings also have implications for future work on developing machine-based tools for parsing and analyzing student work. While more work has been done on analyzing student constructed written responses (e.g., Baker 2014; Gobert et al. 2013; Leeman-Munk et al. 2013), much less work has been done on student drawing recognition (e.g., Forbus et al. 2011). Our findings point to the importance of continued research focusing work on drawing recognition if such formative assessment tools are going to be utilized in the earlier grades such as the context we have investigated.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant No. DRL-1020229. We would also like to acknowledge the help and support of our cooperating classroom teachers.

Compliance with ethical standards

Conflicts of interest The authors have no known potential conflicts of interest pertaining to the research reported in this manuscript.

Research involving human participants and/or animals All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This research with human participants was conducted in this reported research and approved by our overseeing Institutional Review Board.

Informed consent Informed consent was obtained from all individual participants included in this study.

Appendix 1

Example posttest items

5. When a piece of iron is very close to a magnet

   a. Nothing happens to the particles in the steel.
   b. **All the magnetic particles orient the same way.**
   c. The magnetic particles orient in different ways.
   d. Some magnetic particles orient one way, and others orient the opposite way.

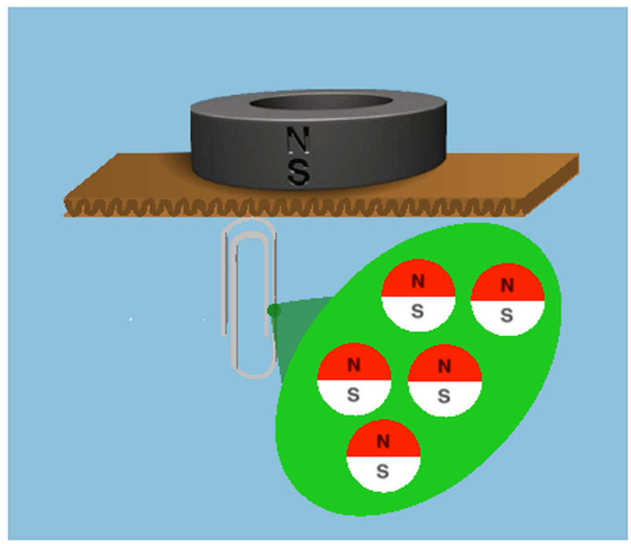
6. Which of the following statements best describe what materials are made of?

   a. All materials contain magnetic particles
   b. All materials contain only one kind of particle
   c. Some materials do not contain smaller particles
   d. **All materials are made of many, many small particles that cannot be seen with your eyes.**

7. Temporary magnets

   a. Have particles that cannot rotate
   b. Do not contain magnetic particles
   c. **Contain magnetic particles**
   d. Have particles that change from non-magnetic to magnetic

8.



Look at the picture. Choose the best description of what is happening in this image

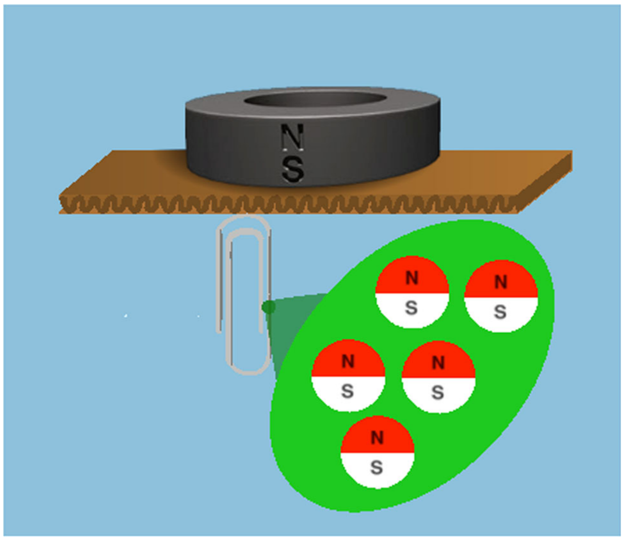
   A. The paperclip is going to fall because the cardboard is blocking the magnetic force
   B. The paperclip has not become magnetized because the cardboard is blocking the magnetic force
   c. The paperclip is being attracted to the magnet through the cardboard
   d. The cardboard has become magnetized

15. Non-magnetic particles

   a. **Do not orient in magnetic fields**
   b. Can turn into magnetic particles

c. Can be in materials that contain magnetic particles
d. Only exist in plastic

17.



Look at the picture. Choose the best description of what is happening in this image

a. **The magnetic field lines pass through the cardboard and magnetize the paperclip**
b. The magnetic field lines only go upwards away from the cardboard
c. The particles in the paperclip have not been affected by the magnetic field
d. The paperclip will not be attracted to the magnet

18. Magnetic field lines

a. Will not pass through non-magnetic material
b. **Pass through aluminum foil and paper**
c. Are best represented as a single circle around a magnet
d. Do not affect magnetic particles

References

Acher A, Arca M, Sanmarti N (2007) Modeling as a teaching learning process for understanding materials: a case study in primary education. Sci Educ 91:398–418
Ainsworth SE, Loizou AT (2003) The effects of self-explaining when learning with text or diagrams. Cogn Sci 27(4):669–681
Alesandrini KL (1981) Pictorial—verbal and analytic—holistic learning strategies in science learning. J Educ Psychol 73:358–368
Aschbacher P, Alonzo A (2006) Examining the utility of elementary science notebooks for formative assessment purposes. Educ Assess 11(3):179–203
Baker RS (2014) Educational data mining: an advance for intelligent systems in education. Intell Syst IEEE 29(3):78–82
Bell B, Cowie B (2001) The characteristics of formative assessment in science education. Sci Educ 85(5):536–553
Bennett RE (2011) Formative assessment: a critical review. Assess Educ: Princ Policy Pract 18(1):5–25. doi:10.1080/0969594X.2010.513678
Berland LK, Hammer DM (2012) Framing for scientific argumentation. J Res Sci Teach 49(1):68–94. doi:10.1002/tea.20446
Berland LK, Reiser BJ (2008) Making sense of argumentation and explanation. Sci Educ 93:26–55
Bransford J, Brown AL, Cocking RR (2000) How people learn: brain, mind, experience, and school, 2nd edn. National Academy Press, Washington
Campbell B, Fulton L (2003) Science notebooks: writing about inquiry. Heinemann, Portsmouth
CAPSI Research Group (2006) Elementary Science Notebooks: Impact on classroom practice and student achievement in science and literacy. http://www.capsi.caltech.edu/research/Projects.htm#Elementary. Accessed May 2008
Creswell JW, Clark Plano VL, Gutman ML, Hanson WE (2003) Advanced mixed methods research designs. In: Tashakkori A, Teddlie C (eds) Handbook of mixed methods in social and behavioral research. Sage, Thousand Oaks
Donnelly DF, Linn MC, Ludvigsen S (2014) Impacts and characteristics of computer-based science inquiry learning environments for precollege students. Rev Educ Res. doi:10.3102/0034654314546954
Falk A (2012) Teachers learning from professional development in elementary science: reciprocal relations between formative assessment and pedagogical content knowledge. Sci Educ 96(2):265–290. doi:10.1002/sce.20473
Forbus K, Usher J, Lovett A, Lockwood K, Wetzel J (2011) CogSketch: sketch understanding for cognitive science research and for education. Top Cogn Sci 3(4):648–666. doi:10.1111/j.1756-8765.2011.01149.x
FOSS Project (2013) Welcome to FOSSWeb. http://www.fossweb.com/. Accessed 15 Oct 2013
Fulantelli G, Taibi D, Arrigo M (2014) A framework to support educational decision making in mobile learning. Comput Hum Behav. doi:10.1016/j.chb.2014.05.045
Furtak EM (2012) Linking a learning progression for natural selection to teachers' enactment of formative assessment. J Res Sci Teach 49(9):1181–1210. doi:10.1002/tea.21054
Gobert JD, Sao Pedro M, Raziuddin J, Baker RS (2013) From log files to assessment metrics: measuring students' science inquiry skills using educational data mining. J Learn Sci 22(4):521–563
Groisman A, Shapiro B, Willinsky J (1991) The potential of semiotics to inform understanding of events in science education. Int J Sci Educ 13(3):217–226
Han JY, Roth W-M (2005) Chemical inscriptions in Korean textbooks: semiotics of macro- and microworld. Sci Educ 90(2):173–201. doi:10.1002/sce.20091
Harlen W (2006) Making formative assessment work: effective practice in the primary classroom. Educ Rev 58(1):116–118
Harlow DB (2010) Structures and improvisation for inquiry-based science instruction: a teacher's adaptation of a model of magnetism activity. Sci Educ 94(1):142–163
Honey MA, Hilton M (eds) (2011) Learning science: computer games, simulations, and education. Committee on Science Learning: Computer Games, Simulations, and Education. National Research Council, Washington
Leeman-Munk S, Wiebe EN, Lester J (2013) Mining student science argumentation text to inform an intelligent tutoring system. Paper presented at the AERA Annual Meeting, San Francisco

Lemke JL (1998) Multiplying meaning: visual and verbal semiotics in scientific text. In: Martin JR, Veel R (eds) Reading science: critical and functional perspectives of the discourses of science. Routledge, New York, pp 87–111

Lester J, Mott B, Wiebe EN, Carter M (2010) The Leonardo Project. http://projects.intellimedia.ncsu.edu/cyberpad/. Accessed Nov 2013

Martinez JF, Borko H, Stecher B, Luskin R, Kloser M (2012) Measuring classroom assessment practice using instructional artifacts: a validation study of the QAS notebook. Educ Assess 17(2–3):107–131

Minogue J, Wiebe E, Madden L, Bedward J, Carter M (2010) Graphically enhanced science notebooks. Sci Child 48(3):52–55

Minogue J, Bedward JC, Wiebe EN, Madden L, Carter M, King Z (2011) An exploration of upper elementary students' story-boarded conceptions of magnetism. Paper presented at the NARST Annual Meeting, Orlando

Nehm RH, Ha M, Mayfield E (2012) Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations. J Sci Educ Technol 21(1):183–196

Ruiz-Primo MA, Li M, Ayala C, Shavelson RJ (2004) Evaluating students' science notebooks as an assessment tool. Int J Sci Educ 26(12):1477–1506

Sadler DR (1989) Formative assessment and the design of instructional systems. Instr Sci 18:119–144

Sampson V, Clark DB (2008) Assessment of the ways students generate arguments in science education: current perspectives and recommendations for future directions. Sci Educ 92(3):447–472

Schnotz W, Bannert M (2003) Construction and interference in learning from multiple representations. Learn Instr 13(2):141–156

Schwarz CV, Reiser BJ, Davis EA, Kenyon L, Acher A, Fortus D, Krajcik J (2009) Developing a learning progression for scientific modeling: making scientific modeling accessible and meaningful for learners. J Res Sci Teach 46(6):632–654

Sederberg D, Bryan L (2009) Tracing a prospective learning progression for magnetism with implications at the nanoscale. Paper presented at the learning progressions in science (LeaPS) conference, Iowa City

Sharples M (2013) Shared orchestration within and beyond the classroom. Comput Educ 69:504–506. doi:10.1016/j.compedu.2013.04.014

Shelton A, Wiebe E, Smith PA (2015) What drawing and argumentative writing says about student understanding. Paper presented at the NARST Annual Meeting, Chicago

Siemens G, Long P (2011) Penetrating the fog: analytics in learning and education. Educ Rev 46(5):30–32

Sil A, Shelton A, Ketelhut DJ, Yates A (2012) Automatic grading of scientific inquiry. In Proceedings of the NAACL-HLT 7th workshop on innovative use of nlp for building educational applications (BEA-7), Montreal

Tashakkori A, Teddlie C (2003) Handbook of mixed methods in social and behavioral research. Sage, London

Tempelaar DT, Rienties B, Giesbers B (2014) In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Comput Hum Behav. doi:10.1016/j.chb.2014.05.038

Timms M (2014) Teacher dashboard design. WestEd, San Francisco

Tomanek D, Talanquer V, Novodvorsky I (2008) What do science teachers consider when selecting formative assessment tasks? J Res Sci Teach 45(10):1113–1130. doi:10.1002/tea.20247

Wiliam D, Thompson M (2008) Integrating assessment with learning: what will it take to make it work? In: Dwyer CA (ed) The future of assessment: shaping teaching and learning. Erlbaum, New York, pp 53–82

Wiser M, Smith CL (2008) Learning and teaching about matter in grades K-8: when should the atomic-molecular theory be introduced? In: Vosniadou S (ed) International handbook of research on conceptual change. Routledge, London, pp 205–239

Yin RK (2009) Case study research: design and methods. Sage, Los Angeles