International Journal of Science Education

ISSN: 0950-0693 (Print) 1464-5289 (Online) Journal homepage: http://www.tandfonline.com/loi/tsed20

Written justifications to multiple-choice concept questions during active learning in class

Milo D. Koretsky, Bill J. Brooks & Adam Z. Higgins

To cite this article: Milo D. Koretsky, Bill J. Brooks & Adam Z. Higgins (2016): Written justifications to multiple-choice concept questions during active learning in class, International Journal of Science Education, DOI: 10.1080/09500693.2016.1214303

To link to this article: http://dx.doi.org/10.1080/09500693.2016.1214303

Published online: 16 Aug 2016.

INTERNATIONAL JOURNAL OF SCIENCE EDUCATION, 2016
http://dx.doi.org/10.1080/09500693.2016.1214303

Written justifications to multiple-choice concept questions during active learning in class

Milo D. Koretsky, Bill J. Brooks and Adam Z. Higgins

School of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, OR, USA

ABSTRACT

Increasingly, instructors of large, introductory STEM courses are having students actively engage during class by answering multiple-choice concept questions individually and in groups. This study investigates the use of a technology-based tool that allows students to answer such questions during class. The tool also allows the instructor to prompt students to provide written responses to justify the selection of the multiple-choice answer that they have chosen. We hypothesize that prompting students to explain and elaborate on their answer choices leads to greater focus and use of normative scientific reasoning processes, and will allow them to answer questions correctly more often. The study contains two parts. First, a crossover quasi-experimental design is employed to determine the influence of asking students to individually provide written explanations (treatment condition) of their answer choices to 39 concept questions as compared to students who do not. Second, we analyze a subset of the questions to see whether students identify the salient concepts and use appropriate reasoning in their explanations. Results show that soliciting written explanations can have a significant influence on answer choice and, when it does, that influence is usually positive. However, students are not always able to articulate the correct reason for their answer.

ARTICLE HISTORY
Received 3 January 2016
Accepted 14 July 2016

KEYWORDS
Conceptual learning; written explanation; clicker questions

Introduction

This study investigates undergraduate students' responses to *concept questions* in thermodynamics as they actively participate during class using technology. For this study, we use the term concept questions to describe qualitative, multiple-choice questions that require students to identify foundational concepts and then apply them in new situations. Concept questions are sometimes called 'ConcepTests' (Mazur, 1997) and form a common type of clicker question (Duncan, 2006).

Concept questions form the core of several concept-based active learning pedagogies such as peer instruction (Crouch & Mazur, 2001; Mazur, 1997). The primary objectives of these pedagogies are to make students aware of their need for conceptual understanding and then foster that understanding. Concept-based active learning has been demonstrated to increase student achievement (Freeman et al., 2014; Hake, 1998) and engagement

CONTACT Milo D. Koretsky milo.koretsky@oregonstate.edu, milo.koretsky@orst.edu School of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, OR 97331-2702, USA

(Deslauriers, Schelew, & Wieman, 2011) and to reduce the performance gap of underrepresented students (Haak, HilleRisLambers, Pitre, & Freeman, 2011; National Research Council, 2011, 2012). However, there is still debate over best practices and what implementation strategies are most effective (Caldwell, 2007; Vickrey, Rosploch, Rahmanian, Pilarz, & Stains, 2015).

For several decades, clickers have been used as the primary technological device to facilitate the delivery of concept questions during class (Caldwell, 2007; Kay & LeSage, 2009; MacArthur & Jones, 2008). Clickers allow all students in class to individually respond to concept questions in real time. They also enable data collection that holds students accountable for their answers and instantaneously provides the instructor performance results to facilitate classroom discussion and formative feedback. However, clicker technology has typically limited delivery exclusively to multiple-choice responses.

As the capabilities of technological tools increase, there is the opportunity to extend pedagogical practices associated with the use of concept questions. For example, researchers have hypothesized that soliciting individual written justifications of students' answer choices improves their understanding (Taylor & Nolen, 2007). In addition to their use as a tool to foster learning, written explanations can also provide valuable information to concurrently assess that learning. However, to our knowledge, there has not yet been a report of the influence on individual writing on students' initial responses to concept questions.

The tool used in this study, the *Concept Warehouse*, allows the instructor to prompt students to provide written responses to justify their selection of a multiple-choice answer. Using this tool, we empirically investigate the influence of writing on students' answer choices to concept questions by comparing responses of students who are prompted to write justifications (treatment) with those who are not (comparison).

Specifically, this study addresses the following research questions:

(1) Does requesting students to provide written explanations for their answer choices to multiple-choice concept questions influence their answer choices in an undergraduate introductory thermodynamics class? Is writing beneficial?
(2) Is the reasoning in the written responses consistent with their answer choices; that is, do students choose correct answers for the right reasons?

Theoretical framework

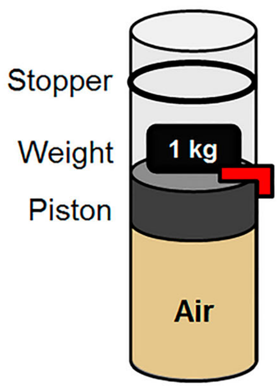
Concept questions and conceptual understanding

Concept questions are designed to be conceptually challenging and typically require little or no computation so that students cannot algorithmically rely on equations to obtain the answer (Beatty, Gerace, Leonard, & Dufresne, 2006; Wood, 2004). They ideally focus on the most important concepts in a subject, target a specific learning goal, uncover misconceptions, and elicit a range of responses (Caldwell, 2007; Crouch & Mazur, 2001; Tanner & Allen, 2005). Concept questions seek to shift engagement from asking the student 'What can you remember?' to 'What do you understand?' (Nurrenbern & Robinson, 1998).

Conceptual understanding is needed to consistently answer the concept questions correctly. We operationalize conceptual understanding as the connected organization of abstract knowledge (Bransford, Brown, & Cocking, 1999; Chi, Feltovich, & Glaser, 1981; Rittle-Johnson & Alibali, 1999) that may have practical utility when applied to science and engineering tasks. For example, Figure 1 shows a sample concept question as it was presented to students in class in the treatment condition in this study. To answer this question correctly, students should apply their understanding of conservation of energy recognizing that the work done by the gas on the surroundings lowers its internal energy and, therefore, its temperature. They do not need to calculate any values, but rather reason through their answer choice based on their understanding of appropriate scientific principles.

However, researchers have found that many science and engineering students who can answer quantitative, algorithmic problems correctly perform poorly when they are asked to answer concept questions on the very same topic and at the same time (Haláková & Prokša, 2007; Koretsky et al., 2011; McDermott, 2001; Papaphotis & Tsaparlis, 2008). VanLehn and Van de Sande (2009) characterize the ability to answer concept questions

Air at high pressure and ambient temperature is contained in a perfectly insulated piston-cylinder device. If the locks holding the piston in place are removed, the piston moves upwards to a stopper. The temperature of the air ______________.



- increases
- remains the same
- decreases
- need more information

Please explain your answer in the box below.

Submit

Figure 1. *Expanding Piston* question as presented in the treatment condition of Cohort A. Students are not committed to a particular answer when writing an explanation. Rather, they can change their answer selection as they write.

correctly as 'conceptual expertise' and assert 'novices often have too little practice of the right kind to develop conceptual understanding' (p. 357). Increasingly, instructors of large, introductory courses are reforming their instructional designs to attend to the development of conceptual understanding. To provide students the opportunity for such development, instructors often have students engage in answering concept questions during class using active learning pedagogies (Crouch & Mazur, 2001; Knight & Wood, 2005; McConnell et al., 2006). These pedagogies are often supported by technology such as clickers (Kay & LeSage, 2009).

Peer instruction as a stimulus for interactions to develop conceptual understanding

We specifically focus on Mazur's (1997) peer instruction pedagogy, arguably the most well-known and widely used technology-mediated active learning pedagogy in post-secondary science classes (Borrego, Cutler, Prince, Henderson, & Froyd, 2013; Crouch, Watkins, Fagen, & Mazur, 2007). However, the ideas are general and apply to other active pedagogies as well.

Mazur's pedagogy, peer instruction, consists of a structured instructional practice and is described as follows. First, a concept question is presented to the class. Students answer the question individually. Vahey, Tatar, and Roschelle (2007) would characterize this individual response as a *private interaction*. They define private interactions as 'those interactions in which students can engage with their materials and sense-making processes individually in a focused way' (p. 189). The next step in peer instruction encourages students to discuss the answer choices in small groups. These group discussions could be considered *public interactions*, defined as 'those interactions in which students engage in active or implicit discourse while they are simultaneously engaged with, or talking about, the product or materials of their work' (Vahey et al., 2007, p. 190). Next, students individually submit a final answer to a second poll of the same concept question (another private interaction). Finally, this sequence may be followed by a class-wide discussion when needed.

Most research on the use of concept questions in the active learning classroom has focused on their use in the peer instruction pedagogy. Studies of student performance during peer instruction show that students' frequency of correct answers usually increases after group discussion (Lasry, Charles, Whittaker, & Lautman, 2009; Smith et al., 2011; Van Dijk, Van Der Berg, & Van Keulen, 2001). Researchers generally link the improved performance to conceptual learning and attribute it to the public interactions during the group discussion between the individual student responses. For example, groups are reported to demonstrate what Singh identifies as 'co-construction of knowledge' where students respond correctly after discussion even when none of the students in the group had the correct answer initially (Singh, 2005; Smith et al., 2009). Alternatively, Hammer, Elby, Scherr, and Redish (2005) present a resource-based framework where the ability to correctly answer a conceptual question reflects a cognitive state that involves activating multiple resources. When viewed from this perspective, group discussion can provide stimulation to activate resources through public interactions. In this study, we seek to determine if private interactions also can help students activate resources.

While public interactions during peer instruction have been well characterized, less attention has been given to the influences of the private interactions. In their review of the peer instruction literature, Vickrey et al. (2015) recommend that it is important to elicit students' individual commitments to an answer before peer discussion. Nielsen, Hansen, and Stav (2014) relate the success of the peer interactions discussed above to the initial thinking period and suggest that deeper individual thinking increases students' time spent on argumentation and improves the quality of discussion. Lucas (2009) argues that prompting students to first write individual explanations lessens the likelihood that high-status students will dominate group discussion. In this article, we seek to contribute to this conversation by examining the influence of soliciting individual written explanations to multiple-choice concept questions (a private interaction) on performance and reasoning. While we take the perspective of writing as a private interaction, we acknowledge that it is not exclusively private if the writer has a specific audience in mind.

Writing as a focusing phenomenon to direct interactions

Our study explores if individual written explanations promote students' thinking 'in a focused way.' In particular, we are interested in the 'sense-making processes' of students. In answering the concept questions explored in this study (e.g. Figure 1), we hypothesize that there is a private interaction between what is inside of a student's mind, the answer choices available, and the question statement. That interaction is different and richer when the student must also construct an explanation in words on the computer screen. On a coarse level, we can imagine that when students answer a multiple-choice question, they may select their answer choice based on a guess, intuition, experience, or explicit scientific reasoning (Heckler & Scaife, 2014; Kahneman, 2003). When asked to concurrently provide a written explanation, they will 'focus' on more explicit use of scientific reasoning, as that type of explanation would typically align with their conception of what the instructor values. On a fine level, we can imagine that the act of writing enables students to more fully develop reasoned arguments (Renkl, 2015; Renkl, Stark, Gruber, & Mandl, 1998). Through elaborating their thoughts in the process of writing, explaining, and constructing a logical argument in defense of an answer, it becomes more likely that student will identify and correct flaws in their logic.

Our hypothesis is supported by learning scientists who have argued for decades that writing enhances thinking (Applebee, 1984; Gere & Abbott, 1985; McDermott & Hand, 2010; Odell, 1980; Sesenbaugh, 1989). The scope of writing in our study is similar to VanOrden (1987, 1990) who gave chemistry students short writing assignments that required the students to first solve a chemistry problem and then further explain and discuss their results in writing. She suggests that students should 'be able to do more than recognize or recall information' and hypothesized that the writing helps students to: '(1) integrate concepts, (2) apply the integrated concepts to real life, and (3) communicate those concepts' (1990, p. 583). Similarly, a recent study compared the effectiveness of open individual writing to group discussion as a synthesis activity toward conceptual learning in an active learning classroom, and found that students asked to write individually performed better on exams (Linton, Pangle, Wyatt, Powell, & Sherwood, 2014).

A few studies have used students' short written explanations to examine the reasoning behind their multiple-choice answer selections (Chandrasegaran, Treagust, & Mocerino, 2007; Tamir, 1989, 1990; Xie & Lee, 2012; Yarroch, 1991). However, the analysis of the explanations is typically directed toward assessment, either assessing the written explanations themselves, or, more commonly, using the explanations in the process of item development for summative assessment (Nelson, Geist, Miller, Streveler, & Olds, 2007; Tamir, 1990). There has been little reported on the role of writing explanations to multiple-choice questions from the perspective of engaging student thinking and learning.

An overarching teaching strategy that emphasizes the relation between thinking and writing by prompting students to create 'texts that explore the relationships among ideas' has been termed writing-to-learn (WTL) (Klein, 1999). WTL generally takes a 'constructivist' perspective (Klein, 1999; Rowell, 1997) that emphasizes that knowledge is *actively* constructed. From this perspective, students learn new knowledge by actively building upon and integrating prior knowledge. Some even argue that writing is inherently more about understanding than communication (Howard & Barton, 1988). Writing has been hypothesized to help promote thinking by prompting students to take multiple perspectives (Tierney, Soter, O'Flahavan, & McGinley, 1989) and encouraging conceptual reprocessing (Scardamalia & Bereiter, 1986) both of which help build understanding by reinforcing connections between related concepts (McDermott & Hand, 2010). In other words, the writer cannot communicate in the absence of organizing their thinking.

Methods

Our quasi-experimental study empirically investigates students' in-class responses to concept questions intended to elicit demonstration of and development of conceptual understanding. The first part of the study uses a crossover design to answer research question 1 by comparing answer selections of students who are prompted to provide written explanations (treatment) with those who are not (comparison). This design allows us to determine if written explanations influence students' private interactions through differences in the answer choices in comparison to a control group. The premise is that as students think about the concept questions more fully, they better learn the concepts and practices of science and engineering. We expect students' in-class responses to concept questions including written explanations to be correct more frequently compared to students' in-class responses to concept questions without written explanations. In the second part of the study, we use stratified sampling to analyze the written content and consistency with answers of five concept questions to address research question 2. We use an emergent coding process to relate the quality of the explanation to the correctness of the multiple-choice answer selection.

Participants and setting

This study is based on the data obtained from two cohorts (labeled Cohort A for year 1 of the study and Cohort B for year 2) enrolled in a required, sophomore-level, undergraduate thermodynamics course at a large public university. The course is titled 'energy balances,' a term that describes the methods engineers use to apply the first law of thermodynamics (conservation of energy) in their solutions to process-related problems. Energy balances is required for students majoring in chemical, biological, and environmental engineering. A total of 302 students (138 in Cohort A and 164 in Cohort B) participated in the study.

The students from each cohort attended a common lecture and self-selected into one of two weekly, one-hour discussion sections (labeled Section 1 and Section 2). The lectures and sections for both cohorts were taught by the same instructor. For each cohort, the second section (Section 2) was scheduled immediately after the first (Section 1) and held in the same room. The lecture and section rooms were consistent across both cohorts as well.

Table 1 shows the demographics according to cohort and discussion section as self-reported by the participants who volunteered this information (65% response rate). The greater number of Asian students in Cohort B is the result of a university recruitment initiative. The results of this study should be interpreted with these demographic differences in mind. The Institutional Review Board approved the research and participants signed informed consent forms.

Concept questions were posed to students during the discussion section as part of peer instruction pedagogy. The length of time available to answer questions was determined as follows. Once approximately 60% of the students had answered, the instructor typically gave a verbal warning of '30 seconds' and a verbal countdown for the last several seconds. The time spent on each concept question with the treatment condition, the written explanation prompt, is longer than is typical (Brooks, Demaree, and Koretsky, 2016; Mazur, 1997), largely because students are asked to provide short written explanations of why they selected a particular answer choice. All students in the course received full credit toward their final grade for submitting an answer (5% of course total points), and received extra-credit for each correct answer submitted (additional 5% of course total). This low-stakes approach was intended to encourage students' conceptual thinking processes while still providing incentive for students to respond correctly (James, 2006; Willoughby & Gustafson, 2009).

Over the term, the students in each cohort completed three midterm exams and a final exam. The exams were written by the course instructor, and all the students within a cohort took the same exams at the same time. They contained both quantitative problems and concept questions. The quantitative problems asked students to demonstrate procedural knowledge of energy balances, had a correct numerical answer, and were similar to those typically asked in this type of engineering course. The midterm exams also contained three concept questions that were worth 21% and 30% of the total points for Cohorts A and B, respectively. Three concept questions were also asked on the final exam and were worth 10% and 18% of the total points for Cohorts A and B,

Table 1. Self-reported demographic data by cohort and section; there was a 65% response rate from study participants.

		Cohort A		Cohort B	
		Section 1	Section 2	Section 1	Section 2
Major	Chemical	68%	67%	73%	57%
	Biological	18%	18%	18%	23%
	Environmental	13%	16%	9%	20%
	N reporting	38	51	45	61
Gender	Male	76%	61%	62%	69%
	Female	24%	39%	38%	31%
	N reporting	38	51	47	61
Race	White – Non-Hispanic	74%	70%	38%	48%
	Asian/Pacific Islander	11%	17%	49%	20%
	Other	13%	9%	7%	18%
	White – Hispanic	3%	4%	7%	14%
	N reporting	38	46	45	56

respectively. Cohort B received more credit for the concept questions since they were required to provide written explanations on the exams while Cohort A was not. Additionally, Cohort B was asked the concept questions at the beginning of the exam while the concept questions for Cohort A appeared at the end.

Study design

Data for this study include (1) multiple-choice responses for the concept questions (treatment and comparison); (2) exam results and overall course grades (all students); and (3) written explanations for five concept questions (treatment).

Influence of written explanations on multiple-choice responses

We quantitatively analyzed the influence of written explanations on student answers to multiple-choice questions. To minimize the effect of potential differences between sections, a crossover study was conducted. The design matrix can be seen in Table 2. Each section alternated between participation in the comparison condition (no written explanation) and in the treatment condition (with written explanation) in two- to three-week intervals during the 10-week term. For the first three weeks of class, Section 2 from each cohort was placed in the treatment condition and the computer system displayed a prompt and space for students to explain their answer choices. Section 1 was the comparison and was not provided such a prompt. In the fourth week, the treatment and comparison conditions were switched, as well as in weeks 7 and 9.

The student responses to concept questions that were collected provide the core of the data for this study. Concept questions, including the multiple-choice answers and written explanations, were answered using the *Concept Warehouse* (Koretsky et al., 2014). Students were asked to first answer individually on their laptops, smartphones, or tablets, without consulting their neighbors or the instructor. Figure 1 depicts the question *Expanding Piston* as it was presented in the treatment condition for Cohort A. For the comparison condition, the same question was asked but without the text input box or prompt that says, 'Please explain your answer in the box below.' Since students were *not* committed to a particular multiple-choice answer selection when the writing prompt was available, they could freely change their multiple-choice selection if the process of writing an explanation changed which answer they thought might be correct. Hypothetically, a student could select an answer, write an explanation defending it, notice a flaw in her/his argument,

Table 2. Crossover experimental design matrix listing treatment and comparison conditions for each discussion section in each cohort by week number.

Cohort	Week	Section 1	Section 2	Questions
A (Year 1)	1–3	Comparison	Treatment	1–25
	4–6	Treatment	Comparison	
	7–8	Comparison	Treatment	
	9–10	Treatment	Comparison	
B (Year 2)	1–3	Comparison	Treatment	9–39
	4–6	Treatment	Comparison	
	7–8	Comparison	Treatment	
	9–10	Treatment	Comparison	

Note: The treatment group received a prompt to explain their answer in writing while the comparison group did not receive that prompt.

and correct the explanation and the answer. Alternatively, a student could start by writing an explanation and then select an answer choice based on it. To receive credit, students in the treatment condition were required to select an answer choice but writing an explanation was optional. In the treatment condition, approximately 80% of the student responses contained some type of written explanation (Cohort A, $\bar{x} = 80.9\%$, SD = 7.8%; Cohort B, $\bar{x} = 79.2\%$, SD = 7.2%).

In the two-year study, a set of 58 unique questions was delivered to both treatment and comparison conditions. This number does not count questions administered near the end of the class and only answered by students in one of the two sections. The questions were selected or written by the course instructor with the criteria of (i) high-quality questions following guidelines from Taylor and Nolen (2007) and (ii) appropriate alignment with the week's learning objectives. The responses analyzed in this study are from a subset of the questions asked and were selected as follows. First, only the initial student responses to the concept questions were analyzed; we excluded responses to the second poll used as part of peer instruction. Second, each question was reviewed for clarity in two ways. One author evaluated the questions for ambiguous language. Separately, another author read through written responses to identify questions that were clearly misinterpreted by more than one student. In all cases, the misinterpreted questions matched the questions independently identified as having ambiguous language. Ten questions were removed from the set using the ambiguous language criterion. Finally, the instructor asked a set of nine questions that directed students to select valid equations for situations in the question statement. Since such questions do not match our classification of a 'concept question,' they were eliminated. Of the remaining pool of 39 individual concept questions, 21 were written by the instructor and 18 were peer-reviewed questions that contributed to the *Concept Warehouse* by others. Cohorts A and B were asked 25 and 31 of the 39 concept questions, respectively. Seventeen questions were used with both cohorts. Response distributions to the 17 overlapping questions generally aligned (Cohort A, $\bar{x} = 63\%$, SD = 17%; Cohort B, $\bar{x} = 60\%$, SD = 17%).

To determine statistical significance of difference in answer choices between treatment and comparison, we used a non-parametric Chi-squared test with an alpha level of 0.05. Chi-squared is appropriate because it can compare two distributions and is not dependent upon an assumption of normality. One-way analysis of variance (ANOVA) is used to compare differences in exam performance between discussion sections within the same cohort.

Content of written explanations

In the second part of the study, written responses to five concept questions were qualitatively analyzed. The questions we analyzed were selected based on stratified sampling. First, we based selection on the results of the crossover study described above. We sought to compare questions that showed a statistically significant positive treatment effect where students in the group that provided written explanations were more likely to answer the multiple-choice question correctly to those questions that have a significantly negative effect. For each case, we intended to identify one easy question, $\bar{x} > 75\%$, one moderate question, $75\% > \bar{x} > 50\%$, and one difficult question, $\bar{x} < 50\%$. However, there were no easy questions that showed a significantly negative effect, resulting in five questions.

We analyzed the written explanations for the treatment condition for these five questions using open coding, a process used to infer categories of meaning. We use a technique similar to that of Newcomer and Steif (2008) in their analysis of written explanations to a concept question in statics. The process involves proposing a code, coding individually, comparing among the coders, modifying the code, and repeating until convergence. Three researchers participated in this process. All three have domain knowledge and had taken either this course or a similar course at another institution. A hierarchical coding scheme was created that incorporates and ranks the observed concepts and misconceptions. Similar to Tamir (1990), the codes ascend from 1 (poorly reasoned) to 4 (well reasoned) with a higher code indicating more appropriate reasoning in the explanation. The hierarchy emerges from the nature of the content and specific student responses to these particular conceptually difficult problems and each code-set is therefore question-specific. For the coding process, only the written explanation itself is considered irrespective of which multiple-choice answer the student selected. While it might be expected that higher code values would correspond to correct answer selections, that is not always the case (Koretsky, Brooks, White, & Bowen, 2016). Code descriptions and sample explanations for the concept question shown in Figure 1 are provided in Table 3.

Once the code categories were developed, two researchers initially coded a subset of the written explanations. Explanations with disagreement between the coders were discussed with the research team until the team reached consensus on the appropriate code for the given explanation. The two researchers then coded all the responses and obtained an inter-rater reliability using the Cohen's κ statistic of 0.88. Approximately 3% of the explanations per coded question had a discrepancy of more than one code category. The highest code rating of 4 was used as the criterion that students had identified the salient concepts and used appropriate reasoning in their explanations.

Results

Influence of written explanations on multiple-choice responses

Figure 2 presents two histograms that illustrate the variation in the correct responses to the concept questions analyzed in this study. Figure 2(a) bins the correct answer percentage

Table 3. Code values, descriptions, and sample student explanations for the *Expanding Piston* concept question.

Code	Level description	Sample student explanation
4	Conservation of energy, including a term for work. The explanation accounts for the work of the gas in the system on the surroundings as the system boundary expands.	'There is no Kinetic Energy or Heat transfer so the only source of energy to move the piston is the air's internal energy. A decrease in internal energy correlates to a decrease in temperature.'
3	No work term is identified. The student recognizes that an energy balance can be applied and identifies that heat is zero, but neglects work.	'Since the device is perfectly insulated, there is no heat transfer between the system and surroundings. This means that the temperature is not changing since there is no heat entering into the system.'
2	Ideal gas law only; the response solely uses the ideal gas law to explain how temperature changes with no mention of any type of energy related term	'if the pressure decreases and the volume increases then there will be no change in the temperature.'
1	Other incorrect reasoning; incorrect reasons that cannot be assigned to either codes '2' or '3' above.	'The temperature increase made the air molecule move faster upward, then push the piston goes up.'

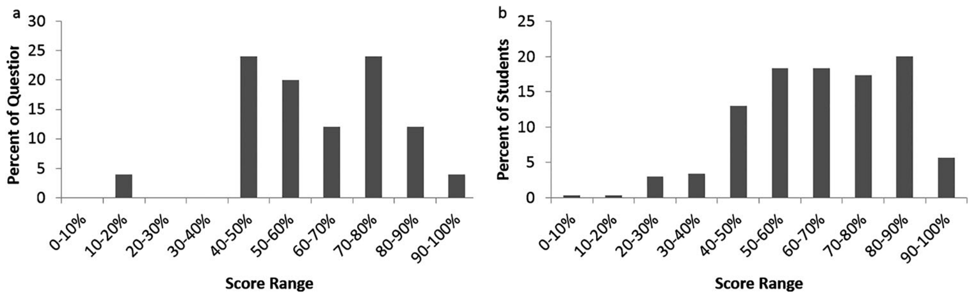


Figure 2. Histograms binned by aggregate correct answer percentage of (a) the percent of questions and (b) the percent of students.

for each of the 39 questions averaged over all of the student participants. There is clearly a wide distribution in the question difficulty with a large fraction ranging from 40% correct to 90% correct. Figure 2(b) bins the correct answer percentage for the 302 students averaged over the questions that they attempted. Again, there is a wide range in performance. A large number of student scores appear across five deciles with some students answering the majority of questions correctly (77 students had scores greater than 80%) and other students struggling (60 students less than 50%). Thus, the 39 selected questions in this study represent a wide range of difficulty and are representative of the distribution of questions that might be used for peer instruction.

Table 4 provides a summary of the cumulative results from the multiple-choice responses study. The results are divided into three rows based on the statistical significance of the difference between the treatment and comparison conditions, using a criterion of $p < .05$ from the Chi-squared test. The top row, labeled 'Tr > Com,' shows the case for questions where students in the treatment condition chose the correct answer significantly more often than the students in the comparison. The case for questions where differences in selection of answer choices between treatment and comparison were not found to be statistically different is reported in the next row and labeled 'Tr = Com.' Finally, questions where the students in the treatment chose the correct answer significantly less often than the students in the comparison are shown in the bottom row of each set and labeled 'Tr < Com.' Data reported in this table include the number of questions for each of the three cases (*n* questions), the corresponding average percent correct for the treatment (Treatment <%>) and comparison group (Comparison <%>), and the average number of students in each group (Treatment <*n* students> and Comparison <*n* students>).

The cumulative data show that students in the treatment condition more frequently selected the correct answer to the multiple-choice question than those in the comparison

Table 4. Summary of cumulative results on multiple-choice performance to the 39 concept questions.

Effect	*n* questions	Treatment <%>	Comparison <%>	Treatment <*n* students>	Comparison <*n* students>
Tr > Com	15	65%	50%	89	92
Tr = Com	19	58%	57%	104	106
Tr < Com	5	56%	69%	83	84

Notes: The rows Tr > Com and Tr < Com indicate the questions where students in the treatment condition scored statistically significantly better or worse, respectively, than the comparison condition, to an alpha of 0.05 according to the Chi Square test. For the row Tr = Com, there was no statistically significant difference.

condition; there is a statistically significant increase for 15 of the 39 questions (with our selected alpha level pure random chance would result in around 2 out of 39). However, surprisingly, students in the treatment condition also answer incorrectly more often for five questions. Overall, the cumulative results suggest that soliciting a written explanation will generally increase the likelihood that students will answer correctly. The results for the 17 overlapping questions delivered to both cohorts are generally consistent with the cumulative results in Table 4.

Since we employed a crossover design *within each cohort*, it is possible that the difference in scores is due to different ability, on average, of students in one section within a cohort as compared to the other section. In order to investigate this alternative explanation, we examined the results from the four exams as well as the total course score. Table 5 provides a summary of the student performance for each section (1 and 2) of each cohort (A and B). One-way ANOVA shows no statistically significant difference between sections within a cohort. We also analyzed individual concept questions from the exams for difference in performance between sections within a cohort. For the exams, the questions were asked to each section in the same way. Only 1 out of 18 questions (6%) showed a statistically significant difference between sections, a proportion much lower than shown in Table 4. Therefore, we can interpret those data from Table 4 to likely result from the treatment and the question characteristics rather than differences in performance ability of students.

Content of written explanations

Table 6 provides coding results for easy, moderate, and difficult concept questions. For each question, the number of written responses (N), number of correct multiple-choice responses ($N_{correct}$), and number of explanations with the highest code value of 4 ($N_{code\,=\,4}$) are shown. The rows Tr > Com and Tr < Com indicate the questions where students in the treatment condition scored statistically significantly better or worse, respectively, than the comparison condition. The number of correct answers always exceeds the number of explanations with a code value of 4. Thus, a significant proportion of students who obtained the correct answer were not able to use normative scientific reasoning to justify their answer. This proportion tended to increase with question difficulty and with the questions that showed the anomalous treatment effect (Tr < Com).

Table 5. Average ($\bar{x}$) and standard deviation (SD) of student performance for the four exams and the total course for each section (1 and 2) in each cohort (A and B).

		Exam 1		Exam 2		Exam 3		Final		Total course score	
Cohort	Section	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
A	1	73	12	70	16	78	14	70	19	82	10
	2	71	14	69	19	77	14	71	20	83	10
ANOVA		F	0.51	F	0.09	F	0.19	F	0.05	F	0.19
		p	0.48	p	0.76	p	0.67	p	0.82	p	0.66
B	1	69	14	68	20	85	20	71	20	81	14
	2	70	12	71	15	89	13	74	14	83	9
ANOVA		F	0.23	F	1.16	F	1.80	F	0.70	F	1.70
		p	0.63	p	0.28	p	0.18	p	0.41	p	0.19

Table 6. Coding results for easy, moderate, and difficult concept questions.

Effect		Easy $\bar{x} > 75\%$	Moderate $75\% > \bar{x} > 50\%$	Difficult $\bar{x} < 50\%$
Tr > Com	N (% correct)	87 (85%)	60 (65%)	77 (23%)
	$N_{correct}$	72	39	18
	$N_{code=4}$	53	21	7
Tr < Com	N (% correct)	NA	79 (59%)	71 (22%)
	$N_{correct}$	NA	47	16
	$N_{code=4}$	NA	11	1

Note: For each question, the number of written responses (N), number of correct multiple-choice responses ($N_{correct}$) and number with the highest code value of 4 ($N_{code=4}$) are shown.

For example, the concept question shown in Figure 1 whose code values are reported in Table 3 is the moderate question (65% correct) in the Tr > Com effect row. Rather than considering an energy balance, a set of students focused exclusively on the ideal gas law, for example, 'because the lower the pressure the air is in the lower the temperature,' or 'Temperature must go down to maintain PV = nRT relationship.' For most questions, a set of students serendipitously chose the correct answer through clearly faulty reasoning.

Discussion

Overall, the results from the first part of the study show that soliciting a written explanation to concept questions in introductory thermodynamics can have a significant influence on answer choice and when it does, students are more likely to answer correctly (research question 1). The distribution of multiple-choice answers to approximately half of the concept questions in this study is significantly different when students were prompted to write an explanation than when they were not (see Table 4). We contrast these results with the overall performance on exams (Table 5) and to responses to individual concept questions on exams. During exams, both sections are asked questions in the same way. When the discussion sections are presented with a different prompt, as in the treatment *vs.* comparison conditions, they often respond differently. On the other hand, the performances between sections on exams are remarkably similar. Such contrast provides support that the differences in response of the treatment section are due to the prompt for a written justification of the question and not due to inherent differences in the composition of students in each section.

The 'focused' nature of private interactions provides a frame to consider how students are encouraged to use more explicit scientific reasoning and can explain the positive influence of writing on answer choice. Researchers in chemistry have found that students often respond to concept questions by using shortcut strategies that circumvent activation of domain-specific knowledge (Graulich, 2015; Talanquer, 2014). In contrast, when students are asked for written justification in our study, they are encouraged to frame the written responses in terms of domain-specific knowledge to the degree they believe it is expected by the instructor. More generally, the results from our study are consistent with studies in cognitive neuroscience that show that a prompt for a confidence judgment decreases the likelihood of guessing (Voss & Paller, 2010) and with the WTL literature that suggests that writing encourages the restructuring and reorganization of knowledge (Klein, 1999; Rivard, 1994). These results are also consistent with the view of Chi, De Leeuw, Chiu, and LaVancher (1994) who indicate that the active process of explaining encourages students to integrate new knowledge with existing knowledge and leads to richer conceptual understanding. By extension, a richer conceptual understanding should lead to more frequent correct answer choice selections.

Students' written explanations also provided insight into how they conceptualize content. For the five questions that were coded, a proportion of students who selected the correct answer used faulty or incomplete reasoning to justify their choice (research question 2). Written explanations allow the instructor to examine explicit student reasoning and potentially identify questions where students apply inappropriate reasoning to get the correct answer. With such understanding, the instructor can directly address the faulty reasoning and help students make connections to normative scientific reasoning. The recent development of methods such as lexical analysis (Haudek, Prevost, Moscarella, Merrill, & Urban-Lurain, 2012; Urban-Lurain et al., 2013) and computer tools such as the *Concept Warehouse* (Koretsky et al., 2014) can help facilitate the ability for instructors to identify student reasoning processes in their writing without requiring overwhelming effort. However, development work is needed to refine these methods and tools so they can readily provide practical information to instructors.

Limitations and future research

There are several limitations to this study that should be kept in mind when considering the results and implications. The study examined student responses to concept questions delivered in an active learning environment 'in the wild,' with a quasi-experimental design and associated issues of real-time data collection in a live class. Additionally, the results are limited to one course, context, and instructor. A randomized control experiment amongst several populations is needed for generalization. In the treatment condition, only 80% of the student responses contained some type of written explanation, which limits the resolution of the study. A clinical setting with 100% response would be desirable. Additionally, 19 questions did not display a significant difference. It would be useful to determine if this result would change with larger sample sizes (i.e. if it is based on statistical power or if there is an inherent difference in the nature of these concept questions). It would be useful to validate this approach by comparing student explanations to analogous, open free response short answer questions without multiple choice. While the coding process used four levels to evaluate student responses, each set of codes depended on the specific content of the question; the cross-question comparisons shown in Table 6 should be interpreted with this limitation in mind. Finally, time on task needs to be considered. While students in both the treatment and comparison conditions were given a 30-second warning after 60% had completed the task (essentially those 60% were given all of the time they wanted), the time allocated between the two conditions differed by an average of approximately 1 minute, which could account for some differences in performance.

Conclusions

This study examined changes in student answer selection that appear to result from soliciting reflective written explanations to multiple-choice concept questions. In the context of the data presented, students tended to pick the correct answer more frequently when provided with a prompt for a written explanation. However, there were five questions where the students prompted for an explanation were less likely to choose the correct answer. We studied selected questions from each case using qualitative methods. Students who answer correctly do not always provide a scientifically normative explanation, and it appears that they have a harder time properly explaining questions in this latter set.

The results suggest that asking students for written explanations helps their thinking and learning, and we encourage instructors to solicit written explanations when they use multiple-choice concept questions; however, the results are preliminary and more research is needed. The results from this study also suggest that instructors should be aware that with multiple-choice concept questions, students may get the correct answer for an incorrect reason. Instructors using concept questions are encouraged to look for cases where their students choose the correct answer using incorrect reasoning, perhaps by simply sampling the explanations from the students who selected the correct answer.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors acknowledge support from the National Science Foundation's Course, Curriculum and Laboratory Improvement Program [grant number NSF 1023099]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Notes on Contributors

Milo D. Koretsky is a professor in the School of Chemical, Biological, and Environmental Engineering at Oregon State University. He received his B.S. and M.S. degrees from UC San Diego and his Ph.D. from UC Berkeley, all in chemical engineering. He is interested in integrating technology into effective educational practices and in promoting the use of higher-level cognitive skills in engineering problem solving. His research interests particularly focus on what prevents students from being able to integrate and extend the knowledge developed in specific courses in the core curriculum to the more complex, authentic problems and projects they face as professionals.

Bill J. Brooks is a postdoctoral scholar in the School of Chemical, Biological, and Environmental Engineering at Oregon State University. He received his B.S. and his Ph.D. from Oregon State University, both in chemical engineering. His Ph.D. research involved the use of written explanations to concept questions to investigate technology mediated active learning in the undergraduate chemical engineering classroom. He is interested in using technology to enhance educational practices in promoting conceptual understanding. He is the primary programmer of the Concept Warehouse and his current focus is on its continued development, specifically creating and integrating Interactive Virtual Labs.

Adam Z. Higgins is an associate professor in the School of Chemical, Biological, and Environmental Engineering at Oregon State University. He received his B.S. degree in Bioengineering and his BA in International Studies from Oregon State University and his Ph.D.in Bioengineering from the Georgia Institute of Technology. Research in the Higgins laboratory fits within the broad theme of biomedical process engineering, with a particular emphasis on processes that involve bio-transport phenomena.

References

Applebee, A. N. (1984). Writing and reasoning. *Review of Educational Research*, *54*(4), 577–596.

Beatty, I. D., Gerace, W. J., Leonard, W. J., & Dufresne, R. J. (2006). Designing effective questions for classroom response system teaching. *American Journal of Physics*, *74*(1), 31–39. doi:10.1119/1.2121753

Borrego, M., Cutler, S., Prince, M., Henderson, C., & Froyd, J. E. (2013). Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses. *Journal of Engineering Education*, *102*(3), 394–425. doi:10.1002/jee.20020

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brooks, B. J., Demaree, D. N., and Koretsky, M. D. (2016). Student response times to in-class thermodynamics concept questions. *International Journal of Engineering Education*, *32*(1), 30–38.

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, *6*(1), 9–20. doi:10.1187/cbe.06-12-0205

Chandrasegaran, A., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, *8*(3), 293–307. doi:10.1039/B7RP90006F

Chi, M. T., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. doi:10.1207/s15516709cog1803_3

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*(2), 121–152.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977. doi:10.1119/1.1374249

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. In E. F. Redish and P. Cooney (Eds.), *Research-based reform of University Physics* (Vol. 1. Issue no. (1), pp. 40–95). College Park, MD: American Association of Physics Teachers. doi:10.1.1.114.9397

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(6031), 862–864. doi:10.1126/science.1201783

Duncan, D. (2006). Clickers: A new teaching aid with exceptional promise. *Astronomy Education Review*, *5*(1), 70–88.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M.P. (2014) Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111* (23), 8410–8415. doi:10.1073/pnas.1319030111

Gere, A. R., & Abbott, R. D. (1985). Talking about writing: The language of writing groups. *Research in the Teaching of English*, *19*(4), 362–385.

Graulich, N. (2015). Intuitive judgments govern students' answering patterns in multiple-choice exercises in organic chemistry. *Journal of Chemical Education*, *92*(2), 205–211. doi:10.1021/ed500641n

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213–1216. doi:10.1126/science.1204820

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. doi:10.1119/1.18809

Haláková, Z., & Prokša, M. (2007). Two kinds of conceptual problems in chemistry teaching. *Journal of Chemical Education*, *84*(1), 172–174. doi:10.1021/ed084p172

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of learning: Research and perspective* (pp. 89–120). Greenwich: Information Age Publishing.

Haudek, K. C., Prevost, L. B., Moscarella, R. A., Merrill, J., & Urban-Lurain, M. (2012). What are they thinking? Automated analysis of student writing about acid–base chemistry in introductory biology. *CBE-Life Sciences Education*, *11*(3), 283–293. doi:10.1187/cbe.11-08-0084

Heckler, A. F., & Scaife, T. M. (2014). Patterns of response times and response choices to science questions: The influence of relative processing time. *Cognitive Science*, *39*(3), 1–42. doi:10.1111/cogs.12166

Howard, V., & Barton, J. H. (1988). *Thinking on paper*. New York, NY: William Morrow Paperbacks.

James, M. C. (2006). The effect of grading incentive on student discourse in peer instruction. *American Journal of Physics*, *74*(8), 689–691. doi:10.1119/1.2198887

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720. doi:10.1037/0003-066X.58.9.697

Kay, R. H., & LeSage, A. (2009). Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*, *53*(3), 819–827. doi:10.1016/j.compedu.2009.05.001

Klein, P. D. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, *11*(3), 203–270.

Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298–310. doi:10.1187/05-06-0082

Koretsky, M. D., Brooks, B. J., White, R. M., & Bowen, A. S. (2016), Querying the questions: Student responses and reasoning in an active learning class. *Journal of Engineering Education*, *105*(2), 219–244. doi:10.1002/jee.20116

Koretsky, M. D., Falconer, J. L., Brooks, B. J., Gilbuena, D. M., Silverstein, D. L., Smith, C., & Miletic, M. (2014). The AiChE concept warehouse: A web-based tool to promote concept-based instruction. *Advances in Engineering Education*, *4*(1).

Koretsky, M. D., Miller, R. L., Falconer, J. L., Prince, M. J., Vigeant, M. A., Krause, S. J., & Siverstein, D. L. (2011). *Educational methods and tools to encourage conceptual learning*. Proceedings of the 2011 American Society for Engineering Education Annual Conference & Exposition. Vancouver, BC.

Lasry, N., Charles, E., Whittaker, C., & Lautman, M. (2009). *When talking is better than staying quiet*. Paper presented at the AIP conference proceedings, Ann Arbor, MI.

Linton, D. L., Pangle, W. M., Wyatt, K. H., Powell, K. N., & Sherwood, R. E. (2014). Identifying key features of effective active learning: The effects of writing and peer discussion. *CBE-Life Sciences Education*, *13*(3), 469–477.

Lucas, A. (2009). Using peer instruction and i-clickers to enhance student participation in calculus. *Primus*, *19*(3), 219–231. doi:10.1187/cbe.13-12-0242

MacArthur, J. R., & Jones, L. L. (2008). A review of literature reports of clickers applicable to college chemistry classrooms. *Chemistry Education Research and Practice*, *9*(3), 187–195. doi:10.1039/B812407H

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., … McGrew, H. (2006). Using ConcepTests to assess and improve student conceptual understanding in introductory geoscience courses. *Journal of Geoscience Education*, *54*(1), 61–68.

McDermott, L. C. (2001). Oersted Medal Lecture 2001: 'Physics education research – The key to student learning'. *American Journal of Physics*, *69*(11), 1127–1137. doi:10.1119/1.1389280

McDermott, M. A., & Hand, B. (2010). A secondary reanalysis of student perceptions of non-traditional writing tasks over a ten year period. *Journal of Research in Science Teaching*, *47*(5), 518–539. doi:10.1002/tea.20350

National Research Council. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: National Academies Press.

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. (S.R. Singer, N.R. Nielsen, & H.A. Schweingruber, Eds.). Washington, DC: National Academies Press.

Nelson, M. A., Geist, M. R., Miller, R. L., Streveler, R. A., & Olds, B. M. (2007). *How to create a concept inventory: The thermal and transport concept inventory*. Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

Newcomer, J. L., & Steif, P. S. (2008). Student thinking about static equilibrium: Insights from written explanations to a concept question. *Journal of Engineering Education*, *97*(4), 481–490. doi:10.1002/j.2168-9830.2008.tb00994.x

Nielsen, K. L., Hansen, G., & Stav, J. B. (2014). How the initial thinking period affects student argumentation during peer instruction: Students' experiences versus observations. *Studies in Higher Education*, 1–15. doi:10.1080/03075079.2014.915300

Nurrenbern, S. C., & Robinson, W. R. (1998). Conceptual questions and challenge problems. *Journal of Chemical Education*, *75*(11), 1502–1503. doi:10.1021/ed075p1502

Odell, L. (1980). The process of writing and the process of learning. *College Composition and Communication*, *31*(1), 42–50.

Papaphotis, G., & Tsaparlis, G. (2008). Conceptual versus algorithmic learning in high school chemistry: The case of basic quantum chemical concepts. Part 1. Statistical analysis of a quantitative study. *Chemistry Education Research and Practice*, *9*(4), 323–331. doi:10.1039/B818468M

Renkl, A. (2015). Different roads lead to Rome: The case of principle-based cognitive skills. *Learning: Research and Practice*, *1*(1), 79–90. doi:10.1080/23735082.2015.994255

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. doi:10.1006/ceps.1997.0959

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175. doi:10.1037/0022-0663.91.1.175

Rivard, L. O. P. (1994). A review of writing to learn in science: Implications for practice and research. *Journal of Research in Science Teaching*, *31*(9), 969–983. doi:10.1002/tea.3660310910

Rowell, P. M. (1997). Learning in school science: The promises and practices of writing. *Studies in Science Education*, *30*(1), 19–56. doi:10.1080/03057269708560102

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. Wittrock (Ed.), *Handbook of research on teaching* (Vol. 3, pp. 778–803). New York, NY: Macmillan.

Sesenbaugh , R. (1989). ERIC/RCS: Writing across the curriculum: Evolving reform. *Journal of Reading*, *32*(5), 462–465.

Singh, C. (2005). Impact of peer interaction on conceptual test performance. *American Journal of Physics*, *73*(5), 446–451. doi:10.1119/1.1858450

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323* (5910), 122–124. doi:10.1126/science.1165919

Smith, M. K., Wood, W. B., Krauter, K., & Knight, J. K. (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. *CBE-Life Sciences Education*, *10*(1), 55–63. doi:10.1187/cbe.10-08-0101

Talanquer, V. (2014). Chemistry education: Ten heuristics to tame. *Journal of Chemical Education*, *91*(8), 1091–1097. doi:10.1021/ed4008765

Tamir, P. (1989). Training teachers to teach effectively in the laboratory. *Science Education*, *73*(1), 59–69. doi:10.1002/sce.3730730106

Tamir, P. (1990). Justifying the selection of answers in multiple choice items. *International Journal of Science Education*, *12*(5), 563–573. doi:10.1080/0950069900120508

Tanner, K., & Allen, D. (2005). Approaches to biology teaching and learning: Understanding the wrong answers – Teaching toward conceptual change. *Cell Biology Education*, *4*(2), 112–117.

Taylor, C., & Nolen, S. (2007). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson-Merrill-Prentice Hall.

Tierney, R. J., Soter, A., O'Flahavan, J. F., & McGinley, W. (1989). The effects of reading and writing upon thinking critically. *Reading Research Quarterly*, *24*(2), 134–173. doi:10.2307/747862

Urban-Lurain, M., Prevost, L., Haudek, K. C., Henry, E. N., Berry, M., & Merrill, J. E. (2013). Using computerized lexical analysis of student writing to support just-in-time teaching in large enrollment STEM courses. *Frontiers in Education*. doi:10.1109/FIE.2013.6685130

Vahey, P., Tatar, D., & Roschelle, J. (2007). Using handheld technology to move between private and public interactions in the classroom. In M. Hooft & K. Swan (Eds.), *Ubiquitous computing in education: Invisible technology, visible impact* (pp. 187–210). Mahwah, NJ: Erlbaum.

Van Dijk, L., Van Der Berg, G., & Van Keulen, H. (2001). Interactive lectures in engineering education. *European Journal of Engineering Education*, *26*(1), 15–28. doi:10.1080/03043790123124

VanLehn, K., & Van de Sande, B. (2009). Expertise in elementary physics, and how to acquire it. In K. Ericsson (Ed.), *The development of professional performance: Toward measurement of expert performance and design of optimal learning environments* (pp. 356–378). New York, NY: Cambridge University Press.

VanOrden, N. (1987). Critical-thinking writing assignments in general chemistry. *Journal of Chemical Education*, *64*(6), 506–507. doi:10.1021/ed064p506

VanOrden, N. (1990). Is writing an effective way to learn chemical concepts? Classroom-based research. *Journal of Chemical Education*, *67*(7), 583–585. doi:10.1021/ed067p583

Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE-Life Sciences Education*, *14*(1), es3. doi:10.1187/cbe.14-11-0198

Voss, J. L., & Paller, K. A. (2010). What makes recognition without awareness appear to be elusive? Strategic factors that influence the accuracy of guesses. *Learning & Memory*, *17*(9), 460–468. doi:10.1101/lm.1896010

Willoughby, S. D., & Gustafson, E. (2009). Technology talks: Clickers and grading incentive in the large lecture hall. *American Journal of Physics*, *77*(2), 180–183. doi:10.1119/1.3013542

Wood, W. B. (2004). Clickers: A teaching gimmick that works. *Developmental Cell*, *7*(6), 796–798.

Xie, C., & Lee, H.-S. (2012). A visual approach to nanotechnology education. *International Journal of Engineering Education*, *28*(5), 1006–1018.

Yarroch, W. L. (1991). The implications of content versus item validity on science tests. *Journal of Research in Science Teaching*, *28*(7), 619–629. doi:10.1002/tea.3660280707