Article

Guiding Students to Develop an Understanding of Scientific Inquiry: A Science Skills Approach to Instruction and Assessment

Elisa M. Stone

Berkeley Science & Math Initiative, University of California, Berkeley, Berkeley, CA 94720-3860

New approaches for teaching and assessing scientific inquiry and practices are essential for guiding students to make the informed decisions required of an increasingly complex and global society. The Science Skills approach described here guides students to develop an understanding of the experimental skills required to perform a scientific investigation. An individual teacher's investigation of the strategies and tools she designed to promote scientific inquiry in her classroom is outlined. This teacher-driven action research in the high school biology classroom presents a simple study design that allowed for reciprocal testing of two simultaneous treatments, one that aimed to guide students to use vocabulary to identify and describe different scientific practices they were using in their investigations—for example, hypothesizing, data analysis, or use of controls—and another that focused on scientific collaboration. A knowledge integration (KI) rubric was designed to measure how students integrated their ideas about the skills and practices necessary for scientific inquiry. KI scores revealed that student understanding of scientific inquiry increased significantly after receiving instruction and using assessment tools aimed at promoting development of specific inquiry skills. General strategies for doing classroom-based action research in a straightforward and practical way are discussed, as are implications for teaching and evaluating introductory life sciences courses at the undergraduate level.

INTRODUCTION

Instruction and assessment are often designed to teach and measure the science concepts students learn but are less likely to address the skills students must develop in order to answer meaningful scientific questions. To make informed decisions in modern society, students must routinely formulate questions, test ideas, collect and analyze data, support arguments with evidence, and collaborate with peers. To promote such skills, science educators have long recommended

frequent experimental work and hands-on activities (Dewey, 1916) and effective assessment methods for measuring critical scientific-thinking skills and evaluating performance in laboratory exercises (National Research Council [NRC], 2000). It is crucial that we strive to meet the call to improve K–12 and undergraduate inquiry instruction and assessment set out by a number of national science education organizations (NRC, 1996, 2003; American Association for the Advancement of Science [AAAS], 1998, 2009). Most recently, the Next Generation Science Standards (NGSS) provided guidelines for encouraging the practices that scientists and engineers engage in as they investigate and build models across the K–12 science curriculum and beyond (NGSS, 2013).

Many educators have claimed that inquiry is especially important in urban environments and for engaging minority students in making math and science relevant for them (Barnes *et al.*, 1989; Stigler and Heibert, 1999; Moses, 2001; Haberman, 2003; Tate *et al.*, 2008; Siritunga *et al.*, 2011). Inquiry-based instruction includes a variety of teaching strategies, such as questioning; focusing on language; and guiding students to make comparisons, analyze, synthesize,

and model. Skills important for scientific thinking are often taught implicitly; that is, the instructor assumes students learn how to think like a scientist by simply engaging in frequent experimental work in the classroom. However, explicit approaches have been shown to be more effective, for example, in teaching nature of science concepts to both students and science teachers (Abd-El-Khalick and Lederman, 2000; Lederman *et al.*, 2001). The classroom described in this action research study aims to create a learning environment that is explicit about these essential features of classroom inquiry.

An accumulation of evidence exists for how inquiry in the science classroom at both the undergraduate and K–12 levels is effective in promoting student understanding of various content areas in life sciences education. Examples include an increased understanding of a variety of key concepts in the life sciences (Aronson and Silviera, 2009; Lau and Robinson, 2009; Rissing and Cogan, 2009; Ribarič and Kordaš, 2011; Siritunga *et al.*, 2011; Treacy *et al.*, 2011; Zion *et al.*, 2011; Ryoo and Linn, 2012). Moreover, Derting and Ebert-May (2010) have shown long-term improvements in learning for students who experience learning-centered inquiry in introductory biology classes. While several studies have correlated such inquiry-based curricula on specific science topics with improvements in general academic skills (Lord and Orkwiszewski, 2006; Treacy *et al.*, 2011), more research is needed on how students develop and integrate their understanding of specific experimental and scientific inquiry skills, as well as what general strategies are effective for promoting and measuring this understanding.

The theoretical framework that lies at the foundation of this study is knowledge integration (KI). "The knowledge integration perspective . . . characterizes learners as developing a repertoire of ideas, adding new ideas from instruction, experience, or social interactions, sorting out these ideas in varied contexts, making connections among ideas at multiple levels of analysis, developing more and more nuanced criteria for evaluating ideas, and formulating an increasingly linked set of views about any phenomenon" (Linn, 2006, p. 243). KI lies at the heart of the curricular design for both concepts and skills taught in the classroom in this study, as well as the design of the research tools for the study itself.

A number of KI rubrics and scoring guides have been developed to measure the extent to which students connect ideas important for understanding key concepts in different content areas (Linn *et al.*, 2006; Liu *et al.*, 2008). For example, Ryoo and Linn (2012) designed a rubric that captures how middle school students integrate their ideas about how light energy is transformed into chemical energy during photosynthesis. In this study, the KI framework is applied to issues of how students integrate their ideas about skills important for scientific inquiry. In particular, this framework goes beyond considering inquiry as an accumulation of compartmentalized ideas. Rather than examining discrete steps in the process of experimentation, such as analyzing data or reaching conclusions (Casotti *et al.*, 2008), the KI construct described below allows for the integration of student ideas about different aspects of experimental work, such as how experimental design is connected to interpretation of data, accounting for the complex ways in which separate skills important for experimentation are interconnected.

The research question for this study was: How does the use of Science Skills instructional and assessment tools that encourage students to identify and explain the skills they are using in laboratory activities improve KI of student ideas about scientific inquiry and experimentation? A successful model for combining inquiry-based instruction with assessment tools for measuring student understanding of concepts related to scientific experimentation in a high school biology class is presented. While this study is set in a high school context, an argument is provided for how it could translate into introductory life sciences courses at the undergraduate level.

METHODS

School Site and Participants

There has been an emphasis in recent years on creating "small schools" within large comprehensive schools that provide a more personalized education for students; build relationships among students, teachers, and parents; give teachers additional opportunities to collaborate; and focus on specific themes, such as health or the arts (Feldman, 2010). Student participants were enrolled in a small school for visual and performing arts students within a large urban high school of more than 3000 students. The total enrollment for this small school was approximately 200 students distributed throughout grades 9 through 12. Arts, humanities, and science teachers collaborated to design an integrated science curriculum in which students learned cell biology, genetics, evolution, and ecology in a ninth-grade biology course, and applied this learning to an in-depth study of human anatomy and physiology, particularly those topics most relevant for visual and performing artists, in a 10th-grade human anatomy and physiology course. Participants in this study included all 10th-grade students for one academic year in two class periods, referred to here as groups 1 and 2.

Students in each of the two class periods for this research study represent a typical group of performing and visual arts high school students. As with any two class periods at most high schools, the two classes for this study differed from one another in some ways. De facto tracking existed in terms of students' interests in specific performing arts, as performing and visual arts classes were scheduled to alternate with the science classes. Students were given a choice about which arts classes they could take; students who preferred drama ended up in group 1, and those who preferred dance ended up in group 2. Additionally, because there were fewer male dancers than female dancers, group 2 was predominantly female (89% compared with 53% for group 1). Students who identified as visual artists were in both groups 1 and 2. It is not clear whether the differences between the two groups may have had any effect on academic performance in a life sciences class, as the overall academic performance as reflected in the grades awarded to assignments in this class was roughly similar between the two groups, falling between 75% and 82% each semester. Student demographic data for this small school were roughly similar to those of the total student body at the high school for the academic year studied (see Supplemental Table S1; Education Data Partnership, 2010).

Permission to do this research was sought and obtained through the local school district; parents received a letter describing the teacher's plans for the research and had the option to give consent for their student's work to be published.

Course Curricula

The students who participated in this research were in two class periods of 10th-grade human anatomy and physiology. The yearlong curriculum was organized by body system, as are many courses in human anatomy and physiology, with the systems brought together under different "big ideas," such as "The brain serves to control and organize all body functions" and "Structure determines function." While there was a strong emphasis on hands-on experience in the course, a variety of instructional approaches were used in these classrooms; these included lecture, group discussions, collaborative research projects, laboratory experiments, and inquiry-driven computer-based curricula. Throughout the course, the instructor made it clear to students that she particularly valued student-driven questions, experimentation, and the excitement of discovery. Key teaching strategies included: guiding students to provide a rationale for their predictions and hypotheses for experiments; leaving data organization and analysis open-ended; discussing as a class the pros and cons of experimental choices made by different student groups; combining class data to expand sample size; grading lab reports on the quality of evidence-based arguments rather than experimental outcomes; highlighting modeling whenever there was an opportunity; having students present findings directly to other classmates, including in scientific conference-style poster sessions; and providing structure for frequent class discussions and scientific discourse between peers in which there was an expectation of challenging and defending ideas. The textbook (Marieb, 2006) was supplemented with curricula designed, collected, and/or modified by the author (e.g., from Kalamuck *et al.*, 2000; National Institutes of Health [NIH], 2000; WISE, 2012; Tate *et al.*, 2008). Students in the course typically engaged in experimental work for approximately 40% of the instructional time each week, with a specific focus on different aspects of the scientific research process and introduction to associated specialized vocabulary on an ongoing basis. A key learning objective for this curriculum is for students to increase awareness of who they are as scientists and develop a more specific vocabulary for discussing experimentation and their strengths in science.

Science Skills Instruments

With the intent of making a number of scientific-thinking and problem-solving skills explicit for students, a list of "A to Z Science Skills" was developed (such as Analyzing, Building in controls, Graphing, Hypothesizing; Figure 1). Not only was this list posted on bulletin boards throughout the classroom, but every student also had quick access to a copy in his or her class binder. The instructor referred to the list whenever a term or method was introduced or required special emphasis in a class activity. It was made clear that this list was not all-inclusive; indeed, the class would often focus on a skill not on the list. Thus, this tool can be used to focus on different terms or parts of the scientific process, depending on the lab or activity for the day.

Students' experimental work was formatively assessed with a postlab reflection, which asked students to use the A to Z Science Skills list to "Pick 3 skills that you think you did well during lab, and describe how you demonstrated this skill." This simple assignment was given approximately one

Figure 1. A to Z Science Skills. Note that this list is, of course, not all-inclusive and can be modified to fit instruction for other disciplines and at other levels from elementary through graduate education. A to Z Science Skills was adapted from Math Alphabilities (P. Tucher, personal communication).

to two times per month to assess 1) how students understood the importance of these terms and methods in relation to their own work doing science investigations, and 2) how they viewed their own development in using such skills. Generally, students completed this self-assessment at the end of a lab exercise as they wrote up their conclusions or after a series of experiments to reflect on their work during the previous week or two, which had the added benefit of promoting metacognition for their science learning. Conclusions for lab work were structured and always followed a similar pattern; students were asked to report their findings, use their data as evidence to support their claims, discuss sources of error, and identify a next experiment that would extend their work. As students completed their conclusions and the postlab reflection, they were encouraged to talk to one another about skills they had used that week (which were frequently different from those employed by their peers), giving the instructor the opportunity to wander around the classroom, checking to make sure their self-assessment matched her own understanding of what they had accomplished.

Group Collaboration Instruments

One skill that is key to successful scientific research is collaboration, which is specifically named in the A to Z Science Skills list. Scientific collaboration was promoted in the classroom on a regular basis with the use of a Group Collaboration rubric (Supplemental Figure S1) and the corresponding reflection, designed to measure successful collaboration skills, such as sharing ideas, distributing work, using time efficiently, and decision making. The goal for using the Group Collaboration tools was to improve awareness of what it takes to collaborate successfully and to help students learn to value group work more as an effective way to accomplish a major project. The Group Collaboration rubric outlined expectations for student group work and was introduced to students with

the first major group project. Specifically, it measured collaboration skills in five categories described in a way that is accessible to high school students: Contributing & Listening to Ideas, Sharing in Work Equally, Using Time Efficiently, Making Decisions, and Discussing Science (Supplemental Figure S1). Over the course of the academic year, there were four times that groups worked together on an extensive project that served to review, connect, and apply concepts from a particular unit. These projects generally corresponded with the end of each quarter and lasted 1–2 wk each. After the projects were complete, each student was instructed "Use the rubric to pick the level that you feel your group reached in its collaboration for each category, and list one or more specific examples in the evidence column for how you reached that level" on an individual Group Collaboration reflection, which was a self-assessment of his or her group's performance. The Group Collaboration tools were also used intermittently for smaller group projects. It was made clear to the students that they would not be evaluated negatively if they identified areas needing improvement, but instead would be evaluated on their ability to provide evidence for their choices and to explain how they would improve on these areas in the future. As with the Science Skills tools, outlining different categories important for collaboration in the Group Collaboration rubric allowed an explicit focus on a specific aspect of good group work depending on the day or project.

Science Skills Assessment and KI Rubric

The extent to which student understanding of science skills and scientific collaboration changed over the course of the year was measured with a simple assessment, the Science Skills assessment (Supplemental Figure S2). Students were told that the assessment was for the instructor's own use to improve the course and would solicit their feedback in different ways about what helped them learn in the class. The assessment was given to students as a pre-, mid- and postassessment during approximately the first few weeks of the first semester, the first week of the second semester, and the last week of the academic year, respectively, and took less than 15 min to administer and complete each time.

A KI framework was used to develop a rubric for scoring open-ended responses to question 2 ("Name three skills that you think are important for doing science well, and explain why you picked them") and question 5 ("What are your strengths in doing science?"). The Science Skills Knowledge Integration (SSKI) rubric was developed for this study to measure the extent to which students made links between specific skills and their importance for science (Figure 2 includes examples of student responses). The SSKI rubric, a five-level latent construct aligned with other KI rubrics (Linn *et al.*, 2006; Liu *et al.*, 2008), maps onto students' increasingly sophisticated understanding of the research process. All scoring levels were represented in student responses. Interrater reliability for the SSKI rubric was greater than 95% for both questions 2 and 5, with two raters, with a high agreement indicated by a Cohen's kappa of 0.976.

For assessing whether students had more awareness of what it takes to collaborate successfully, at midsemester and at the end of the year, an attempt was made to create a KI rubric for collaboration to analyze responses to question 4 on the Science Skills assessment. Unfortunately, the open-ended responses proved difficult to code and categorize, limiting further information about student understanding of scientific collaboration from this study.

Score	Link Levels	Description	Examples[a]
4	Complex	Elaborates a connection between 2 or more science skills	*The hypothesis/observations because its important to know what you think is gonna happen and to observe it. Last the experiment because you can test your hypothesis*
3	Full	Links a science skill to its importance for science by explaining at least one scientific thinking or experimental skill	*Experimenting because it helps to try different things*
2	Partial	Lists 1 or more specific science skills, but explanation absent or incomplete; or, explanation given, but no skill named	*Observing, hypothesizing*
1	No	Lists general academic skills, but does not link them to importance in science	*Memorization, because there is a lot of names for things*
0	Blank/Off Task	No answer; irrelevant; does not make sense; or, claim there are none	*Knowing lang an stuff*

Figure 2. SSKI rubric, with student examples. Examples are responses to the question "Name three skills that you think are important for doing science well, and explain why you picked them."

Table 1. Goals of Science Skills instruments

Instrument	Purpose	Teaching tool	Assessment tool
A–Z Science Skills[a]	Make key terms for describing skills needed to do experimental work accessible for students; student reference guide for postlab reflection.	+	−
Postlab reflection[a]	Asks students to reflect on what skills they used well, with examples, after an experiment or laboratory activity; allows teacher to formatively assess how students understand the skills required for scientific experimentation.	+	+
Group Collaboration rubric[b]	Makes effective strategies for successful scientific collaboration explicit and outlines expectations for student group work. Aims to promote awareness of what it takes to successfully collaborate, help students learn to value the benefits of teamwork for accomplishing major projects, and improve the quality of group work.	+	−
Group Collaboration reflection[b]	Asks students to self-assess the quality of their collaboration skills and group work among three levels for five major categories, list evidence for why they choose that level, and reflect on how they would improve in next group project.	+	+
Science Skills assessment	Allows teacher-researcher to evaluate how well students can identify important experimental skills (including collaboration), synthesize their understanding about the skills, and self-assess their performance in science class. This instrument was the pre-, mid-, and postassessment.	−	+
SSKI rubric	Allows teacher-researcher to assess the extent to which students are able to connect different ideas about skills required for experimental work at five levels of complexity. Used primarily to score responses to question 2 on Science Skills assessment.	−	+

[a]Science Skills tools.
[b]Group Collaboration tools.

While some of the items on the Science Skills assessment are self-assessments, the SSKI rubric does not measure how much the students perceive they learned about science skills nor their skill level, but instead measures the ability of the student to identify specific skills important for scientific experimentation and the extent to which they are able to connect different ideas about the scientific research process. See Table 1 for a complete list of the different instruments used and the purpose intended for each.

Study Design

The study was designed to simultaneously test two treatments, Science Skills or Group Collaboration, with two different groups of students and each group serving as a "no-treatment" comparison for other (Figure 3). This experimental design can be applied to any classroom situation in which the student population can be divided into two groups for two independent interventions. In this case, the groups represented two different periods of students taking the same course, a typical teaching assignment for high school science teachers.

During the first half of the year, group 2 was instructed to use the Science Skills assessment tools (A to Z Science Skills and postlab reflection, Table 1); at the same time, group 1 was instructed to use the Group Collaboration assessment tools (Group Collaboration rubric and reflection, Table 1). Thus, group 1 served as a no-treatment comparison for the group 2 Science Skills treatment, and group 2 served as a no-treatment comparison for the group 1 scientific collaboration treatment. Throughout the second half of the year, both groups were encouraged to develop scientific thinking and collaboration by using both types of assessment tools and therefore received instruction in both areas by the end of the course. KI gains were measured by scoring the Science Skills assessment responses to question 2 using the SSKI rubric (see above and Figure 2).

The work presented here began when the author participated in Project IMPACT, a university program that provides a professional learning community structure to take on action research with regular feedback from teacher colleagues (Curry, 2008). Teachers involved in this professional
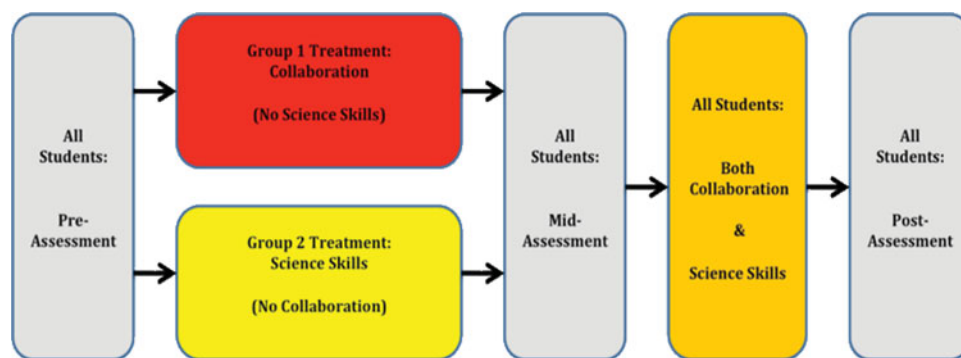

Figure 3. The study design allows for reciprocal testing of two simultaneous treatments.

learning community worked on individual research projects connected by the general theme of how to promote social justice in a science or mathematics classroom. Teachers worked in groups of five to six with guidance from one another and a trained facilitator, posing a research question, designing a study, and testing changes they implemented in their classrooms. The Science Skills and Group Collaboration tools were piloted by the author with 120–140 high school life sciences students over the 2-yr period prior to the academic year in which this study took place.

Statistical Methods

Statistical analysis was done using Stata Data Analysis and Statistical Software (2012) and the R Software Environment (2013), as indicated in relevant figure legends and associated text (see *Results*).

RESULTS

Snapshots of Inquiry in the Classroom Using the Science Skills Approach

Many science teachers in K–12 urban schools struggle to create a learning environment that contains the essential features of classroom inquiry. This includes a learning environment in which students engage in scientifically oriented questions, formulate explanations from evidence, and communicate and justify their proposed explanations (NRC, 2000; NGSS, 2013). Secondary life sciences teachers also struggle to unify inquiry-based lessons that span the entire curriculum in a cohesive manner. To meet the goal of promoting student inquiry, a new approach was designed and tested in the author's classroom, called "Science Skills," a term that was accessible for high school students.

To illustrate how student inquiry was built into the curriculum in this classroom, a snapshot for one particular inquiry lesson is described. This lesson, which was taught within a unit on the senses and the nervous system, was modified from *The Brain: Understanding Neurobiology through the Study of Addiction*, a unit within a published curricular series (NIH, 2000). For this lesson, all students in this study observed the effect of caffeine on the human body, generating their own hypotheses about the influence caffeinated soda drinks might have on their heart rates. They were instructed to choose controls to address different variables they expected to be important, record data in tables, and make graphs to analyze their data. As students worked together to make conclusions and compare findings with those of other student teams, students in one of the groups in this study were also asked to complete a postlab reflection (group 2; see the following section and *Methods* for details). One student focused on "observing" as a skill she used in this investigation, and remarked,

> When we did the caffeine lab I observed my heart rate and pulse every two minutes and made notes on it. Observing makes it easier for me to understand [the effects of caffeine].

Instruction for the routine skills required for all experimental work was integrated in a similar fashion into each of the different curricular units and content areas. Additional examples of student responses to the postlab reflection reveal the variety of scientific processes emphasized by different students and for other experiments and activities (see Supplemental Table S2). Student reflections were first assessed informally by checking in with each student individually before his or her completed reflection was accepted. This informal assessment served to corroborate students' self-assessment or to ask them to revise their responses until they matched the instructor's assessment of their mastery. Individual attention to students as they wrote their reflections was easy to manage, could be done quickly, and helped push students further in their understanding. For example, another student, who also focused on observation for a different laboratory activity, said,

> I think observing is a more nature skill everyone has [*sic*]. I personally specialize with that skill because I'm always check[ing] new things and experiments out whenever something seems interesting.

The student was asked to be more specific after the instructor checked in, at which time the student added,

> For example, when I had to dissect a cow's eye. I had to first examine and observe its outside.

While this student struggled to express himself articulately, this practice helped him advance his understanding of what approaches are important in scientific experiments. All of the students, in both groups 1 and 2, engaged in the same inquiry lessons at the same point in the curriculum. For example, in the case of the caffeine experiment, all students were instructed similarly to choose controls to address different variables, record data in tables, and analyze data with graphs. The two groups only differed in their use of additional tools that supplemented each inquiry lesson (see the following section).

Evidence of Gains in SSKI

For determining whether the Science Skills tools designed for this study were effective in promoting student inquiry and scientific thinking, student responses were evaluated for the pre-, mid-, and postassessment question 2 on the Science Skills assessment (see *Methods*). Two different groups of students participated in the study: group 2 was instructed to use the Science Skills tools in the first half of the year; during this same instructional time, group 1 was being taught to use the Group Collaboration tools (Figure 3). The learning goals, sequence of content taught, group projects, experiments, and activities were otherwise identical in both groups. Thus, group 1 served as a no-treatment comparison for the Science Skills treatment group 2, until the second half of the year, when both groups used both types of instruction and assessment tools (Table 1). Student responses for both groups were analyzed using the SSKI rubric (Figure 2) to measure the extent to which students made links between specific skills and their importance for science.

Results of this analysis revealed that students who used the Science Skills tools are better able to name and explain scientific-thinking strategies than their peers who did not use the same tools (Table 2). The average scores for responses from both groups started at a similar level, as a two-sample t test revealed there was no significant difference ($p = 0.55$). When the average pre- and midassessment KI scores for each student were compared, the change in scores for

Table 2. Students increase KI after using Science Skills instruction and assessment tools[a]

A. Average KI scores

Science Skills assessment	Group 1	Group 2
Pre	1.6	1.8
Mid	1.4	2.4
Post	2.7	2.5
n	19	18

B. Change in individual student's scores

Science Skills assessment	Group	Average (SE)	*p* Value	Cohen's *d*
Pre–mid change	1	−0.26 (.21)	0.88	−0.28
	2	0.44 (0.20)	0.021	0.57
Pre–post change	1	1.1 (0.27)	0.00053	1.0
	2	0.67 (0.23)	0.0048	0.70

[a] This analysis was done on student responses to question 2 on the Science Skills assessment ("Name three skills that you think are important for doing science well, and explain why you picked them."). A one-sided paired *t* test with 18 (or 17) degrees of freedom was performed for group 1 and group 2 changes in average KI scores, respectively. The null hypothesis was that there was no change and the average difference from pre- to midassessment or from pre- to postassessment was 0.

the no-treatment comparison group 1 was not statistically significant, as revealed by a one-sided paired *t* test ($p = 0.88$). However, for group 2, which received the Science Skills treatment, each student's score improved on average by 0.44 at midsemester, demonstrating a statistically significant difference in KI ($p = 0.021$). By the end of the school year, after both groups of students had received the Science Skills treatment, the average improvement on the assessment for group 1 was 1.1, and for group 2, it was 0.67. The average improvement in KI for each group pre- to postassessment is highly statistically significant (*p* values were 0.00053 and 0.0048, respectively).

Effect size, which helps determine the extent to which statistically significant changes are likely to be meaningful, was also calculated (Cohen, 1992). Consistently, Cohen's *d* measurements revealed that the Science Skills treatment for group 2 had a modest effect size midyear ($d = 0.57$). By the end of the year, when both groups had received the treatment, an even larger effect size was seen ($d = 1.0$ and $d = 0.70$, respectively; see Table 2B). Similar patterns were seen from analysis of question 5 responses, in which students clearly identified science skills when describing their own particular strengths in science, confirming the results obtained with question 2 (unpublished data).

Further analysis of student responses for the Science Skills assessment was done to gain additional insight on the distribution of the student scores for pre-, mid-, and postassessment of these two groups (Figure 4). A box plot reveals that the distribution was similar for both groups at the beginning of the year and for group 1 at midyear, before receiving the Science Skills treatment. However, the distribution shifted higher for group 2 midyear and for both groups at the end of the year, after both received the Science Skills treatment, confirming that students show gains in KI only after using these instruction and assessment tools.

Not only did students show increases in average KI scores for treatment groups, but comparison of individual students' scores for different assessments using scatter plots (Figure 5) revealed that most students showed an increase from a lower KI score to a higher score after they had used the Science Skills instructional and assessment tools. A comparison of pre- and midassessment scores (Figure 5A) shows that most students in the no-treatment group 1 tended to score the same or worse, whereas most students in the Science Skills treatment group 2 had immediately increased their understanding of science skills at the midyear point. In comparing pre- and postassessment scores, Figure 5B shows that students in both groups had improved KI scores at the end of the year. Thus, there is a clear relationship between increased student understanding of science skills and the use of the Science Skills tools.

Comparing Science Skills and Group Collaboration Treatments

Among many skills important for successful inquiry is group collaboration because the outcome of scientific projects and experiments depends on how well groups in classrooms or research laboratories function. In addition to helping students see collaboration as an important skill for scientific research, a goal for the introduction of Group Collaboration tools was for students to learn to appreciate collaborative group work more as an effective way to accomplish a significant science project. For assessing whether this goal was met, the five-level Likert-scale responses to the simple query in question 3 on the Science Skills assessment were analyzed by comparing the average change in KI scores for each student. For group 1, who had received the Group Collaboration treatment from the beginning of the year, an average of only 21% of the students had immediately increased their appreciation of group work at the midyear point; by the end of the year, 37% of the students on average increased their appreciation of group work as they continued to use the tools (as seen by comparing pre- and midassessments, and pre- and postassessments, respectively). For the no-treatment group 2, a surprising average of 33% of the students appreciated group work more when midyear responses were compared with those from the beginning of the year, but by the end of the year, this number had on average regressed back to the preassessment average (i.e., 0% of the students showed an increased appreciation of group work, the reason for which cannot be explained at
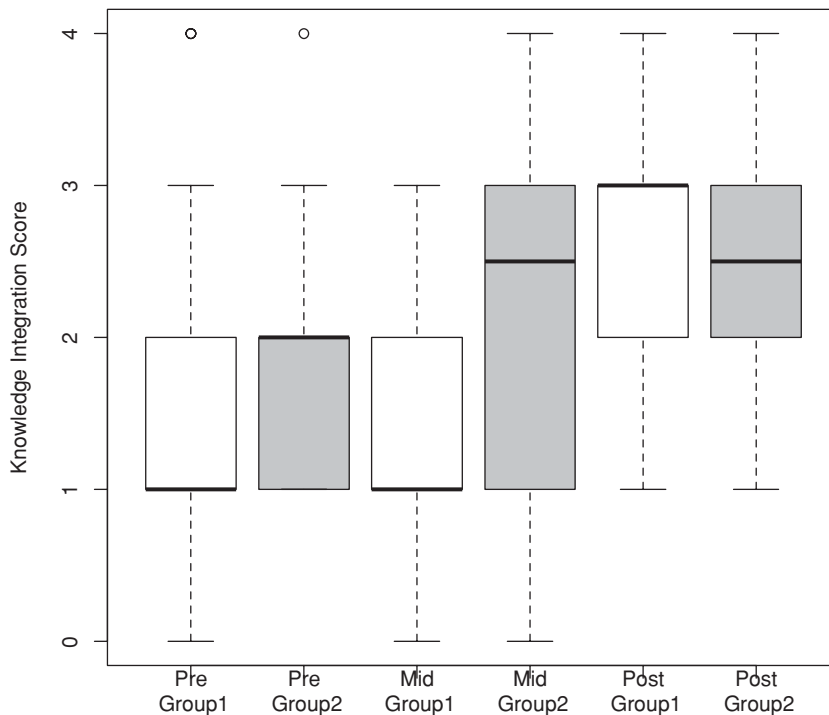
Figure 4. Box plots reveal differences in the distribution of KI scores after students use Science Skills instruction and assessment tools. Group 1 (white boxes) used the Science Skills tools only after they were assessed midyear ($n = 19$); group 2 (gray boxes) used Science Skills tools from the beginning of the course, were assessed at midyear, and continued use of the tools through the end of the year ($n = 18$). Note that the dark lines represent the median; the boxes include 50% of the data, representing the 1st to the 3rd quartile; 90% of the data is within the whiskers; and open circles represent outliers. Figure 4 shows the distribution of the same data set as is analyzed in Table 2 and Figure 5.

this time). Thus evaluation of pre-, mid-, and postassessment of the scientific collaboration treatment did not reveal any relationship between the Group Collaboration treatment and gains made in the appreciation students had for group work.

Analysis of Group Collaboration Reflection Responses

Another goal for promoting scientific collaboration with the use of Group Collaboration tools was to improve awareness of what it takes to collaborate successfully, including that successful collaboration means sharing ideas, distributing work, using time efficiently, and decision making. Responses to the Group Collaboration reflection revealed that students' self-assessment was both reflective and honest (see Table 3 for illustrative examples from each of these categories). The accuracy of their responses matched that of the teacher's own assessment, which frequently aligned well with students' self-assessments. For example, Student E discussed a common issue for group work, referring to the group needing to pace itself in order to better meet the project deadline. Additionally, different students on the same team often responded similarly, even though they had completed their reflections independently. Almost every student discussed some things they did well, at the exemplary level, and some things they could improve on, at the developing or beginner level. Interestingly, different groups of students answered differently, emphasizing that the categories outlined on the rubric are each important for collaborative work in a science class. Moreover, each assessment revealed that only about half of the students indicated they had reached what they considered to be the exemplary level on any of the five categories.

DISCUSSION

New approaches for teaching and assessing scientific skills and practices are critical for producing scientifically literate citizens (NGSS, 2013). This work shows that student understanding of scientific inquiry can be significantly increased by using instruction and assessment tools aimed at promoting development of specific inquiry skills. The success of the Science Skills approach can be attributed to being explicit with students about what skills are particularly important for progress in science, introducing specific terminology for experimentation, encouraging student self-assessment, and assessing scientific thinking in addition to content. Such strategies also place an emphasis on developing the academic language necessary for communicating in science and improving literacy (Snow, 2010).

Overall the analysis of responses to the pre-, mid-, and postassessments revealed that students expressed an increasing awareness of who they are as scientists and developed a more specific vocabulary for discussing experimentation and their strengths in science compared with students who had not used the same tools, meeting one of the key learning outcome goals for this classroom. Moreover, this new assessment approach enabled the teacher to work individually with struggling students to help them master critical skills; it was these students who often showed the greatest gains in learning how to talk about their experimental work (unpublished data). Not only do such assessments evaluate the extent to which students understand experimental skills, but they also serve as a tool for learning the skills and vocabulary themselves; assessments that accomplish both goals simultaneously have been dubbed "learning tests" by Linn and Chiu (2011).
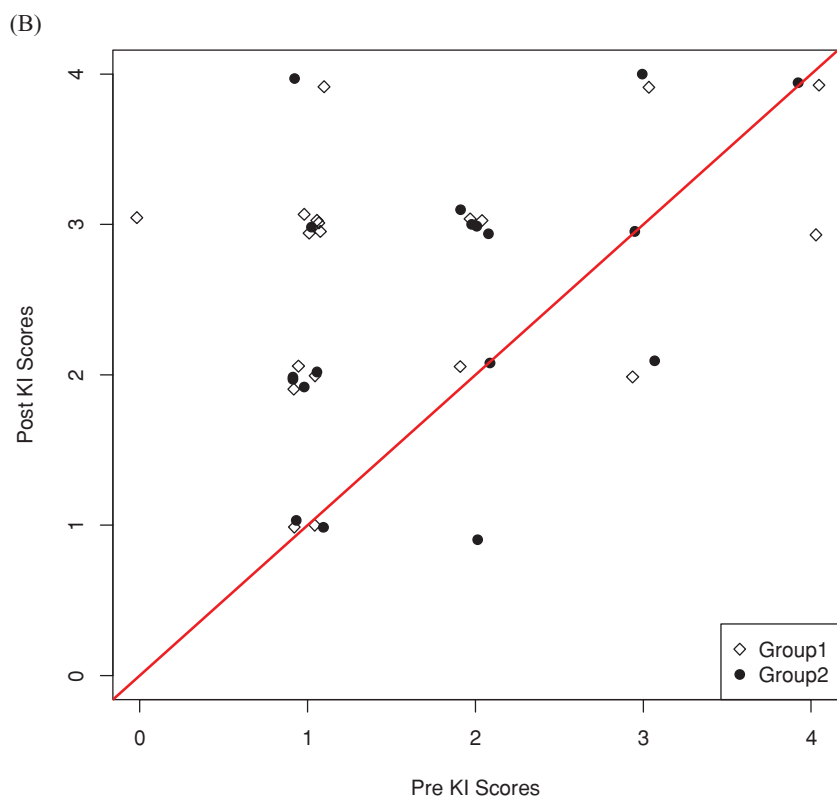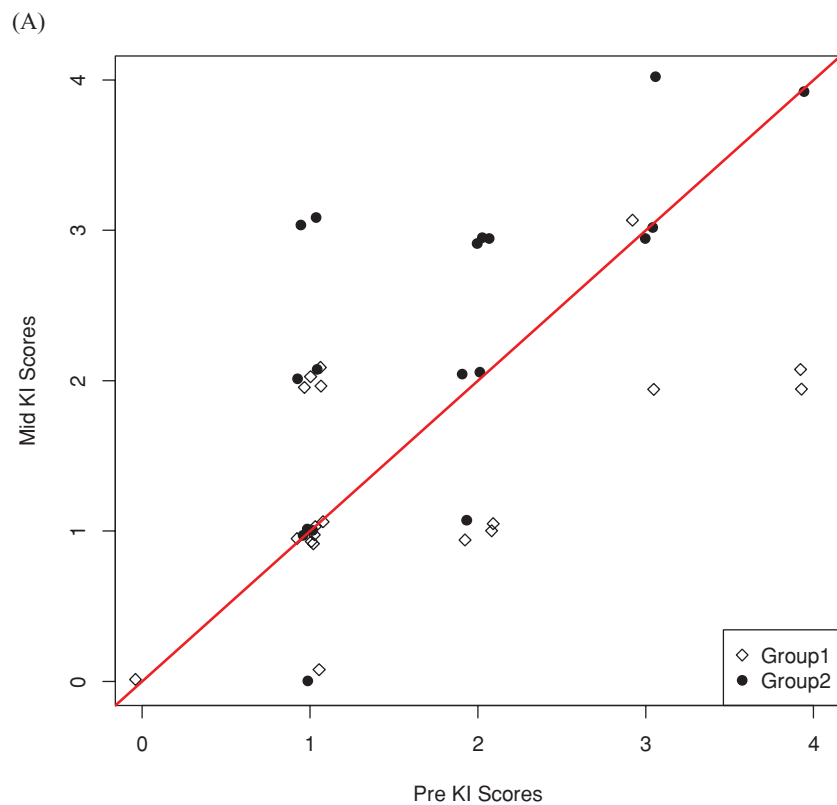
(A)

(B)

Figure 5. Scatter plots reveal increases in individual student's KI scores after using Science Skills instruction and assessment tools. (A) Comparison of pre- and midassessment scores for each student. (B) Comparison of pre- and postassessment scores for each student. Data points were jittered in the R Software Environment so that all were visible. A red line with a slope of 1 indicates no improvement; points above the line indicate increased improvement; points below the line indicate decreased improvement. Note that the data set is the same as that analyzed in Table 2 and Figure 4.

Table 3. Example Group Collaboration reflection responses

Category	Example	Level chosen	Response[a]
Contributing to ideas	A	Developing	I think that we did a good job overall but I think we could have talked more about it like explain stuff to each other.
	B	Exemplary	We worked together in the structure of our pamphlets. I thought of the phone number, 1-800-XIT-XTC and some stuff about the clinic. I would improve by doing a rough draft next time.
Sharing in work equally	C	Developing	Everyone was on their own for a little while.
	D	Exemplary	My group did a great job with sharing our work especially when some of us didn't have the info we needed.
Using time efficiently	E	Developing	We could have worked harder in the beginning so we wouldn't have to rush in the end.
	F	Exemplary	We finished on time.
Making decisions	G	Beginner	We didn't coordinate what we were going to write on the brochure very well.
	H	Exemplary	We didn't argue about the project, and any decisions were easily made.
Discussing science	I	Beginner	Our team talked mostly about other things than science. If we talked more about science then our work would have been better quality.
	J	Exemplary	We clarified our research on the neurons and how to visualize it. I think we could even communicate more next time.

[a]Note that misspelled words were corrected for clarity.

While clear learning gains were made for KI of scientific inquiry skills, the tools designed to promote a better understanding of group collaboration were not as successful. Several interpretations could account for why the collaboration treatment was not as effective, including: 1) in contrast to science skills, group work is something with which students already have a lot of experience, as well as the vocabulary for describing strengths and limitations of good collaboration; 2) the maturity level of high school students makes the social interactions required for negotiating tasks like sharing work and making decisions difficult; 3) the measure was not optimal, for example, the pre–post questions did not fully elicit what students understood about scientific collaboration; and 4) implementation of the Group Collaboration rubric was not ideal. Plans for improving the measure and its implementation in the future include probing student understanding of group work with other questions, guiding students to be more specific in their responses, and performing more formative assessment and collecting suggestions for improvement from the students during the project group work. Despite limited success with these measures, from the teacher's perspective, the author found that listening to students discuss the Group Collaboration rubric and reading student responses for the Group Collaboration reflection were useful for understanding class patterns of what was working and not working for the students during collaborative work, as well as for uncovering problems with group dynamics for particular student teams. Although classroom group work is known to be difficult to implement effectively (Cohen *et al.*, 1999), it is also an important component of successful teaching and is thus worthy of further investigation.

Action Research on Classroom Practices

When introducing a new teaching strategy, it is often difficult to determine its impact in isolation from other instructional approaches used in the classroom. While many instructors are interested in testing new teaching approaches in their own classrooms, questions of ethics quickly arise when one considers exposing some students to new strategies designed to improve learning, while a control group may not benefit from those same experimental strategies. Moreover, randomized field trials, the current gold standard for educational research, are impractical for the typical K–12 or college classroom instructor. Nonetheless, there is a need for improvement of scientific approaches to science education (Wieman, 2007; Asai, 2011). Teacher research, also known as teacher inquiry or action research, is an intentional and systematic approach to educational research in which data are collected and analyzed by individual teachers in their own classrooms to improve their teaching practices (Cochran-Smith and Lytle, 1993). This teacher-driven action research project served to improve the author's own teaching practice and is an example for other instructors on how to manage effective educational research while teaching high school, undergraduate, or graduate classes. Not only did these findings provide evidence for the educational benefits of the Science Skills approach to promoting scientific inquiry, but research in the context of the author's own classroom also allowed her to question the Group Collaboration approach and plan next steps for making it more effective. Being involved in an action research group can be a valuable professional development opportunity for any instructor, as it provides an opportunity to reflect on teaching strategies, engage in data analysis of student work, learn from colleagues, and consequently improve teaching practice.

Applications for Undergraduate Life Sciences Education

While this study is set in a high school context, lessons learned can easily transfer to introductory life sciences courses at the undergraduate level. The experimental design shown in Figure 3 allows the instructor to simultaneously test two treatments with two different groups of students, with each group serving as a no-treatment comparison for the other. This experimental design can be applied to any classroom situation in which the student population can be divided into two

groups for two independent interventions. Examples of other contexts for which this design could be useful are parallel discussion or laboratory sections for the same undergraduate course or a large lecture format that can be divided into two groups to test the impact of implementing two independent teaching strategies. The KI perspective described here is a promising framework with which to evaluate the effectiveness of both K–12 and undergraduate student learning in life sciences education.

ACKNOWLEDGMENTS

REFERENCES

Abd-El-Khalick F, Lederman NG (2000). Improving science teachers' conceptions of nature of science: a critical review of the literature. Intl J Sci Ed 22, 665–701.

American Association for the Advancement of Science (AAAS) (1998). Blueprints for Reform: Science, Mathematics, and Technology Education, Oxford University Press, New York, chap. 8.

AAAS (2009). Vision and Change: A Call to Action. www.visionandchange.org (accessed 15 November 2012).

Aronson BD, Silviera LA (2009). From genes to protein to behavior: a laboratory that enhances student understanding in cell and molecular biology. CBE Life Sci Educ 8, 291–308.

Asai DJ (2011). Measuring student development. Science 332, 895.

Barnes D, Britton J, Torbe M (1989). Language, the Learner and the School, Portsmouth, NH: Boynton Cook- Heinemann.

Casotti G, Rieser-Danner L, Knabb MT (2008). Successful implementation of inquiry-based physiology laboratories in undergraduate major and nonmajor courses. Adv Physiol Educ 32, 286–296.

Cochran-Smith M, Lytle S (1993). Inside Outside: Teacher Research and Knowledge, New York: Teachers College Press.

Cohen EG, Lotan RA, Scarloss BA, Arellano AR (1999). Complex instruction: equity in cooperative learning classrooms. Theor Pract 38, 80–86.

Cohen J (1992). Statistical power analysis. Curr Dir Psychol Sci 92, 98–101.

Curry MW (2008). Critical Friends Groups: the possibilities and limitations embedded in teacher professional communities aimed at instructional improvement and school reform. Teachers College Record 110, 733–774.

Derting TL, Ebert-May D (2010). Learner-centered inquiry in undergraduate biology: positive relationships with long-term student achievement. CBE Life Sci Educ 9, 462–472.

Dewey J (1916). Democracy and Education, New York: The McMillan Company.

Education Data Partnership (2010). Fiscal, Demographic, and Performance Data on California's K–12 Schools. www.ed-data.k12.ca.us (accessed 15 November 2012).

Feldman J (2010). Large and small high schools: Some key comparisons. In: Teaching without Bells: What We Can Learn from Powerful Practice in Small Schools, Boulder, CO: Paradigm, pp. 21–46.

Haberman M (2003). The Pedagogy of Poverty versus Good Teaching, New York: Education News.

Kalamuck KE, the Exploratorium Teacher Institute (2000). Human Body Explorations: Hands-on Investigations of What Makes Us Tick, Dubuque, IA: Kendall/Hunt.

Lau JM, Robinson GL (2009). Effectiveness of a cloning and sequencing exercise on student learning with subsequent publication in the National Center for Biotechnology GenBank. CBE Life Sci Educ 8, 326–337.

Lederman NG, Abd-El-Khalick F, Bell RL (2001). If we want to talk the talk, we must also walk the walk: the nature of science, professional development, and educational reform. In: Professional Development: Planning and Design, ed. J Rhoton and P Bowers, Arlington, VA: National Science Teachers Association.

Linn MC (2006). The knowledge integration perspective on learning and instruction. In: The Cambridge Handbook of the Learning Sciences, ed. RK Sawyer, New York: Cambridge University Press, 243–264.

Linn MC, Chiu J (2011). Combining learning and assessment to improve science education. Res Pract Assess 5, 4–13.

Linn ML, Lee H-S, Tinker R, Husic F, Chiu JL (2006). Teaching and assessing knowledge integration in science. Science 313, 1049–1050.

Liu OL, Lee H-S, Hofstetter C, Linn MC (2008). Assessing knowledge integration in science: construct, measures and evidence. Educ Assess 13, 33–55.

Lord T, Orkwiszewski T (2006). Moving from didactic to inquiry-based instruction in a science laboratory. Am Biol Teach 68, 342–345.

Marieb EN (2006). Essentials of Human Anatomy and Physiology, 8th ed. San Francisco: Pearson/Benjamin Cummings.

Moses B (2001). Radical Equations: Civil Rights from Mississippi to the Algebra Project, Boston: Beacon.

National Institutes of Health (2000). The Brain: Understanding Neurobiology through the Study of Addiction, Curriculum Supplement Series, Colorado Springs, CO/Seattle, WA: BSCS Press/Videodiscovery.

National Research Council (NRC) (1996). National Science Education Standards, Washington, DC: National Academies Press.

NRC (2000). Inquiry and the National Science Education Standards, Washington, DC: National Academies Press.

NRC (2003). BIO2010: Transforming Undergraduate Education for Future Research Biologists, Washington, DC: National Academies Press.

Next Generation Science Standards (2013). NGSS home page. www.nextgenscience.org (accessed 25 June 2013).

R Software Environment (2013). The R Project for Statistical Computing. www.r-project.org (accessed 25 June 2013).

Ribarič S, Kordaš M (2011). Teaching cardiovascular physiology with equivalent electronic circuits in a practically oriented teaching module. Adv Physiol Educ 35, 149–160.

Rissing SW, Cogan JG (2009). Can an inquiry approach improve college student learning in a teaching laboratory. CBE Life Sci Educ 8, 55–61.

Ryoo K, Linn MC (2012). Can dynamic visualizations improve middle school students' understanding of energy in photosynthesis. J Res Sci Teach 49, 218–243.

Siritunga D, Montero-Rojas M, Carrero K, Toro G, Vèlez A, Carrero-Martìnez FA (2011). Culturally relevant inquiry-based laboratory module implementations in upper-division genetics and cell biology teaching laboratories. CBE Life Sci Educ 10, 287–297.

Snow CE (2010). Academic language and the challenge of reading for learning about science. Science *328*, 450–452.

Stata Data Analyis and Statistical Software (2012). Stata/IC, Version 11. www.stata.com (accessed 15 November 2012).

Stigler J, Heibert J (1999). The Teaching Gap: The Best Ideas from the World's Teachers for Improving Education in the Classroom, New York: Free Press.

Tate ED, Clark DB, Gallagher JJ, McLaughlin D (2008). Designing science instruction for diverse learners. In: Designing Coherent Science Education: Implications for Curriculum, Instruction, and Policy, ed. Y Kali, MC Linn, and JE Roseman, New York: Teachers College Press, 65–93.

Treacy DJ, Sankaran SM, Gordon-Messer S, Saly D, Rebecca Miller R, Isaac RS, Kosinski-Collins MS (2011). Implementation of a project-based molecular biology laboratory emphasizing protein structure–function relationships in a large introductory biology laboratory course. CBE Life Sci Educ *10*, 18–24.

Wieman C (2007). Why not try a scientific approach to science education? Change (Sept/Oct), 9–15.

WISE (2012). The Web-based Inquiry Science Environment. wise.berkeley.edu (accessed 15 November 2012).

Zion M, Spektor-Levy O, Orchan Y, Shwartz A, Sadeh I, Kark S (2011). Tracking invasive birds: a programme for implementing dynamic open inquiry learning and conservation education. J Biol Educ *45*, 3–12.