J Sci Educ Technol (2012) 21:392–402
DOI 10.1007/s10956-011-9332-1

Assessing Multimedia Influences on Student Responses Using a Personal Response System

Kyle Gray · Katharine Owens · Xin Liang · David Steer

Published online: 20 August 2011

Abstract To date, research to date on personal response systems (clickers) has focused on external issues pertaining to the implementation of this technology or broadly measured student learning gains rather than investigating differences in the responses themselves. Multimedia learning makes use of both words and pictures, and research from cognitive psychology suggests that using both words and illustrations improves student learning. This study analyzed student response data from 561 students taking an introductory earth science course to determine whether including an illustration in a clicker question resulted in a higher percentage of correct responses than questions that did not include a corresponding illustration. Questions on topics pertaining to the solid earth were categorized as *illustrated* questions if they contained a picture, or graph and *text-only* if the question only contained text. For each type of question, we calculated the percentage of correct responses for each student and compared the results to student ACT-reading, math, and science scores. A within-groups, repeated measures analysis of covariance with instructor as the covariate yielded no significant differences between the percentage of correct responses to either the text-only or the illustrated questions. Similar non-significant differences were obtained when students were grouped into quartiles according to their ACT-reading, -math, and -science scores. These results suggest that the way in which a conceptest question is written does not affect student responses and supports the claim that conceptest questions are a valid formative assessment tool.

Keywords Personal response systems · Clickers · Multimedia · Conceptest · ACT

Introduction

Over the past two decades, two avenues of research involving the use of computers in the classroom have yielded complimentary findings. Research from cognitive psychology demonstrates that humans most efficiently process audio-visual data when words and pictures are combined in a single presentation (Mayer 2009). During the same time period, wide acceptance of personal response systems (colloquially known as 'clickers') in undergraduate science courses has transformed lecture courses by providing a means of utilizing an active-learning pedagogy in an otherwise passive learning environment. Numerous studies demonstrate effective ways of using personal response systems in a variety of disciplines (Beatty and Gerace 2009; Crouch and Mazur 2001; McConnell et al. 2006) but very few studies have assessed whether student responses to the associated conceptually-based questions may be influenced by factors other than question content or student conceptual knowledge (King and Joshi 2008; Steer et al. 2009). Given that personal response systems use a

K. Gray (✉)
Department of Earth Science, University of Northern Iowa, Latham Hall, Room 124, Cedar Falls, IA 50614-0335, USA
e-mail: kyle.gray@uni.edu

K. Owens
Department of Curricular and Instructional Studies, University of Akron, Akron, OH, USA

X. Liang
Department of Educational Foundations and Leadership, University of Akron, Akron, OH, USA

D. Steer
Department of Geology and Environmental Science, University of Akron, Akron, OH, USA

multimedia format, it is possible that student responses are influenced by design principles related to developing a good multimedia presentation. This study addresses whether the way in which a question is written influences student responses. It also assesses whether differences in student prior achievement influence their answers to conceptually-based questions used in conjunction with a personal response system.

Research on Conceptest Questions

Within undergraduate education, personal response systems are commonly used within Mazur's (1997) Peer Instruction pedagogy in which students listen to a short lecture, and then use their transmitter to answer a conceptually-based question. If more than 75% of the students correctly answer the question, the instructor continues with the lecture, otherwise, the students discuss the question within small groups and answer the question a second time. The concept-based questions (Fig. 1) are multiple-choice questions that require students to use their understanding of an underlying concept in order to determine the correct response (Crouch and Mazur 2001). Thus, students must engage in the higher-order thinking skills of application, analysis, and synthesis (Mazur 1997). Conceptual understanding is further enhanced by having students discuss their answers with their peers. Originally called ConcepTests by Mazur (1997), we choose to follow Piepmeier (1998), Gray (2009) and Mollborn and Hoekstra (2010) who used the term conceptest questions to refer to this type of question. Conceptest questions are available for a wide range of scientific and non-scientific disciplines wide range of disciplines including geoscience (Greer and Heany 2004; McConnell et al. 2006), chemistry (Butcher et al. 2003), math (Schlatter 2002), physics (Beatty et al. 2006), Biology (Brewer 2004), nursing (Debourgh 2007), packaging (Auras and Bix 2007), and economics (Maier and Simkins 2008).

Past research on conceptest questions have focused primarily on three broad themes including: (1) Implementing personal response systems in undergraduate courses (Beekes 2006; d'Inverno et al. 2003); (2) Student attitudes towards using the devices (Bombaro 2007; Trees and Jackson 2007); or (3) Impacts on student learning (Beatty et al. 2006; Fies and Marshall 2006). These studies typically do not report the student responses themselves, but rather use independent instruments such as course exams (Mayer et al. 2009), surveys (Trees and Jackson 2007), or domain-specific assessments like the Force Concept Inventory (Hestenes et al. 1992) to measure the objectives of the studies. Gray (2009) found 88 studies that used the term conceptest, but only 19 of those papers reported any type of student response data. Of those 19 studies, most used student responses as secondary data sources that augmented conclusions drawn from other instruments such as the Force Concept Inventory or course exams (e.g. Freeman et al. 2007; Hatch et al. 2005; Lucas 2007; Stowell and Nelson 2007). Those papers that did include student response data to individual questions (e.g. Piepmeier 1998; Suchman et al. 2006) used the response data to assess student learning or evaluate the peer instruction methodology and aggregated the data across one or more courses. Possible variations in student responses to individual questions or groups of questions were not explored, and these studies did not discuss any possible differences in student responses based on student demographics.

Similarly, few studies have investigated whether responses to conceptest questions are related or influenced by individual student characteristics. For example, Steer et al. (2009) analyzed over 4,700 student responses to geosciences conceptest questions and identified differences by race and gender in the percentage of correct responses. They concluded that further work is needed to evaluate whether conceptest questions accurately assess student learning or whether other confounding variables may influence the answers provided by our students. Other studies have called for similar types of research, but have focused on student response rates (King and Joshi 2008) or student attitudes towards the technology (MacGeorge et al. 2008b; Trees and Jackson 2007). Given the paucity of reported data, it is not known whether student responses vary according to how the question is worded or whether different student populations provide different responses.

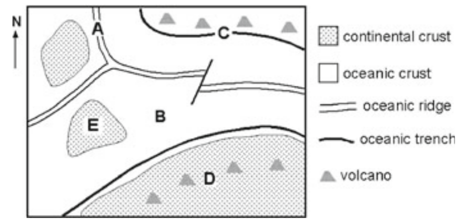
A Earthquake Conceptest Question

An earthquake occurred on the Erie Fault, 5 km beneath Ashtabula, Ohio. Damage from the earthquake was greatest in nearby Chardon. The farthest report of shaking was recorded in Akron. Where was the earthquake's epicenter?

A. The Erie Fault C. Chardon
B. Ashtabula D. Akron

B Plate Tectonics Conceptest Question

How many plates are in this image?



A. 2
B. 3
C. 4
D. 5
E. 6

Fig. 1 Examples of a text-only **a** and an illustrated **b** conceptest question

Conceptest Questions and Prior Achievement

Many of the studies described above (e.g. Crouch and Mazur 2001; Mayer et al. 2009) attempted to connect the use of personal response systems to improved student achievement on a course exam or improved grades for the course, and assessed student prior knowledge or prior achievement as a domain-specific construct. These studies considered personal response systems and conceptest questions as the treatment condition and did not analyze the student responses themselves.

Few studies have investigated whether a potential relationship exists between student academic preparation and answers on conceptest questions, but none of these studies analyzed the actual responses to questions asked in class. The two most common measures of student preparation are the Scholastic Aptitude Test (SAT) and the American College Test (ACT). For example, MacGeorge et al. (2008a) compared SAT scores to student attitudes towards personal response systems, and Lass et al. (2007) used data from the SAT as a measure of student 'innate ability'. Results from the ACT have been used in a similar manner. For example Gonzalez-Espada and Bullock (2007) compared student scores to answers on conceptest questions, whereas Johnson and Robson (2008) used the ACT as a measure of student intelligence. Studies also used ACT and SAT scores to verify homogeneity between research groups (Beuckman and Rebello 2007; Mayer et al. 2009). None of these studies compared actual student responses to measures of student academic achievement. Gonzalez-Espada and Bullock did compare ACT scores and response times to questions and found a significantly negative correlation, which suggests that a student's general academic achievement may influence their responses to conceptest questions. Yet that study's small sample size ($n = 27$) suggests that their findings should be considered tentative at best. Further research in the area is needed to determine whether students with a high level of academic achievement respond differently to conceptest questions than students with lower levels of achievement.

Connections to Multimedia Learning

Mayer (2009) defined multimedia learning as learning that includes both words and pictures. Extensive research by Mayer and his associates documented that text, illustrations, and sound can be arranged to optimize student learning, and their findings may provide insight into the use of conceptest questions and personal response systems. First, the inclusion of an illustration improves student learning. Studies show that students retain more information when reading an annotated text than when reading text without a corresponding illustration (Mayer 1989; Mayer and Gallini 1990; Moreno and Mayer 2002). Taken together, students who viewed an illustration while reading an accompanying text correctly answered 23% more questions on a summative assessment than students who just read the text (0.23 effect size). Mayer (2009) called this finding the Multimedia Effect and argued that the mind uses different cognitive operations to process visual and verbal inputs, so when presented with a diagram and words, the student retains more information than from the text or diagram individually. Second, student learning gains improve when illustrations are placed next to the associated text, which Mayer (2009) called the Spatial Contiguity Principle. Experimental results supported this hypothesis. Participants scored significantly higher for both memory retention (Moreno and Mayer 1999; Mayer 1989) and transfer of knowledge to a novel situation (Moreno and Mayer 1999; Mayer et al. 1995; Mayer 1989) when the text and illustration were on the same page rather than on separate pages. Mayer (2009) used these results to suggest that the presence of words and text together allows students to utilize both a visual and a verbal pathway within their working memory, whereas just reading text (a verbal process) or analyzing an illustration (a visual task) accesses only a single pathway. Therefore, students appear to learn best when they use illustrated texts rather than texts alone.

Research in the area of multimedia learning also suggests that individual student differences may regulate how effectively multimedia principles impact their learning (Mayer 2009). For example, students with low levels of content knowledge benefited more from viewing multimedia presentations than students with higher levels of content knowledge (Mayer and Gallini 1990; Mayer et al. 1995). The inclusion of an illustration in conjunction with text or verbal narration appears to provide contextual clues that allow students with low levels of content knowledge to better understand the target material whereas students with high levels of content knowledge do not require such visual clues.

These findings from multimedia learning may provide insight into the use of conceptest questions and personal response systems. Many conceptest questions used in undergraduate science courses include an image, diagram, or graph that must be interpreted before the correct answer can be determined (see Fig. 1b for an example); however, questions from physics (Mazur 1997), chemistry (Landis et al. 2001), and earth science (McConnell et al. 2006) also include examples where the question only contains text and does not include an illustration. Given that some, but not all, questions include an illustration, it is possible that Mayer's principles of multimedia learning may also influence student responses when answering conceptest questions. That is, students may find that illustrated questions are easier to answer than the text-only questions because they can access both visual and verbal pathways within their working memory. If this occurs, then students would correctly answer a higher percentage of illustrated questions than text-only questions. Such a finding would suggest that the illustrations allow students to devote a greater amount of mental effort towards answering illustrated questions, whereas text-only questions would require students to first identify and store that salient factual relationships embedded within the question before they could determine the correct answer.

It should be noted that Mayer's principles of multimedia learning may not impact student responses to conceptest questions. Previous studies on multimedia learning evaluated whether the design of instructional materials (i.e. the presence or absence of an illustration) affected one's ability to learn and recall *new* information acquired within a multimedia context. In these studies, the illustrations served as an additional learning opportunity that allowed the students to develop a deeper understanding of the target concept. If these principles apply to answering conceptest questions, then illustrated questions would yield a higher percentage of correct responses than questions that only contain text. In contrast, conceptest questions are intended to assess information that the student *previously* learned in a variety of contexts including reading the text or listening to lecture. In this case, the illustrations do not aid the reader in understanding new material but rather may provide a barrier towards ascertaining the correct response. If the principles of multimedia do not apply to conceptest questions, then the presence or absence of an illustration should not have any bearing on whether students select the correct response. Such a finding would suggest that other factors such as student conceptual understanding, and not the format of the questions, account for any observed differences in student responses.

This study examined whether Mayer's multimedia principle impacts student responses to conceptest question by examining student responses to illustrated and text-only questions. In addition, the study assessed whether differences in student prior achievement influence their responses to conceptest questions.

Research Questions

The following research questions were investigated in this study.

Research Question #1 Does including an illustration improve the percentage of correct responses to conceptest questions?

Research Question #2 Does student prior achievement in reading, math, or science (as measured by the ACT) relate to the percentage of correct responses to conceptest questions?

Research from multimedia learning suggests that the mode in which information is presented to the student (i.e. the presence or absence of an illustration) can influence the level of student understanding. We hypothesized that conceptest questions containing an illustration would yield a larger percentage of correct responses than questions that only included text. In addition, we hypothesized that differences in correct student responses to conceptually-based questions would vary by student prior achievement in reading, math, and science. Specifically, the inclusion of an illustration would provide a greater benefit to students with low ACT scores than to students with high ACT scores.

Methods

This study analyzed student responses to conceptest questions asked during nine sections of a large, introductory earth science course for non-science majors offered at a four-year, public university in the Midwestern United States. The course did not include a lab component and covered concepts pertaining to astronomy, geology, hydrology, oceanography, and meteorology in that sequence, and contained approximately 150 students per section. The conceptually-based questions analyzed for this paper came from a three-week time period near the beginning of the course that included topics related to the solid earth (plate tectonics, earthquakes, volcanoes, and rock formation). Two tenured geology faculty members taught all sections of this course. Together, these instructors had 26 years of combined experience teaching this course and 17 years of experience incorporating active-learning pedagogies into this course.

Participants

A total of 1,236 students enrolled in the nine courses included in this study. For this study, we only included students who signed consent letters and provided ACT scores to the university. A small number of those students (n = 31) repeated the course during the course of the study. In these cases we only analyzed data from the students' initial enrollment in the course. We also excluded high school students enrolled in the course for dual credit as well as students who answered fewer than five text-only and illustrated questions. We removed students who answered fewer than five questions of either type because answering a small number of questions yielded very large changes in the percentage of correct responses. A total of 561 students met all of these criteria and were included in the study. Many of these students (74%) were freshman, so this earth science course represented their first undergraduate science experience. Student demographics (gender, race, grade level, ACT scores, degree program) did not significantly vary across each of the nine courses. During the first or second week of each semester, the instructor distributed a consent letter obtaining permission to analyze their answers to the course's conceptest questions. The instructors permanently placed the students into four-person teams and assigned a designated seating location for each group. Every class session utilized Mazur's peer instruction pedagogy where students individually answered conceptest questions. Students also participated in a variety of in-class activities intended to reinforce learning or completed worksheets within these groups.

Instruments

During all nine course sections, the instructors followed the peer instruction pedagogical model (Mazur 1997) by asking conceptest questions throughout their lectures. If fewer than 75% of the students correctly answered the question, the students discussed the question within their teams and answered a second time. For this study we chose to exclude responses given after student discussions because they represented the collective understanding of the group rather than the conceptual understanding of each individual student. All of the questions analyzed in this study were multiple-choice questions that offered two to five possible solutions and are listed in Gray (2009). The personal response system software (eInstruction 2009) automatically tabulated each student's responses for further analysis.

We operationally defined prior achievement using the results from the American College Test (ACT). The test consists of four separate sub-tests that cover mathematics, science, reading, and English. All four sub-tests are scored on a 36-point scale with a nationally normalized average at or near 20. A fifth score (ACT-Composite) is the mean of the four sub-tests. We chose to use the ACT because it is designed to measure a student's general academic preparedness for succeeding in a university environment after taking a typical college-preparatory curriculum (achievement) rather than assessing the student's general verbal and mathematical knowledge (aptitude) (American College Test 2008a). In addition, student responses might be influenced by their level of scientific knowledge, and the ACT (unlike the SAT) includes a test for science knowledge. Furthermore, at our university, a majority of students provided ACT scores rather than SAT scores.

Data Analysis

The 561 students provided a total of 20,453 responses to 155 separate questions. To test whether the presence of an illustration influenced student responses, we categorized the conceptest questions based on the format of the question. *Text-only* questions contained only words whereas *Illustrated* questions contained words plus a map, diagram, graph, or hand specimen that required interpretation before the correct answer could be determined (Fig. 1). Nineteen questions contained only text but were classified as illustrated questions because they all referred to a physical specimen (e.g. a rock sample) or worksheet that the students had at their desks that required an interpretation before the correct answer could be determined. In contrast, none of the text-only questions required the students to interpret any type of diagram, graph, or object. Had these nineteen questions included a photograph of the referent object, the question would clearly have been an illustrated question. Using these criteria resulted in 43 text-only questions and 112 illustrated questions. For each student, we calculated the percentage of correct responses for answers to the text-only and illustrated conceptest questions.

Answering conceptest questions required students to use several academic skills that were assessed by the ACT. When presented with a conceptest question, students had to accurately read and comprehend the text before applying their knowledge of earth science to select the correct answer. This process suggested that student reading ability might impact their ability to answer conceptest questions. Some questions also required students to utilize math skills such as reading a graph or calculating a rate (such as the rate of a moving tectonic plate). It was not known whether student proficiency in any of these three domains is linked to student success in answering conceptest questions. In addition, these conceptest questions assessed students for understanding of science concepts, so student prior achievement in science might impact their answers on science-related conceptest questions. Ideally students of all achievement levels answer a similar percentage of text-only questions and illustrated questions. Such a finding would further support the claim that conceptest questions formatively assess student learning.

To test this hypothesis, results from the ACT-Reading, Math, and Science subtests were compared to the student response data. For the text-only questions, it was hypothesized that students who were good readers (as measured by the ACT-Reading sub-test) would correctly answer a larger percentage of text-only questions than illustrated questions. Similarly, correctly interpreting the illustrated questions may have required interpretive and analytical skills typically taught in math and science courses; therefore it was hypothesized that students who had demonstrated a high level of achievement in math and/or science would correctly answer a significantly higher percentage of illustrated questions than text-only questions. We divided the student ACT scores into four achievement categories

Table 1 Defining ACT values and number of students in each achievement category

Achievement category	Quartile ranges[a]			Number of students		
	Reading	Math	Science	Reading	Math	Science
High	26–36	25–36	24–36	123	91	136
High-average	21–25	20–24	21–23	174	177	156
Low-average	17–20	17–19	18–20	147	169	173
Low	1–16	1–16	1–17	117	124	96

[a] Based on data from ACT (2008b) on students who took the test in 2007

(Table 1) based on the published quartile levels for each test (American College Test 2008b).

To evaluate both hypotheses we used a within-subjects (Stevens 2007), repeated measures, mixed analysis of covariance (ANCOVA) for each of the three ACT tests included in this study. In within-subjects designs, all participants provide data for each dependent variable, which improves the power of the results and eliminates the need for an independent control group (Davis 2002). Thus the results from the within-subjects analysis identified any differences in student responses to the two different types of questions. The multimedia principle (Mayer 2009) states that including an illustration would improve student comprehension. If this principle applied to conceptest questions, then students would answer a significantly higher percentage of illustrated questions than text-only questions.

The repeated measures analysis also assessed the between-subjects effects, which determined whether student reading, math, or science achievement could explain their responses to text-only and illustrated conceptest questions. To further assess differences in student responses between different achievement categories, we used a post-hoc Bonferroni, pairwise comparison of student responses to each type of question for every combination achievement categories. Lastly, a series of post-hoc, paired *t* tests assessed differences in student responses between the text-only and illustrated conceptest questions within each achievement category. Even with 561 participants, we set our level of significance at $\alpha = .05$ to ensure that any small effect sizes would be significant (Cohen 1988).

Results

Covariate Analysis

Before analyzing the data, we tested for the presence of covariates that might influence the final results. Independent-samples *t* tests using instructor as the independent variable found significant differences for the percentage of correct responses to both text-only questions ($t_{(375)} = -2.35$, $p = .02$) and illustrated questions ($t_{(375)} = -4.84$, $p < .001$). The presence of a significant difference in student responses from classes taught by two different instructors suggested that factors pertaining to how each instructor incorporated the conceptest questions into his courses may have affected the percentage of correct responses for each type of question. By including the instructor as a covariate, we eliminated this observed difference and included it in all subsequent analyses.

Results Between Different Types of Questions

The repeated measures mixed-ANCOVA for each ACT test (Reading, Math, and Science) yielded no significant difference between the percentages of correct responses to text-only questions compared to illustrated questions. The within-subjects effects (Table 2) indicated that no significant differences existed between the type of conceptest question (text-only vs. illustrated) and scores on any of the three ACT tests included in this study.

Results Between Different Achievement Levels

In contrast, the between-subjects analysis (Table 3) yielded significant differences in student responses between the four ACT achievement categories. The Bonferroni post-hoc analysis (Table 4) revealed that significant differences in responses between all combinations of categories for all three tests with effect sizes (Cohen's d) ranging from 0.32 to 1.05. The only groups that did not yield significant differences were comparisons of responses from students in the high-average and low-average categories. The Bonferroni analysis considered all student responses and did not separate out responses by the type of question. A series of paired *t* tests evaluated the influence of question type by calculating whether the percentage of correct responses to text only and illustrated conceptest questions significantly varied for students at each level of ACT achievement (Table 5). Only one *t* test (students with high achievement scores on the ACT-Math test) yielded a significant difference. These students correctly answered 3.1% more illustrated questions than text-only questions with an associated effect size of 0.21. All other achievement levels yielded

Table 2 Results of the within-subjects analysis of student responses for question type (QT) and ACT achievement scores

Variable	df	F	p	η^2
Reading				
Question type	1	0.08	.78	0.01
QT × Instructor	1	0.57	.45	0.03
QT × ACT quartile	3	0.52	.67	0.05
Error	556	−152.75		
Math				
Question type	1	0.01	.93	0
QT × Instructor	1	0.4	.53	0
QT × ACT quartile	3	1.11	.34	0.01
Error	556	−152.26		
Science				
Question type	1	0.05	.83	0
QT × Instructor	1	0.49	.48	0
QT × ACT quartile	3	0.32	.81	0
Error	556	−152.91		

Values in parentheses are mean square errors

Table 3 Results of the between-subjects analysis of student responses for question type (QT) and ACT achievement scores

Variable	df	F	p	η^2
Reading				
Instructor	1	31.72	<.01	0.03
ACT quartile	3	22.74	<.01	0.06
Error	556	−280.93		
Math				
Instructor	1	32.73	<.01	0.03
ACT quartile	3	30.09	<.01	0.08
Error	556	−271.33		
Science				
Instructor	1	37.27	<.01	0.03
ACT quartile	3	29.67	<.01	0.08
Error	556	−271.87		

Values in parentheses are mean square errors

non-significant differences between the percentage of correct responses to text-only and illustrated questions. Figure 2 graphically illustrates the results from all three analyses.

Discussion

Research Question #1

The results from this study indicate that Mayer's (2009) multimedia principle does not influence student responses to illustrated conceptest questions. The similarity in the percentage of correct responses to both text-only and illustrated questions suggests that factors other than the presence or absence of an illustration influence student responses to these questions. These results are consistent with the hypothesis that the multimedia principle operates in a domain when students use words and pictures to learn *new* material rather than assessing *prior learning*, and suggests that instructors do not need to consider the multimedia principle when writing conceptest questions.

The results from the post hoc analyses also support this conclusion. The consistent lack of any significant differences between responses to text-only and illustrated questions across all levels of achievement fails to support the hypothesis that students at one end of the achievement continuum might be influenced by multimedia factors whereas students elsewhere on the continuum might not be affected. This trend is consistent across both achievement level and ACT content domain. The small, non-significant differences in the means between responses to text-only and illustrated questions for all but one level of ACT achievement suggest that all students use similar mental processes when answering both types of questions rather than following cues based upon the way in which the question is presented. In addition, the one group of students (high achieving math students) with a significant difference in the percentage of correct responses only produced a three percent difference in the percentage of correct responses between the two types of questions. A three percent difference equates to one additional correct response to an illustrated question every four class sessions. It is highly unlikely that an instructor would detect such a difference while giving a lecture, and coupled with the associated small effect size (d = 0.21), this significant difference does not invalidate the conclusion that the form in which a conceptest question is written (text-only versus illustrated) does not influence how students select their answer.

An alternate interpretation of the data might be that students with superior math skills are better at interpreting visual data displays such as graphs and therefore are better at interpreting illustrated conceptest questions than students with weaker math skills. This interpretation is not supported by the data. The students with the *lowest* ACT Math scores had a similar difference (2.99%) between the percentage of correct responses to both types of questions as students with the *highest* ACT Math scores, which suggests that prior math achievement is not a contributing factor in correctly responding to text-only questions compared to illustrated questions.

Research Question #2

The significant results from the between-subjects analysis (Table 3) indicate that students with different levels of

Table 4 Bonferroni post hoc pair-wise comparisons of correct student responses categorized by ACT achievement category

ACT achievement category	High		High-average		Low-average	
	p	d	*p*	d	*p*	d
ACT-reading						
High-average	.004	0.34				
Low-average	<.001	0.56	.11	–		
Low	<.001	0.82	<.001	0.47	.026	0.27
ACT-math						
High-average	<.001	0.60				
Low-average	<.001	0.63	1.00	–		
Low	<.001	1.05	<.001	0.47	<.001	0.42
ACT-science						
High-average	<.001	0.57				
Low-average	<.001	0.65	1.00	–		
Low	<.001	0.95	.001	0.36	.006	0.32

Effect sizes (d) were not calculated for non-significant comparisons

Table 5 Paired *t* test post hoc analysis comparing student responses to text-only versus illustrated questions for each achievement category

ACT achievement category	Mean difference[a]	SD	t	df	*p*	d
ACT-reading						
High	−0.3	16.37	−0.2	122	.84	−0.02
High-average	−1.78	17.43	−1.34	173	.18	−0.10
Low-average	−1.32	17.11	−0.94	146	.35	−0.08
Low-average	−2.95	19.04	−1.67	116	.10	−0.15
ACT-math						
High	−3.09	14.56	−2.02	90	.05	−0.21
High-average	−1.57	18.43	−1.13	176	.26	−0.09
Low-average	0.27	17.91	0.19	168	.85	0.02
Low-average	−2.99	17.28	−1.93	123	.06	−0.17
ACT-science						
High	−0.46	15.49	−0.35	135	.73	−0.03
High-average	−2.25	19.09	−1.47	155	.14	−0.12
Low-average	−1.48	17.2	−1.13	172	.26	−0.09
Low-average	−2.24	17.9	−1.23	95	.22	−0.13

Paired *t* tests do not include the variable Instructor as a covariate

[a] Mean Difference = (Text-Only) − (Illustrated). Negative values indicate the mean % of correct responses to illustrated questions is larger than the mean % of correct responses to text-only questions

reading, math, and science achievement yielded different percentages of correct responses. Furthermore, the percentages of correct responses given by students at each level of achievement for all three tests significantly differed from students in other achievement groups (Table 4). Only the students from the two middle categories (high-average and low-average) were not significantly different. These differences are illustrated in Fig. 2. Students in the middle two categories yielded response values that are similar in size but different from students at either end of the spectrum. On average, students who scored highest on the ACT tests also correctly answered the highest percentage of conceptest questions. Similarly, students who scored the lowest on any of the tests correctly answered the lowest percentage of conceptest questions.

For all three tests (ACT-Reading, ACT-Math, and ACT-Science) students from within each achievement category (high, high-average, low-average, and low) correctly answered a significantly different percentage of questions. For example, students from the high achievement category were more likely to select the correct answer than students in the other three categories. The results are consistent

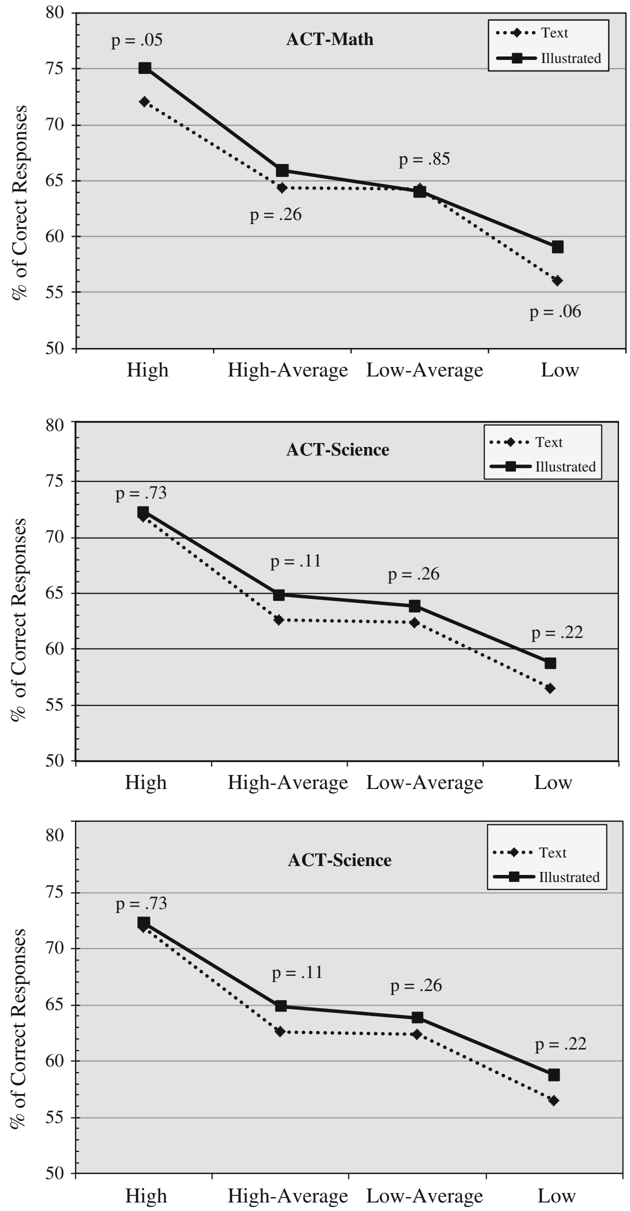


Fig. 2 Percentage of correct student responses for each type of question compared to ACT-reading, Math, and Science achievement categories. *p* values do not include the variable instructor as a covariate

across all categorical combinations except the responses between students in the high-average and low-average categories. It is tempting to conclude that conceptest questions measure student understanding and can be used as a valid formative assessment tool. King and Joshi (2008) found that a high percentage of students who correctly answered conceptest questions in a chemistry course also correctly answered similar questions on a subsequent exam. Conversely, they found that significantly fewer students incorrectly answered the conceptest questions but gave correct answers on the exam. Other studies such as McConnell et al. (2006) and Mayer et al. (2009) link the peer instruction pedagogy and conceptest questions to gains in student performance. Many authors also describe student responses to conceptest questions as a formative assessment tool (Beatty and Gerace 2009; Dufresne and Gerace 2004; MacArthur and Jones 2008), however only King and Joshi provide data comparing answers to conceptest questions to answers on subsequent exams. Additional work in this area could provide additional evidence that correctly answering a conceptest question (especially the first time the question is asked) accurately assesses student understanding of the underlying concept.

Another possible explanation for the significant differences between answers to conceptest questions and prior achievement is that student responses to conceptest questions are solely based on prior levels of achievement or prior content knowledge. Under this scenario, the students respond to what they already know rather than what they learned in class. This explanation in unlikely because it does not account for the fact that earth science courses typically cover content that is typically unfamiliar to the student (Libarkin et al. 2005).

The most likely explanation is that the percentage of correct responses to conceptest questions reflects differences in the student's general academic preparedness to succeed in an undergraduate earth science course. The stated intent of the ACT tests is to measure student academic preparedness for success at the university level (ACT 2008a). Even though the ACT is not designed to measure a student's general cognitive ability, several studies have used the test in this manner (Born et al. 2002; Thompson and Zamboanga 2004), and Koenig et al. (2008) have shown that the ACT is a valid measure of student academic ability.

The similarity of our results between the different ACT tests (Fig. 2) also supports this conclusion. For each level of academic achievement, students correctly answered similar percentages of conceptest questions. Data for the ACT-English and ACT-Composite tests are not reported here, but yielded similar results. The similarity of student responses across differing achievement levels and different test domains suggests that external factors such as student reading level or math ability are not significant factors in determining a student's response to a conceptest question. Rather, student understanding of the underlying scientific concepts is more likely to have influenced student responses.

Our study used student ACT scores to measure prior achievement in math, science, and reading, yet the data suggest that this variable more likely defines the students' general cognitive ability. Given the similarity of responses to conceptest questions across all three domains, it seems likely that the ACT values measure student ability to succeed in an undergraduate science course rather than measure their science or math knowledge. This is consistent with the findings of Koenig et al. (2008) who demonstrated that the ACT is a valid and reliable measure of general cognitive ability. In addition, the similar relationship between different ACT tests and responses to conceptest questions suggests that future research could use the more general ACT-Composite score rather than data from the three domain specific tests.

Summary

Many instructors utilize Mazur's (1997) peer instruction pedagogy when incorporating personal response systems into non-major's science course. This teaching practice relies upon the conceptually-based conceptest questions to assess student learning and determine whether students should discuss a given topic. To date, few studies have investigated whether factors other than student content knowledge influence their responses. This study suggests that the way a question is written does not influence how students respond. That is, the presence or absence of an illustration does not affect student responses. These questions are conceptually-based rather than factually based, which requires students to use higher-order thinking skills (such as application, analysis, and synthesis) to arrive at the correct response. The lack of a significant difference between questions with and without an illustration suggests that conceptest questions do assess student content knowledge. Similarly, the significant differences in the percentage of correct responses between different levels of student achievement support the claim that conceptest questions are a valid type of formative assessment. Furthermore, the similarities between the different three tests investigated in this study suggest that student reading levels or math ability by themselves do not influence student responses to conceptest questions.

Further Research

Additional studies are needed to determine what factors are present in students who score high on the ACT but lacking in students who score low on the same measures. Once these factors have been identified, a comparison with student responses to conceptest questions could provide meaningful data on the factors that contribute to student responses. Other studies comparing student responses to conceptest questions and conceptual understanding would strengthen the argument that conceptest questions are a valid type of formative assessment.

Acknowledgments Partial support for this work was provided by the National Science Foundation's Division of Undergraduate Education program under Award No. 0716397. 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. We would also like to thank Dawn Del Carlo and Jeff Morgan for helpful suggestions on an earlier draft of the manuscript.

References

American College Test (ACT) (2008a) Facts about the ACT. Retrieved 10 Sept 2008, from http://www.act.org/news/aapfacts.html

American College Test (ACT) (2008b) ACT High School profile: section II, academic achievement. Retrieved 21 Jan 2009 from http://www.act.org/news/data/07/pdf/two.pdf

Auras R, Bix L (2007) WAKE UP! The effectiveness of a student response system in large packaging classes. Packag Technol Sci 20(3):183–195

Beatty ID, Gerace WJ (2009) Technology-enhanced formative assessment: a research-based pedagogy for teaching science with classroom response technology. J Sci Educ Technol 18(2):146–162

Beatty ID, Gerace WJ, Leonard WJ, Dufresne RJ (2006) Designing effective questions for classroom response system teaching. Am J Phys 74(1):31–39

Beekes W (2006) The 'millionaire' method for encouraging participation. Active Learn High Educ 7(1):25–36

Beuckman J, Rebello NS (2007) How does a classroom interaction system affect student performance? Paper presented at the 2007 Annual Meeting of the National Association of Research in Science Teaching, New Orleans, LA. Retrieved from http://perg.phys.ksu.edu/papers/2007/rebello-beuckman-narst.pdf

Bombaro C (2007) Using audience response technology to teach academic integrity: "The seven deadly sins of plagiarism" at Dickinson College. Ref Serv Rev 35(2):296–309

Born WK, Revelle W, Pinto LH (2002) Improving Biology Performance with Workshop Groups. J Sci Educ Technol 11(4):347–365

Brewer CA (2004) Near real-time assessment of student learning and understanding in biology courses. Bioscience 54(11):1034–1039

Butcher DJ, Brandt PF, Norgaard NJ, Atterholt CA, Salido AL (2003) Sparky introchem: a student-oriented introductory chemistry course. J Chem Educ 80(2):137–139

Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Erlbaum, Hillsdale, NJ

Crouch CH, Mazur E (2001) Peer instruction: ten years of experience and results. Am J Phys 69(9):970–977

d′Inverno R, Davis H, White S (2003) Using a personal response system for promoting student interaction. Teach Math Appl 22(4):163–169

Davis CS (2002) Statistical methods for the analysis of repeated measurements. Springer, New York

Debourgh GA (2007) Use of classroom "clickers" to promote acquisition of advanced reasoning skills. Nurse Educ Pract 8(2):76–87

Dufresne RJ, Gerace WJ (2004) Assessing-to-learn: formative assessment in physics instruction. Phys Teach 42:428–433

eInstruction. (2009). eInstruction, simple solutions. Real results. Accessed May 1, 2009 from http://www.einstruction.com

Fies C, Marshall J (2006) Classroom response systems: a review of the literature. J Sci Educ Technol 15(1):101–109

Freeman S, O'Conner E, Parks JW, Cunningham M, Hurley D, Haak D, Wenderoth MP (2007) Prescribed active learning increases performance in introductory biology. CBE Life Sci Educ 6(2):132–139

Gonzalez-Espada WJ, Bullock DW (2007) Innovative applications of classroom response systems: Investigating students' item response times in relation to final course grade, gender, general point average, and high school ACT scores. Electron J Integr Technol Educ 6:97–108

Gray KR (2009) A study of student responses to text-only and illustrated conceptest questions related to plate tectonics: differences by gender and prior achievement. Unpublished Dissertation: University of Akron

Greer L, Heany PJ (2004) Real-time analysis of student comprehension: an assessment of electronic student response technology in an introductory earth science course. J Geosci Edu 52(4):345–351

Hatch J, Jensen M, Moore R (2005) Manna from heaven or "clickers" from hell: experiences with an electronic response system. J Coll Sci Teach 34(7):36–42

Hestenes D, Wells M, Swackhamer G (1992) Force concept inventory. Phys Teach 30:141–158

Johnson M, Robson D (2008) Clickers, student engagement and performance in an introductory economics course: a cautionary tale. Comput High Educ Econ Rev 20:4–12

King DB, Joshi S (2008) Gender differences in the use and effectiveness of personal response devices. J Sci Educ Technol 17(6):544–552

Koenig KA, Frey MC, Detterman DK (2008) ACT and general cognitive ability. Intelligence 36(2):153–160

Landis CR, Ellis AB, Lisenky GC, Lorenz JK, Meeker K, Wamser CC (2001) Chemistry conceptests: a pathway to interactive classrooms. Prentice Hall, Upper Saddle River, NJ

Lass D, Morzuch B, Rogers R (2007) Teaching with technology to engage students and enhance learning (Working Paper No. 2007-1). Retrieved from the University of Massachusetts Amherst's website: http://www.umass.edu/resec/workingpapers

Libarkin JC, Anderson SW, Dahl J, Beilfuss M, Boone W (2005) Qualitative analysis of college students' ideas about the earth: interviews and open-ended questionnaires. J Geosci Edu 53(1):17–26

Lucas A (2007) Using peer instruction and i-clickers to enhance student participation in calculus. Primus. Accessed on 11/20/2008 at http://www.vanderbilt.edu/cft/resources/teaching_resources/technology/crs_biblio.htm#math

MacArthur JR, Jones LL (2008) A review of literature reports of clickers applicable to college chemistry classrooms. Chem Educ Res Pract 9:187–195

MacGeorge EL, Homan SR, Dunning JB, Elmore D, Bodie GD, Evans E, Lichti SM (2008a) The influence of learning characteristics on evaluation of audience response technology. J Comput High Educ 19(2):25–46

MacGeorge EL, Homan SR, Dunning JB Jr, Elmore D, Bodie GD, Evans E, Beddes B (2008b) Student evaluation of audience response technology in large lecture classes. Educ Technol Res Dev 56(2):125–145

Maier M, Simkins S (2008) Learning from physics education research: lessons for economics education. Retrieved 15 Sept 2008 from http://arxiv.org/ftp/arxiv/papers/0807/0807.3534.pdf

Mayer RE (1989) Systemic thinking fostered by illustrations in scientific text. J Educ Psychol 81(2):240–246

Mayer RE (2009) Multimedia learning, 2nd edn. Cambridge University Press, New York

Mayer RE, Gallini JK (1990) When is an illustration worth ten thousand words? J Educ Psychol 82(4):715–726

Mayer RE, Steinhoff K, Bower G, Mars R (1995) A generative theory of textbook design: using annotated illustrations to foster meaningful learning of science text. Educ Tech Res Dev 43(1):31–43

Mayer RE, Almeroth K, Bimber B, Chun D, Knight A, Campbell J (2009) Clickers in college classrooms: fostering learning with questioning methods in large lecture classes. Contemp Educ Psychol 34(1):51–57

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

McConnell DA, Steer DN, Owens KD, Knott JR, Van Horn S, Borowski W, Heaney PJ (2006) Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. J Geosci Edu 54(1):61–68

Mollborn S, Hoekstra A (2010) A meeting of minds: using clickers for critical thinking and discussion in large sociology classes. Teach Sociol 38(1):18–27

Moreno R, Mayer RE (1999) Gender differences in responding to open-ended problem-solving questions. Learn Individ Differ 11(4):355–364

Moreno R, Mayer RE (2002) Learning science in virtual reality multimedia environments. J Educ Psychol 96(3):598–610

Piepmeier E Jr (1998) Use of conceptests in a large lecture course to provide active student involvement and peer teaching. Am J Pharm Educ 62:347–351

Schlatter MD (2002) Writing conceptests for a multivariable calculus class. Primus 12(4):305–314

Steer D, McConnell D, Gray K, Kortz K, Liang X (2009) Analysis of student responses to peer-instruction conceptual questions answered using an electronic response system: trends by gender and ethnicity. Sci Educ 18(2):30–38

Stevens J (2007) Applied multivariate statistics for the social sciences. Lawrence Erlbaum Associates, Hillsdale, NJ

Stowell JR, Nelson JM (2007) Benefits of electronic audience response systems on student participation, learning, and emotion. Teach Psychol 34(4):253–258

Suchman E, Uchiyama K, Smith R, Bender K (2006) Evaluating the impact of a classroom response system in a microbiology course. Microbiol Educ 7:3–11

Thompson RA, Zamboanga BL (2004) Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. J Educ Psychol 96(4):778–784

Trees AR, Jackson MH (2007) The learning environment in clicker classrooms: student processes of learning and involvement in large university-level courses using student response systems. Learn Media Technol 32(1):21–40