HOW WE TEACH | Classroom and Laboratory Research Projects

Inquiry in oral communication: adapting oral examinations for teaching introductory physiology students to evaluate scientific research studies

Andrew R. St. James¹ and Darlene Campbell²

¹Department of Microbiology, Cornell University, Ithaca, New York; and ²Office of Undergraduate Biology, Cornell University, Ithaca, New York

Submitted 2 July 2019; accepted in final form 8 March 2020

St. James AR, Campbell D. Inquiry in oral communication: adapting oral examinations for teaching introductory physiology students to evaluate scientific research studies. Adv Physiol Educ 44: 192–202, 2020; doi:10.1152/advan.00087.2019.—Generating an increasingly skilled and numerous workforces of science, technology, engineering, and mathematics (STEM) professionals is a national priority. Central to this goal is improving the ability of STEM graduates to apply scientific inquiry within oral communication, a necessary skill for STEM professionals. In this case study of an introductory biology course, we observed that providing explanatory feedback to students responding to inquiry-based questions in an oral examination suggests improvements in students’ abilities to critically analyze results, draw conclusions, and discuss the broader implications of data. We found students struggled with generating hypotheses and constraining discussions of scientific limitation and broader implications. We show that low-performing students especially benefit from the feedback intervention. The findings of this study are applicable to college and university instructors who are looking to incorporate methods for teaching students to use scientific inquiry effectively during oral communication, particularly those with access to teaching assistants.

INTRODUCTION

Oral communication (both listening and speaking) is an essential job skill in most professional disciplines. A survey by the National Association of Colleges and Employers (NACE) found that 95.9% of employers rated oral communication skills as “essential,” whereas only 41.6% rated graduates as “proficient” in these skills (28). This is due in part to the fact that less than one-half of colleges and universities require students to take course work in oral communication, and those that do generally focus on formal presentation alone, which does not adequately prepare students for the myriad forms of communication found in the workplace (4, 8).

In another study where over 1,200 science, technology, engineering, and mathematics (STEM) job analysts were asked to rate critically important job skills in their fields, “active listening” was ranked as the third most important skill and “speaking” was the fourth (16). The first was “critical thinking.” These rankings highlight the necessity for teaching scientific inquiry within the context of oral communication. Teaching this skill should benefit students in their future professional careers.

One key aspect of applying scientific inquiry in oral communication is the immediacy of conversation, and methods for improving students’ scientific inquiry abilities within the context of oral communication should mirror this immediacy as much as possible. Numerous studies have investigated methods for improving scientific inquiry instruction, mainly focusing on teaching critical thinking within frameworks that allow students ample time to wrestle with complex scientific evidence (e.g., problem/project/inquiry-based learning) (17, 26, 34, 41, 42). However, there is a gap in the research literature with respect to methods that teach scientific inquiry within the framework of oral communication. Given that STEM professionals often discuss research findings orally, developing methods that explicitly train students to critically evaluate scientific research presented in this rapid-fire manner should be important.

Oral Examinations and Differentiated Instruction

Oral examination (also known as “viva voce”) is a commonly used assessment tool in postgraduate medical education and doctoral defenses but is less common in the U.S. undergraduate curriculum. As an assessment tool, oral examinations allow for more thorough evaluation of students’ content mastery, problem-solving skills, and conceptual misunderstandings and provide students the opportunity to practice one-on-one oral communication (29, 30, 32). The conversational aspect of the examinations allows for instructors to ask probing questions for concept clarification and enable students to self-correct and think more carefully about self-explanation (32), a practice that helps students integrate new knowledge into existing knowledge (6). Furthermore, by incorporating multiple oral examinations within a course, students become more comfortable with critically evaluating academic material in an impromptu, conversational manner, which could decrease anxiety and increase confidence when students advance to the professional workplace (30).

Another major benefit of oral examinations that is leveraged in this study is the ability to immediately provide individualized feedback within the exam environment. Extensive research on individualized, differentiated instruction has been conducted within the context of K–12 education. Fewer studies reporting methods of differentiated instruction exist within the context of college classrooms, although those that do provide...
evidence that differentiated instruction strongly benefits college students (22). In this study, we test a novel differentiated instruction tool: the effectiveness of explanatory feedback within the oral exam environment to improve the ability of undergraduates to use the process of scientific inquiry in the context of oral communication. Feedback has generally been linked to improved performance on exams (7, 36), with explanatory feedback mechanisms being viewed as more effective than corrective feedback mechanisms (27). One in-depth review of meta-analyses on feedback demonstrated that the most effective forms of explanatory feedback are those that provide reinforcements to learners, build upon previous feedback, are goal-oriented, and are visual or auditory (13).

Literature on the efficacy of oral examinations has demonstrated that they are appropriate as a method of assessment only when issues of validity, reliability, and fairness are overcome (24). Studies suggest that standardizing content and attending to structural and procedural aspects of oral examinations should be effective at mitigating many of the features of oral assessment that underlie these issues (e.g., maximizing question structure increases reliability, and improving examiner training increases fairness) (9, 24). However, given the highly interpersonal nature of oral examinations, issues arising from miscommunication or stress could still affect both how a candidate performs and how a candidate’s performance is perceived by an examiner, underscoring the advantage of multiple exposures to oral examination (1).

Objective of the Study

The objective of this case study was to determine whether explanatory feedback within oral examinations could be used as a method for teaching scientific inquiry within the context of oral communication. The objective was assessed by analyzing differences between cohorts of students who did and did not receive a feedback-based intervention in response to being asked inquiry-based questions during an oral examination. This study was divided into three phases. The “Non-intervention” phase had teaching assistants (TAs) measure student performance on a series of inquiry-based questions that were administered during eight oral examinations. The “Intervention 1.0” was identical to the Non-intervention phase in all respects, except that students were provided immediate explanatory feedback on their responses to the inquiry-based questions and engaged in a conversation about each question, with the TA administering the oral examination. At the end of the Intervention 1.0 phase, TAs sat for interviews about their experiences working with students. These data were used to redesign the study for the “Intervention 2.0” phase, which repeated the intervention from the previous phase with modifications, including changes to TA training, reordering of the assessment rubric, and rewriting of some question prompts.

METHODS

IRB Review

This study was granted exemption from full review by Cornell University’s Institutional Review Board (IRB) under protocol no. 1708007345.

Course of Study

This study was conducted at Cornell University, a private land grant Ivy League research university in Ithaca, NY, during the 2017–2018 Academic Calendar Year (ACY) and the Fall semester of the 2018–2019 ACY. The course of study was BIOG 1445: Introduction to Comparative Anatomy and Physiology, Individualized Instruction, which is an introductory-level biology course taken primarily by Biological Sciences majors, many of whom are on a premedicine track. Students in the Biological Sciences major are required to take an investigative biology laboratory course in addition to two courses from three core areas of biology: Comparative Physiology, Cell and Developmental Biology, and Ecology and the Environment. BIOG 1445 is one of two options that qualifies for the Comparative Physiology option. The other course, BIOG 1440, is a traditional lecture-style physiology course with similar content. Students choose which course to take based on their preference in taking the lecture version or the individualized instruction version.

BIOG 1445 course material is divided into 10 core units, which are each broken up into a series of objectives that are provided to students in the Survival Manual course (12). Students complete the objectives using a customized textbook (37), the course website (https://www.biog1445.org/), and demonstration materials available in the Study Center course, a multipurpose suite of rooms that students use for studying, tutoring, testing, viewing demonstration materials, and performing dissections. Individual study on each unit is complemented by ample access to one-to-one tutoring and help from TAs who staff the Study Center 7 days/wk to work with students as they prepare each unit. Student evaluation of each unit is accomplished through written quizzes and oral examinations. Oral examinations are administered by TAs, each of whom write their own examinations for each unit based on a standardized “hit list” of core concepts. Oral examinations last ~30–45 min, are graded pass/no pass, and are designed to assess for mastery learning (2).

Participants

Table 1 includes relevant information about students participating in the study.

Students in the Non-intervention phase had completed the least number of semesters of collegiate work on average. Students in all phases were similarly diverse in ethnicity, sex, and percentage of students who had completed Advanced Placement Biology in high school. The study also consisted of 36 TAs of varied experience, including 5 graduate (GTAs) and 31 undergraduate TAs (UTAs).

Table 1. Demographics of students participating in BIOG 1445 during each phase of the study

|                      | Non-intervention | Intervention 1.0 | Intervention 2.0 |
|----------------------|------------------|------------------|------------------|
| Students, n          | 90               | 47               | 63               |
| Average position (semesters competed) | 0.56             | 1.82             | 0.84             |
| Taken AP Biology, %  | 66               | 64               | 63               |
| Sex, %               |                  |                  |                  |
| Female               | 61               | 55               | 61               |
| Male                 | 39               | 45               | 35               |
| Ethnicity, %         |                  |                  |                  |
| Asian/Pacific Islander | 26              | 21               | 29               |
| Black                | 11               | 13               | 11               |
| Hispanic             | 10               | 13               | 16               |
| Native American      | 1                | 0                | 0                |
| White                | 29               | 34               | 24               |
| No response/other    | 23               | 19               | 20               |

n, No. of students. AP, Advanced Placement.
Use of Inquiry Questions in Oral Examinations

To assess the ability of students to use scientific inquiry in oral communication, we introduced standardized, inquiry-based questions into oral examinations for 8 of the 10 core units in the course (Fig. 1). Three of the eight prompts used in the Non-intervention and Intervention 1.0 phases were rewritten for the Intervention 2.0 phase based on feedback from the TAs at the end of the Intervention 1.0 phase, such that all the questions were based on relevant primary literature articles. Each question followed a structured format whereby students were read an observation based on the content in the unit for which they were being assessed and asked to generate a hypothesis regarding the observation. Students were then read a hypothetical hypothesis related to the observation, given a hypothetical experiment designed to test the hypothesis, and shown a hypothetical data figure from the experiment. Students were then asked to relate the results back to the hypothesis, draw a conclusion from the data, and discuss the limitations and broader implications of the study. In the Non-intervention and Intervention 1.0 phases, limitations and broader implications were asked after students drew conclusions. In the Intervention 2.0 phase, students were asked to provide limitations before drawing conclusions. This was based on feedback from TAs that students frequently gave definitive conclusions throughout the Non-intervention and Intervention 1.0 phases. By asking limitations before conclusions, we hoped to prime students to provide more constrained conclusions. Generally, the length of time students spent answering the inquiry question did not exceed five min.

The inquiry questions were routinely asked as the last question of the oral exam during which the student had already demonstrated mastery of the unit material, thus ensuring that only students who were about to pass their oral examination were asked the question. This was done to ensure that students only responded to each inquiry question once and to ensure that students were only assessed after they had demonstrated mastery of the content of the unit.

To help TAs prepare to assess a wide range of possible student responses, TAs attended a 15-min training session for each unit, led by the first author, that occurred during the course’s weekly TA training meetings. During these trainings in the Non-intervention and Intervention 1.0 phases, TAs were provided the inquiry question, along with sample student responses for each rubric section that they could reference in the testing room (see ENDNOTE for links to additional material). The first author led a discussion about the types of responses that were appropriate for each point level for each section on the rubric, and clarifying questions were fielded and discussed. TA trainings were updated in the Intervention 2.0 phase, so that TAs

Kidney stones are formed when large amounts of calcium are stored in the urine. The prevalence of kidney stones in northeast Thailand is especially high when compared to other regions of the world. Geologic surveys of villages with especially high incidences have often discovered large sodium salt deposits in the soil. In these villages, it is common that drinking water is sourced from deep wells.

1. Using these observations and your knowledge of renal biology, develop a hypothesis to explain the physiological relationship between sodium and calcium.

You set out to test the hypothesis that sodium inhibits the production of enzymes involved in preventing calcium stone formation. To test this hypothesis, you measure the levels of two enzymes involved in preventing calcium stone formation (THP and OPN) in urine in patients consuming a normal-sodium diet (NS) and a high-sodium diet (HS) after 42 days. The results are in the figure.

![Image Credit Adapted from: Huang, H-S. and Ma, M-C. (2015) High sodium-induced stress and poor antcrystallization defense aggravate calcium oxalate crystal formation in rat hyperoxaluric kidneys. PLoS ONE 10: e0134764.)](image)

2. Based on the data in the figure, relate the results back to the hypothesis and discuss what overall conclusions you can draw.

3. Discuss the limitations and broader implications of your results.

Fig. 1. Sample inquiry question prompt used by teaching assistants (TAs) during oral examinations. TAs read the prompt to students and provided students with the image.
responded to each question before the discussion led by the first author. This was done so TAs could engage with the questions more critically, and so a wider variety of potential responses could be used for training and was implemented as a result of TA feedback at the end of the Intervention 1.0 phase requesting more thorough training. During the trainings, TAs conferred in small groups before responding to inquiry questions using the freely available polling platform Poll Everywhere (https://www.polleverywhere.com/). Responses were then used for a whole group discussion, which included determining a rationale for how each answer would be scored and reviewing feedback strategies for how to tutor students when they gave incomplete or incorrect responses to the specific question.

To rate students’ scientific inquiry abilities for each unit, TAs scored student responses on a 15-point, four-category rubric (see ENDNOTE). The four rubric categories on which students were scored were 1) generating hypotheses; 2) relating results to the hypothesis; 3) drawing conclusions; and 4) discussing limitations and broader implications. Hypotheses were assessed based on the qualities of a hypothesis given in the student textbook, which includes narrow scope, testability/falsifiability, and relevance to the observation. Results were assessed for their accuracy in determining whether the data “supported” the hypothesis or not. Conclusions were scored based on the Association of American Colleges and Universities (AACU) “Inquiry and Analysis VALUE Rubric,” which assessed for specificity and how logically responses extrapolated the findings of the experimental data (31). Finally, limitations and broader implications were assessed for their relevance to the experiment and their insight, which was based on the same AACU rubric. In the Intervention 2.0 phase, the rubric was modified slightly to decouple limitations and broader implications, with each section being scored out of two points.

Analysis of Performance Data

Composite rubric scores for students within each Intervention phase were analyzed by comparing to students in the Non-intervention phase using chi-square contingency tables. For the Intervention 2.0 phase, no chi-square tests were performed for units 1, 3, and 7 because the question prompt was changed based on TA feedback.

In the Non-intervention and Intervention 1.0 phases, student scores were collected in a de-identified manner using paper rubrics. After feedback from TAs, collection of student scores was changed to the use of a Qualtrics survey that could individually track each student. Students were tracked using student numbers in accordance with the IRB exemption. Tracking of individual scores facilitated a more thorough statistical analysis of student data using a nonparametric Friedman test for repeated measures performed in IBM SPSS Statistics version 25. Students in phase III were separated into two groups for analysis. Students scoring at or below the pre-intervention median (unit 1 rubric score) were analyzed separately from students scoring above the pre-intervention median. Students in the former group were designated as the “low-performance” group, and students in the latter group as the “high-performance” group.

For analysis of student performance on individual rubric categories within each phase, we used a series of chi-square contingency tables to discern the effect of the intervention on individual rubric sections. Again, no comparisons were tested for units 1, 3, and 7 for the Intervention 2.0 phase.

In the Non-intervention and Intervention 1.0 phases, students were asked to rate their level of comfort in answering inquiry questions at the beginning of the semester and the end of the semesters. Students rated their comfort level using a 1–5 Likert scale.

Explanatory Feedback Intervention

Generally, TAs spent no more than 5 min providing explanatory feedback, which, when added to the time students spent responding to the questions, meant that the total length of the inquiry intervention did not exceed 10 min per student per oral exam. TAs were instructed in methods for providing explanatory feedback during a 30-min training at the beginning of each semester. TAs were then reminded of feedback strategies throughout the semester during training on each individual question. TAs were trained to provide feedback based on the model by Hattie and Timperley (13). Using this model, TAs were trained to provide feedback at four levels: self, task, process, and self-regulation. Self-feedback was used to reinforce positive aspects of the student’s performance, which could be based on delivery or level of thoughtfulness in response, in addition to the quality of the response. Task feedback centered around how well the student answered the question. TAs were encouraged to repeat student answers back when explaining the quality of the responses. Process feedback centered around teaching the student about the pedagogy of each question. Self-regulation feedback centered around providing the students strategies for regulating their performance, such as tips for identifying the key aspects of the prompt.

As an example of how this feedback would ideally work in an intervention, we will use an example student response for discussing limitations. The unit 4 inquiry question (see ENDNOTE) asked students to discuss the limitations of a study that was testing the hypothesis that Leigh syndrome, a rare inherited metabolic disorder, was caused by mutations in enzymes involved in the electron transport chain of cellular respiration. The study compared DNA from two siblings with the syndrome to over 100 healthy controls and found a single point mutation in one of the cytochromes involved in the electron transport chain. A common response from students would be that the researchers only looked at a single mutation, which limits their ability to determine whether other mutations are also important in the syndrome. This answer is not relevant to the study because the methodology involved comparing whole genomes, not a targeted analysis of a specific gene. Self-feedback to this response could emphasize the explanation given by the student, which would be a meaningful limitation if the study had been conducted using a targeted methodology that focused on a single gene. Task feedback would correct the student’s response by repeating the response and explaining why, given the methodology of the study, the response is irrelevant. Process feedback would then remind the student that limitations are rooted in the methodology, whereby the TA could explain the methodology and describe a potential correct response. Finally, self-regulation feedback would provide the student with strategies, which could include revisiting the methods that were used, and training the student to think about other methodologies that could be used to test the hypothesis, and build limitations from what those alternative methodologies might assess that the chosen methodology did not.

TA Feedback

To provide qualitative insights into student performance and suggest modifications to the study to be implemented in the Intervention 2.0 phase, TAs were interviewed by the first author at the end of the Intervention 1.0 phase. These interviews began after unit 6 during the semester and continued through the end of the course. The interviews were semistructured and consisted of questions related to perceptions of student performance, perceptions of TA performance, and perceptions of the challenges associated with teaching scientific inquiry for oral communication (see ENDNOTE). The interview script was designed to elicit feedback from TAs about students’ strengths and weaknesses with respect to performance on each criterion on the inquiry rubric and to understand the TAs’ perceptions of teaching scientific inquiry and their role in that process.

TA responses were recorded, transcribed, and categorized based on a list of codes and analyzed for themes using thematic analysis (5). Codes were generated using open coding, where codes are developed and modified throughout the coding process as opposed to line-by-line coding, and developed into themes (23). For example, in response to the question, “Describe the experience of asking and providing tutoring on the inquiry questions from your perspective as a TA,” several
responses were coded with “improved” and “important.” These responses were collated, and initial themes regarding “improved understanding of inquiry among TAs” and “recognition of the importance of teaching inquiry” were generated. All relevant data relating to these themes were then collected and reviewed, and the theme was formally refined. In this case, these themes were ultimately combined into a single theme, “TAs responded positively to participating in the study,” and relevant quotes were provided for support to elaborate on the theme. The second author certified the first author’s coding choices by reviewing interview scripts and coding decisions. Discrepancies were discussed, resolved, and certified by both authors.

Due to scheduling conflicts, not all TAs were able to be interviewed. However, all of the GTAs were interviewed, and 14 of the UTAs were interviewed. To assess differences between the inquiry abilities of GTAs and UTAs in oral communication, rubric scores were compared using a chi-square contingency table.

RESULTS

Student Performance

Overall student performance. Scores for students in the Intervention 1.0 phase were significantly higher than for those in the Non-intervention phase for units 3 and 9 (Fig. 2). Rubric scores for students in the Intervention 2.0 phase were significantly higher than for those in the Non-intervention phase for units 2, 4, 8, and 9 (Fig. 2).

A non-parametric Friedman test for repeated measures was run to determine if there were differences in composite rubric scores for students in two groups in the Intervention 2.0 phase: “low-performance” (pre-intervention score at or below median) and “high-performance” (pre-intervention score above median). Pairwise comparisons were performed (SPSS Statistics, 2017) with a Bonferroni correction for multiple comparisons. Composite rubric scores were significantly higher at the different time points during the explanatory feedback intervention for students in the low-performance group ($\chi^2 (7) = 28.372$, $P < 0.0005$). Post hoc analysis revealed composite rubric scores were significantly higher between pre-intervention (median = 10) and unit 6 (median = 13; $P = 0.022$), unit 7 (median = 13; $P = 0.033$), unit 8 (median = 13; $P = 0.019$), and unit 9 (median = 13.5; $P = 0.001$) (Fig. 3). Composite rubric scores were not statistically significantly different at the different time points during the explanatory feedback intervention for students in the high-performance group ($\chi^2 (7) = 17.155$, $P > 0.05$).

Students in both Intervention and Non-intervention groups reported low levels of comfort with the oral exam questions early in the semester ($\bar{x}_{\text{intervention}} = 2.59$, $\bar{x}_{\text{non-intervention}} = 2.88$) and high levels of comfort by the end of the semester ($\bar{x}_{\text{intervention}} = 4.00$, $\bar{x}_{\text{non-intervention}} = 4.04$).
Generating hypotheses. Rubric scores for students in the Intervention 1.0 phase were significantly higher than for those in the Non-intervention phase for units 1 and 3 (Fig. 4). Rubric scores for students in the Intervention 2.0 phase were significantly higher than for those in the Non-intervention group for unit 9 (Fig. 4). No TAs identified generating hypotheses as the most or least improved component of the inquiry rubric.

Relating results to the hypothesis. Rubric scores for students in the Intervention 1.0 phase were significantly higher than for those in the Non-intervention group for units 1, 3, 4, 6, 8, and 9 (Fig. 4). Rubric scores for students in the Intervention 2.0 phase were significantly higher than for those in the Non-intervention group for all units where the same question as the Non-intervention group was asked (Fig. 4). Most TAs reported that students made the strongest gains when connecting the results back to the hypothesis, especially with respect to the proper usage of the terms “support” and “does not support.” These gains were attributed to it being “easy to tell them that all of the evidence that we collect in science doesn’t really prove anything, but it points to a direction or a dogma that we’re trying to support” (GTA 3). However, TAs perceived that these gains did not necessarily extend to units where the data provided either did not support the hypothesis or did not adequately address the hypothesis (units 2 and 7). In these instances, it was difficult for students to realize “what they’re trying to find out and what the experiment did” when the experiment was “not really testing the hypothesis that we wanted so it’s not really supporting it or not” (UTA 14).

Drawing conclusions. Rubric scores for students in the Intervention 1.0 phase were significantly higher than for those in the Non-intervention group for units 1, 3, and 9 (Fig. 4). Rubric scores for students in the Intervention 2.0 phase were significantly higher than for those in the Non-intervention group for all units where the same question as the Non-intervention group was asked, except unit 6 (Fig. 4). Few TAs mentioned drawing conclusions as the most or least improved inquiry topics; however, for those TAs who discussed this topic, common criticisms were that responses were “too definitive” and were “given as an analysis of whether the results support the hypothesis or not” (GTA 1) rather than as a conclusion, which could have stemmed from “a lot of issues in differentiating between results and conclusion” (UTA 9). Nevertheless, it was perceived that this was “less common [in the Intervention 1.0 phase than in the Non-intervention phase]” (GTA 1). TAs in the Intervention 2.0 informally reported that students made less definitive conclusions when the order of

![Fig. 4. Inquiry scores by rubric section and unit for students in each phase of the study. Top left: scores for “Generating Hypotheses.” Top right: scores for “Relating Results to the Hypothesis.” Bottom left: scores for “Drawing Conclusions.” Bottom right: scores for “Discussing Limitations and Broader Implications.” Scores for students in Limitations and Broader Implications were combined from each separate rubric section for students in the “Intervention 2.0” to reflect the scoring structure used for students in the “Non-intervention” and “Intervention 1.0” phases. Bars represent median of student scores. Black bars, Non-intervention phase; dark gray bars, Intervention 1.0 phase; light gray bars, Intervention 2.0 phase. Statistical significance was determined by chi-square. No chi-square tests were run between the Non-intervention and Intervention 2.0 phases for units 1, 3, and 7 due to different question prompts being used. *P < 0.05, **P < 0.01, and ***P < 0.005.](image-url)
questioning was changed, so that students were asked to discuss limitations before drawing conclusions.

**Discussing limitations and broader implications.** Rubric scores for students in the Intervention 1.0 phase were significantly higher than for those in the Non-intervention group for unit 9 (Fig. 4). Rubric scores for students in the Intervention 2.0 phase were significantly higher than for those in the Non-intervention group for units 2, 4, and 9 (Fig. 4). Most TAs identified discussing limitations and broader implications as the rubric topic, with the lowest level of improvement throughout the semester. One common mistake made by students when discussing limitations was that their discussions were “way outside the scope [of the experiment] and looking at something that’s completely unrelated” (UTA 5). Students frequently lost sight of either the hypothesis that was being tested or the methodology used in the experiment. Occasionally, students would “repeat whatever they were told was the correct limitation from the previous time” (UTA 13) without fully considering whether they were applicable to the unit being assessed. This was reported to be because limitations are a little bit harder to explain because even as scientists we perform studies without thinking about the limitations of the study we’re performing. It’s really context dependent and each example requires a specific answer to explain to them what a limitation would be for the study (GTA 3).

A similar issue was seen in students’ broader implications, where students tended to provide implications that were too far overreaching or just restatements of the conclusion. Generally, students struggled to come up with broader implications that would lead to “both a practical and meaningful follow-up to the experiment” (GTA 1).

**TA Perceptions of the Study**

**Observed themes in student performance.** Two themes regarding student behaviors emerged from TA interviews and student end-of-semester feedback. The first theme was that some students began to give more structured and formulaic answers as the intervention semester progressed. TAs disagreed about whether the formulaic nature of student responses indicated depth of learning. For instance, while UTA 2 felt that repeat testing meant “students could tell what was coming and be able to formulate a more structured answer”, UTA 12 felt that “they learned what we expected to hear and not necessarily better understanding the questions presented to them.” Students also disagreed about the formulaic nature of the questioning. For instance, while one student reported “the formula is pretty helpful in training students to formulate hypotheses and recognize trends in data,” another student reported, “I just answered all of them the same way I usually did because I was unsure how to answer better.”

The second theme was that not all TAs perceived students as responding to the inquiry questions with a similar degree of seriousness. The consensus among TAs was that some students did not take the tutoring intervention seriously because “sometimes they know that it doesn’t count so they’ll say whatever and they won’t really think about it enough” (GTA 4). Nevertheless, TAs reported that students who did not take the intervention seriously would still “get the concept” (GTA 4), and that those students who did take the intervention seriously improved a lot because they were “trying to run through the criteria we had given them and the feedback we’ve given them while they were answering” (UTA 4).

**Challenges of taking oral exams.** Some of the challenges of taking an oral exam as perceived by the TAs emerged during TA interviews. One challenge was that focusing on the important information in the prompt was difficult when the prompt was delivered orally. As reported by GTA 1, “Definitely the hardest part is listening and picking out the key details. It’s a challenging listening experience as much as it is anything else.” These listening challenges could be a barrier to students who don’t “hear things well” (UTA 6).

Another challenge was that responding to an inquiry question is “pretty nerve-wracking because it’s on the spot” (UTA 9). In fact, having an examiner waiting for a response was reported to lead to “exhaustion” (UTA 8) and be “really, really stressful,” which could lead to students becoming “freaked out” because of having to “hear their silence” (UTA 10). This anxiety was confounded by the fact that the inquiry question topics were often, “something you’ve never heard of” (UTA 4). Nevertheless, TAs reported that the experience for students would likely become less stressful throughout the semester as they were regularly exposed to the intervention because “it’s not the most stressful thing if the students know they’re going to do it every time” (UTA 11).

**Perceptions of TA benefits.** TAs responded positively to participating in the study. They reported that, “as opposed to just teaching the course content, the inquiry questions include concepts that are much broader and theoretical, and explaining that, it’s kind of exciting” (GTA 1). TAs also felt that their own performance as examiners “really improved over the semester” (UTA 2), leading to a “better grasp of inquiry” (UTA 4) and a deeper understanding of “how much the students didn’t know... and how specific [responses] have to get” (UTA 9). Generally, TAs reported that teaching inquiry skills in the context of oral communication was “super important” (UTA 2), because, in their own experiences, TAs have “definitely had to critically think in that way” in their own studies, and, if there had been “someone sitting [them] down and explaining how to do it... there would’ve been a few times where things would have been easier” (UTA 14).

**DISCUSSION**

**Student Performance**

**Overall student performance.** Students in the Intervention 1.0 had completed an additional semester of study compared with the Non-intervention students, which may explain the higher initial scores for unit 1. In the Intervention 1.0 phase, the increase in the number of units with a significant positive improvement suggests that improving the TA training with respect to explanatory feedback and reordering the prompt improved student performance over the course of the study. The repeated-measures test conducted on the Intervention 2.0 group suggests that the explanatory feedback intervention leads to significant overall gains in the ability of low-performing students to apply the process of inquiry within the context of orally communicated data. It also shows that high-performing students maintain a high-performance level over the course of the intervention.

**Generating hypotheses.** The data suggests that students in all phases failed to master the ability to generate narrow, testable
hypotheses. Generally, students’ hypotheses were too vague, a finding that is in line with previous studies, which have also found that vague, confusing hypotheses are common mistakes made by students (35, 43). The hypothesis generation portion of the inquiry question emphasizes students’ abilities to apply hypothetico-deductive reasoning skills, whereas the other sections emphasize argumentation skills. The inability of students to master generating narrow, testable hypotheses suggests that hypothetico-deductive reasoning may be more appropriately taught elsewhere, for instance in a laboratory course where hypothesis generation could be incorporated in a more applied setting with additional instruction in hypothetico-deductive reasoning.

**Relating results to the hypothesis.** The intervention seems to have a positive effect on students’ abilities to relate results to hypotheses, and explanatory feedback and improving TA training further increase this effect. TAs reported that students struggled with identifying when studies did not support their respective hypotheses, and these observations are supported by the lower rubric scores for this section seen in these units. This suggests that student improvements were largely attributable to learning the terminology, rather than improving judgment of the hypothesis. This result is not surprising, given previous research that has shown students are less likely to consider alternative hypotheses when considering prespecified hypotheses, as opposed to when considering self-generated hypotheses (19). In the future, we would suggest incorporating more inquiry questions with nonsupporting data to help students improve their ability to distinguish these studies.

**Drawing conclusions.** The intervention seems to have a positive effect on students’ abilities to draw conclusions in the Intervention 2.0 phase. The increase in performance ability from the Intervention 1.0 phase to the Intervention 2.0 phase is likely attributable to two factors: 1) improved training of TAs, which could have improved the nature of the explanatory feedback; and 2) asking limitations before conclusions. As reported informally by TAs, changing the order of questioning likely led to improvements in student performance by priming students to consider limitations when drawing conclusions, which reduced the overall frequency of student responses that drew definitive conclusions from contextually limited data.

**Discussing limitations and broader implications.** The data of student scores suggest an increase in the overall performance of students in the Intervention 2.0 phase. As TAs pointed out, students struggled with discussing limitations and broader implications. The most common mistake students made during their oral examinations was not discussing limitations and broader implications that were relevant to the study at hand. This mistake is common in scientific studies, where scientists oftentimes make grandiose claims from contextually limited data (10). As was reported, “[Students] just wanted to repeat whatever they were told was the correct limitation from the previous time” (UTA 13), suggesting that students were not critically engaged with discussing limitations. One possible factor leading to the lack of critical engagement with discussing limitations and broader implications could have been insufficient training of TAs. During the Intervention 1.0 phase, TAs were only trained for 15 min per unit, which was not enough to review a wide scope of potential student responses. Skill level of instructors has been consistently linked to improved student outcomes (18, 44), so, if a group of TAs was less proficient in their inquiry abilities, then student outcomes would not reach their full potential. As reported, “The limitations are a little bit harder to explain because [they’re]..really context dependent” (GTA 3). This point emphasizes the necessity for teaching these skills to undergraduates in science courses.

The improved performance of students in the Intervention 2.0 group after improvements to TA training were made suggests that high-quality TA training plays an essential role in the success of the intervention. Further development of TA training could be performed to achieve more meaningful student gains with respect to discussing limitations and broader implications.

**Structure of Student Responses**

As some TAs and students reported, familiarity with the questioning led to responses that were more structured and formulaic as the semester progressed. While some discussed increased structure as a positive improvement, others perceived that this did not always correlate to deeper understanding. In general, the adoption of more structured responses should not be of concern, if the goal is to improve student abilities and the rubric is designed to assess conceptual understanding rather than response structure. Given the rubric designed for this study, the only section that considers structure is the “relates the results to the hypothesis” section, where maximum number of points can only be given to students who use the terminology “support” or “does not support.” All other sections of the rubric require the student to demonstrate deep conceptual understanding to receive full credit. Therefore, we believe that the rubric used in this study is robust, and we believe that the increased structure of student responses is in fact indicative of increased conceptual understanding.

As some TAs noted, not all students took the inquiry questions seriously, because they knew that their responses would not affect their grades; however, students who did take the questions seriously generally showed marked improvement. Generally, though, it was perceived that most students took the questions seriously, and that those who did not were in the minority. Nevertheless, another instructional implication from this data is that students should be held accountable for their responses. We would recommend tracking individual students and holding them accountable for growth rather than for a specific numerical score, as this would encourage students to think more critically about their responses while minimizing the added pressure of having to reach a specific score.

**Student Focus and Anxiety**

Staying focused while listening to inquiry questions was reported to be challenging, and having to provide a response immediately was reported to be stressful. These observations are supported by the literature, which has shown that increased anxiety levels are common for students taking oral examinations (20, 33), and that examination stress can lead to impairments in attention, retrieval of information from memory, and decision making (21, 38).

To help reduce anxiety related to tracking the question, we would recommend incorporating the explicit teaching of listening skills into the feedback protocol. This should teach students how to focus on the key parts of each question and
help them feel more comfortable providing a response. With regards to general stress from the oral examination environment, we feel that the use of numerous oral examinations throughout the semester will help reduce student stress. This is supported by our data showing that students’ levels of comfort improved over the course of the interventions. Another modification could be to introduce inquiry questions during laboratory or lecture that mimic the Intervention 2.0 phase TA trainings. Briefly, students could work in small groups to think of responses and submit them through a polling software, which the instructor could use to generate discussion about each question. This formative assessment would give students more practice in a low-stress situation. Reducing student anxiety is essential, as anxiety has been shown to interfere with information processing (15).

Improving TA Training in the Future

Efforts should be made to improve TA inquiry instruction in the future. Several studies have looked at inquiry-oriented TA training methods for laboratory courses (3, 11, 14, 25, 39, 40), and these studies can provide some insights into methods that could be effectively leveraged for training TAs to teach inquiry within the context of the methodology presented in this study. For example, Wheeler et al. (39) describe a successful training program for both UTAs and GTAs teaching an inquiry-based general chemistry laboratory course. In their program, Wheeler and colleagues focus on three categories of TA training: 1) theoretical foundations; 2) pedagogical strategies; and 3) practical course details. Here, theoretical foundation training could focus both on teaching TAs more generally about inquiry-based learning theory and more specifically about the place of oral communication in science. Pedagogical strategy training could introduce TAs to effective teaching methods (e.g., guided questioning techniques) and could give TAs opportunities to practice using these methods by modeling oral examinations and practicing teaching with feedback from peers and instructors. Peer and faculty observations could also be incorporated throughout the semester to better evaluate TAs interactions with students. Formative assessment of TAs by peers and instructors would allow for interventions to be designed that help improve TA feedback. Finally, making sure that TAs are knowledgeable about the practical details of the course and their expectations in the inquiry intervention would be essential.

We do not recommend that the TA-based assessment and feedback be removed from this method. We feel that the time saved by training TAs more effectively far outweighs the commitment placed on instructors of record taking on the sole responsibility of assessing these abilities. Furthermore, we feel that the learning benefits and general excitement of TAs should not be neglected. Students serve to benefit from having TAs that are excited about teaching, and TAs interested in future academic work serve to benefit from a deeper understanding of student misconceptions regarding inquiry.

Applying the Method across Instructional Frameworks

Given that individualized instruction models and the use of oral examination is not widespread in undergraduate institutions, we hesitate to generalize the efficacy of our findings and methodology to more traditional course frameworks. For instructors wishing to implement these methods within more traditional courses, troubleshooting of the oral examination format would likely be necessary. Nevertheless, we do believe that this form of instruction could be amenable to other course frameworks.

Given the short time frame of the inquiry portion of the oral examinations (~10 min), these exams could feasibly be incorporated into more traditional lecture-based courses or laboratories if enough TAs are available for carrying out the assessments. In the individualized instruction model from the case study presented here, students complete the course of study independently, which means that, while students have the opportunity to interact with TAs on a regular basis in tutoring sessions, the oral exam room is the only instance where students are required to be present with an instructor, and our methodology has focused on applicability within this constrained course framework. One major advantage of this framework is the depth of individualized instruction that can occur during one-on-one interactions with TAs. To improve instruction within these constraints, more objectives about scientific inquiry could be incorporated into the course.

A potential benefit of incorporating this mode of assessment into a more traditional course is the ability to explicitly teach scientific inquiry and have students practice communicating with these skills outside of the stress of an exam environment. By decoupling instruction from assessment, the instructional component is less individualized, but the assessment and explanatory feedback would remain one to one. Student practice with scientific inquiry for oral communication could occur in small groups in either the classroom or laboratory.

A major benefit of this intervention is that the feedback is given during an oral examination, when students are interacting with TAs and instructors one on one. We believed this face-to-face immediate feedback is more beneficial than small-group or whole-group feedback, because it addresses everyone’s needs, and the conversational aspect of the oral exam provides space for deeper explanation of feedback. Therefore, we recommend that any modification of the module for alternative instructional settings retain the oral examination aspect and use small-group and whole-group activities for formative rather than summative assessment.

Limitations of the Study

TA inquiry abilities were not thoroughly assessed throughout the study. With many TAs, there was likely some diversity in the TAs’ conceptual knowledge of inquiry. TAs responded to each inquiry question during training before they were provided rubrics with sample answers. We did not regularly observe low-quality answers from TAs during training. However, because these trainings were completed in small groups, we were unable to assess each individual TA. In the future, we would recommend revising the intervention to more thoroughly develop TA inquiry abilities before students begin the intervention.

Because TA oral exams are unsupervised, there was no way to assess the quality of TA feedback and whether TAs were following the suggestions in training for how to provide feedback. This limits our ability to assess whether students were receiving a similar quality of instruction. This was done for practical reasons. Over 1,000 oral exams were given during the
semester, with multiple exams often occurring simultaneously in different rooms, which made it impossible to supervise all exams. Moving forward, the TA training could be further improved by introducing role-playing scenarios where TAs must practice giving explanatory feedback.

It is important to note that the number of oral examinations administered by each TA was highly variable, which limits the generalizability of TA interpretations, especially for those TAs who administered relatively few examinations and could, therefore, have biased perceptions. Additionally, because interviews started after unit 6, most TAs were interviewed before students had received all of the interventions, which could have biased their interpretations of student performance, limiting the generalizability of their claims. Nevertheless, common themes emerged from several TAs, regardless of how many students they had tested or how early or late in the semester they were interviewed. Given the emergence of thematic commonalities, we do not believe these limitations sufficiently undermine the validity of our conclusions.

Individual tracking of student scores was not implemented until the Intervention 2.0 phase, which meant that a repeated-measures test to compare individual scores across units was only conducted on a subset of students receiving the intervention. While these results demonstrated significant improvement in student performance, the validity of our conclusions could be further strengthened by conducting a larger-scale study involving a broader sample of students. Moving forward, the TA training could be further refined to ensure that TAs are better prepared to provide effective feedback to students.

ACKNOWLEDGMENTS

We thank Dr. Kimberly Williams for helpful insights with the project design and for guidance on education research methods and manuscript review throughout the course of the project. We also thank the many teaching assistants who were willing to collect data for this study and provide invaluable insights into student performance during interviews: Natalie Arimah, David Boak, Joe Cammarata, Tracy Cheung, Chloe Citron, Delaney Ding, Larry Dunn, Kevin Feng, Emma Frawley, Saman Gavrasia, Matthew Guo, Michael Heyan, Nikki Joh-Carnella, Lauren Kam, Ryan Kim, Michelle Laterrade, Zoe Lee-McDermott, Mercedes Lewandrowski, Rachel Lippman, Beverly Lo, Heyang, Nikki Joh-Carnella, Lauren Kam, Ryan Kim, Michelle Laterrade, Zoe Lee-McDermott, Mercedes Lewandrowski, Rachel Lippman, Beverly Lo, Heyang, and Amy Zhao.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

A.R.S.J. and D.C. conceived and designed research; A.R.S.J. performed experiments; A.R.S.J. analyzed data; A.R.S.J. and D.C. interpreted results of experiments; A.R.S.J. prepared figures; A.R.S.J. drafted manuscript; A.R.S.J. and D.C. edited and revised manuscript; A.R.S.J. and D.C. approved final version of manuscript.

ENDNOTE

At the request of the authors, readers are herein alerted to the fact that additional materials related to this manuscript may be found at https://doi.org/10.6084/m9.figshare.8411744, https://doi.org/10.6084/m9.figshare.8411693, and https://doi.org/10.6084/m9.figshare.8411750. These materials are not part of this manuscript and have not undergone peer review by the American Physiological Society (APS). APS and the journal editors take no responsibility for these materials, for the website address, or for any links to or from it.

REFERENCES

1. Abrahamson S. The oral examination: the case for and the case against. In: Evaluating the Skills of Medical Specialists, edited by Loyd JS, Langsley DG. Chicago, IL: American Board of Medical Specialists, 1983, p. 121–124.

2. Bloom BS. Learning for mastery. Eval Comment 1: 1–12, 1968.

3. Bohrer K, Stegenga BD, Ferrier A. Training and mentoring TAs in inquiry-based methods. In: Proceedings of the 28th Workshop/Conference of the Association for Biology Laboratory Education (ABLE). West Lafayette, IN: ABLE, 2007, p. 335–346.

4. Bok D. Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More. Princeton, NJ: Princeton University Press, 2006.

5. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. London: Sage, 1998.

6. Chi MTH, De Leeuw N, Chiu M-H, Lavancher C. Citing self-explanations improves understanding. Cogn Sci 18: 439–477, 1994. doi:10.1016/0364-0213(94)90016-7.

7. Crippen KJ, Schraw G, Brooks DW. Performance-related feedback: the hallmark of efficient instruction. J Chem Educ 82: 641, 2005. doi:10.1021/ed082p641.

8. Crosling G. Transition to university: the role of oral communication in the undergraduate curriculum. J Inst Res 9: 69–77, 2000.

9. Davis MH, Karunathilake I. The place of the oral examination in today’s assessment systems. Med Teach 27: 294–297, 2005. doi:10.1080/01421590500126437.

10. Faber J. Writing scientific manuscripts: most common mistakes. Dental Press J Orthod 22: 113–117, 2017. doi:10.1590/2177-6709.22.5.113-117.sar.

11. Gormally C, Brickman P, Hallar B, Armstrong N. Lessons learned about implementing an inquiry-based curriculum in a college biology laboratory classroom. J Coll Sci Teach 40: 45–51, 2011.

12. Hardy C, Campbell D, Plessia A, Schiller B. Biog 1445—Introduction to Comparative Comparative Anatomy and Physiology, Individualized Instruction Survival Manual. Ithaca, NY: Gnomon Copy Service, 2017.

13. Hattie J, Timperley H. The power of feedback. Rev Educ Res 77: 81–112, 2007. doi:10.3102/00346543007003840.

14. Hughes PW, Ellefson MR. Inquiry-based training improves teaching effectiveness of biology teaching assistants. PLoS One 8: e78540, 2013. doi:10.1371/journal.pone.0078540.

15. Ikeda M, Iwanaga M, Seiwa H. Test anxiety and working memory system. Percept Mot Skills 82, Suppl 3: 1223–1231, 1996. doi:10.2466/pms.1996.82.3c.1223.

16. Jang H. Identifying 21st century STEM competencies using workplace data. J Sci Educ Technol 25: 284–301, 2016. doi:10.1007/s10956-015-9593-1.

17. Jiang F, McComas WF. The effects of inquiry teaching on student science achievement and attitudes: Evidence from propensity score analysis of PISA data. Int J Sci Educ 37: 554–576, 2015. doi:10.1080/09500693.2014.1000426.

18. Kennedy M. How does professional development improve teaching? Rev Educ Res 86: 945–980, 2016. doi:10.3102/0034654315626800.

19. Koehler DJ. Hypothesis generation and confidence in judgment. J Exp Psychol 20: 461–469, 1994. doi:10.1037/0278-7393.20.2.461.

20. Lacey K, Zaharia MD, Griffiths J, Ravindran AV, Merali Z, Anisman H. A prospective study of neuroendocrine and immune alterations associated with the stress of an oral academic examination among graduate students. Psychoneuroendocrinology 25: 339–356, 2000. doi:10.1016/S0306-4530(99)00059-1.

21. LeBlanc VR. The effects of acute stress on performance: implications for health professions education. Acad Med 84, Suppl: S25–S33, 2009. doi:10.1097/ACM.0b013e3181b37bf8f.

22. Lightweiss SK. College success: a fresh look at differentiated instruction and other student-centered strategies. Coll Q 16: 1–9, 2013.

23. Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. All Irel J Teach Learn High Educ 8: 3351–33514, 2017.

24. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examination: establishing conditions for validity, reliability and fairness. Adv Health Sci Educ Theory Pract 15: 277–289, 2010. doi:10.1007/s10459-008-9111-9.

25. Miller K, Brickman P, Oliver JS. Enhancing teaching assistants’ (TAs’) inquiry teaching by means of teaching observations and reflective discourse. Sch Sci Math 114: 178–190, 2014. doi:10.1111/smm.12065.

26. Minner DD, Levy AJ, Century J. Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years
1984 to 2002. J Res Sci Teach 47: 474–496, 2010. doi:10.1002/tea.20347.

27. Moreno R. Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback in discovery-based multimedia. Instr Sci 32: 99–113, 2004. doi:10.1023/B:TRUC.0000021811.66966.1d.

28. NACE. Job Outlook 2018. Bethlehem, PA: NACE, 2018.

29. Oakley B, Hencken C. Oral examination assessment practices: effectiveness and change within a first year undergraduate cohort. J Hosp Leis Sport Tour Educ 4: 3–14, 2005. doi:10.3794/johlste.41.88.

30. Pearce G, Lee G. Viva voce (oral examination) as an assessment method: insights from marketing students. J Mark Educ 31: 120–130, 2009. doi:10.1177/0273475309334050.

31. Rhodes T. Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics. Washington, DC: Association of American Colleges and Universities, 2010.

32. Roecker L. Using oral examination as a technique to assess student understanding and teaching effectiveness. J Chem Educ 84: 1663, 2007. doi:10.1021/acs.jchemed.6b00831.

33. Schoofs D, Hartmann R, Wolf OT. Neuroendocrine stress responses to an oral academic examination: No strong influence of sex, repeated participation and personality traits. Stress 11: 52–61, 2008. doi:10.1080/10253890701453943.

34. Schroeder CM, Scott TP, Tolson H, Huang T-Y, Lee Y-H. A meta-analysis of national research: effects of teaching strategies on student achievement in science in the United States. J Res Sci Teach 44: 1436–1460, 2007. doi:10.1002/tea.20212.

35. Strangman I., Knowles E. Improving the development of student’s research questions and hypotheses in an introductory business research methods course. Int J Scholarsh Teach Learn 6: 2, 2012. doi:10.20429/ijsotl.2012.060224.

36. Swanson HL, Lussier CM. A selective synthesis of the experimental literature on dynamic assessment. Rev Educ Res 71: 321–363, 2001. doi:10.3102/00346654071002321.

37. Urry L, Cain M, Wasserman S, Minorsky P, Reece JB. Campbell Biology (11th ed.). Hoboken, NJ: Pearson Education, 2017.

38. Vedhara K, Hyde J, Gilchrist ID, Tytherleigh M, Plummer S. Acute stress, memory, attention and cortisol. Psychoneuroendocrinology 25: 535–549, 2000. doi:10.1016/S0306-4530(00)00008-1.

39. Wheeler LB, Clark CP, Grisham CM. Transforming a traditional laboratory to an inquiry-based course: importance of training TAs when redesigning a curriculum. J Chem Educ 94: 1019–1026, 2017. doi:10.1021/acs.jchemed.6b00831.

40. Wheeler LB, Maeng JL, Whitworth BA. Teaching assistants’ perceptions of a training to support an inquiry-based general chemistry laboratory course. Chem Educ Res Pract 16: 824–842, 2015. doi:10.1039/C5RP00104H.

41. Wilson CD, Taylor JA, Kowalski SM, Carlson J. The relative effects and equity of inquiry-based and commonplace science teaching on students’ knowledge, reasoning, and argumentation. J Res Sci Teach 47: 276–301, 2010. doi:10.1002/tea.20329.

42. Wise KC, Okey JR. A meta-analysis of the effects of various science teaching strategies on achievement. J Res Sci Teach 20: 419–435, 1983. doi:10.1002/tea.3660200506.

43. Woolf BP, Marshall D, Mattingly M, Lewis J, Wright S, Jellison M, Murray T. Tracking student propositions in an inquiry system. In: Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies, edited by Hoppe U, Verdejo F, Kay J. Amsterdam: IOS, 2003, p. 21–28.

44. Zhang D, The Effects of Teacher Education Level, Teaching Experience, and Teaching Behaviors on Student Science Achievement. Logan, UT: Utah State University, 2008.