

## EXPLORING THE ROLE OF CONTEXT IN STUDENTS' UNDERSTANDING OF SAMPLING

JACQUELINE R. WROUGHTON  
*Northern Kentucky University*  
*wroughtonj1@nku.edu*

HERLE M. MCGOWAN  
*North Carolina State University*  
*hmmcgowa@ncsu.edu*

LEIGH V. WEISS  
*Capital University*  
*lweiss@capital.edu*

TARA M. COPE  
*State University of New York Adirondack*  
*copet@sunyacc.edu*

### ABSTRACT

*Context provides meaning for data analysis and the evaluation of evidence but may be distracting to students. This research explores the role of context in students' reasoning about sampling: specifically, the relationship between the strength of students' opinions about a topic, which provides the context for a study, and their ability to judge the quality of the sampling method and the scope of the conclusions in the study. Data were collected at four diverse institutions in both a testing environment and through individual interviews. Student responses were analyzed using a grounded theory approach. Testing environment results showed little evidence of the use of context whereas interview results showed more evidence of reliance on context-based opinions rather than statistical principles.*

**Keywords:** *Statistics education research; Statistical reasoning; Sampling; Confirmation bias; Belief bias*

### 1. INTRODUCTION

The need for using unbiased sampling methods is one of the most important topics taught in an introductory statistics course. Having data that is reliably obtained is necessary to ensure appropriate inference (Rossman, 2008). Thus, developing an ability to assess the quality of the sampling methods used in a research study is critical for developing a complete understanding of statistical inference. It is particularly important for students to recognize when the selection method is likely to lead to bias, or when a study's conclusions are too broad for the sample that was selected.

In addition to emphasizing the importance of random sampling methods, most statistics educators believe it is important to discuss data in context with students: "It is important to use real data in teaching statistics to be authentic to consider issues related to how and why the data were produced or collected, and to relate the analysis to the problem context" (American Statistical Association, 2005, p. 16). Although there is support for using context when teaching statistics, there is also evidence that context may be distracting to students (e.g., Watson & Moritz, 2000a). Depending on the particular context of a problem and how strongly a student feels about that topic, the student may rely on his or her own opinion to make judgments about the quality of the data obtained, rather than focusing on statistical sampling principles. Thus, the role of context in students' reasoning should be more

thoroughly explored, especially when it comes to a concept as foundational to statistics as sampling. This paper explores the interaction between the strength of students' opinions about a topic (where this topic provides the context of the study) and their ability to judge the quality of the sampling method and the scope of the conclusions in the study. Specifically, we explore the role that two psychological phenomena—confirmation bias and belief bias—play in students' conceptual understanding of sampling.

## 2. LITERATURE REVIEW

### 2.1. THE IMPORTANCE OF CONTEXT

Many authors have discussed the importance of data context to the discipline of statistics—that context is what separates statistics from mathematics (Moore & Cobb, 2000) and provides meaning for data analysis (ASA, 2005; Konold & Higgins, 2003). In our paper, we use the definition of data context that is put forward by Langrall, Nisbet, and Mooney (2006, p. 1): “the real-world phenomena, settings, or conditions from which [data] are drawn or about which data pertain” and which is consistent with how other statistics educators have used the term. The *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report* (ASA, 2005, p. 11) states that the “desired result of all introductory statistics courses is to produce statistically educated students, which means that students should develop statistical literacy and the ability to think statistically.” To satisfy these overall goals of the introductory statistics course, educators are encouraged to use real data and real-world (contextual) problems. These recommendations are found both in the GAISE college report and in many papers written by statistics education researchers. These authors contend that using real-world problems in the classroom engages the students (ASA, 2005), discourages the use of a formulaic approach to teaching (Konold & Higgins, 2003), and helps the students to develop a more “mature statistical understanding” (Schwartz, Goldman, Vye, & Barron, 1998, p. 234).

It is argued that the use of contextual problems helps students become more active learners by engaging them further in the material that is being taught in the classroom. When students are provided with contextual problems that are both interesting and relevant to their lives, they are more active participants in their education. Teaching students using contextual problems also helps them see how material they are being taught applies to real-world situations (ASA, 2005).

Beyond applicability, using contextual problems encourages students to develop a conceptual understanding of the material and gives meaning to the interpretation of results (delMas, Garfield, & Zieffler, 2009; Konold & Higgins, 2003). Konold and Higgins (2003) further argue that using data without a context lends itself to a formulaic, rather than student-centered, approach to teaching statistics.

When educators help students develop a conceptual understanding of the material and focus on interpreting the results of statistical analyses, they help the students to become more mature in their statistical thinking. Mooney (2002) argues, as he outlines his framework for middle school students' statistical thinking, that the most sophisticated thinkers are the students who can evaluate both quantitative and contextual information to make reasonable conclusions. Schwartz, Goldman, Vye and Barron (1998) argue that it is important for students to see problems in context because they need to use their intuition about the context to help develop a deeper statistical understanding of the material. They also argue that students who experience incongruence, a dissonance between common sense based on intuitions and mathematical understanding, can especially develop a higher level of statistical thinking. Unfortunately, this incongruence between a student's prior beliefs or intuition and the statistical results can also lead to difficulties in the statistics classroom.

### 2.2. THE CONSEQUENCE OF CONTEXT

One such potential difficulty, and an unintended consequence of using data in context, is that the context may distract students from the statistical principles they are trying to learn. Often contexts are unstructured and require students to make judgments about a problem. The student then defends those judgments by expressing personal opinions or beliefs and demonstrates uncertainty about what

concepts or principles are necessary to solve the problem (Lampen, 2010). Several examples of this appear in the literature.

**Research specific to sampling** Jacobs (1999) found that children’s evaluations of sampling methods were influenced by how well the results corresponded with what they would have expected to happen in the real-world context. Watson and Moritz (2000a) found that males were more likely to answer a question situated in the context of buying a car based on the brands of the cars rather than on the statistical sample quality. In a study of how students’ notion of the concept of sample develops as they progress through school, Watson and Moritz (2000b) also found that “most students offered their own examples of samples that were based on experiences” (p. 56). This reliance on their own experience lessened as the students matured.

While the above examples demonstrate that students may ignore the statistical issues of sampling in favor of relying on personal experience, there is also research that shows they may not “believe” the principles of good sampling. In a study of 12 high school students, Rubin, Bruce, and Tenney (1991) found that students had a general distrust of using a random mechanism to assign people to groups, feeling that it would not produce balanced samples. The authors speculated that this could be due to students’ over-reliance on personal experience with small samples, which are more likely to be unbalanced than large samples.

**Other research on context** The distracting nature of context is observed not only in relation to the principles of sampling. Schwartz et al. (1998) gave examples of students who were unable to reason properly about a concept when presented in a detailed context but were later able to reason properly about the same concepts when presented in a context-free manner. Langrall, Nisbet and Mooney (2006) found that for group discussion questions that related to a topic on which one particular student was an expert, there was more context-related discussion although the statistical knowledge used in the group discussion remained the same (regardless of context or the presence of a student who was a context expert).

Students’ reliance on personal experience extends beyond the principles of sampling to general interpretation of data. For example, a study by Russell, Schifter, and Bastable (2002) of students in a kindergarten class revealed that their interpretations of a data plot were influenced by their memories of the data collection (as cited in Konold & Higgins, 2003, pp. 14-15). In their development of a framework for studying informal inferential reasoning, Zieffler, Garfield, delMas, and Reading (2008) recognized that is important to consider the student’s use of informal knowledge (defined as “everyday knowledge of the problem context, prior knowledge about statistical concepts, real world knowledge and experience, and statistical language”, p. 53), as well as if the use of context has “over-ridden the use of data in making inferences” (p. 53).

Finally, students often believe events should be explained causally or intuitively (Rossman, 2008; Schwartz et al., 1998) and thus may rely on their knowledge of the context to provide such explanations. In a study of fifth graders, Hancock, Kaput and Goldsmith (1992) reported that students had trouble seeing the data as a group, and would rather focus on the particulars of an individual student. The researchers even had students strip out identifying markers from their datasets, but students were able to remember which data came from which student and still relied on their knowledge of the particular student from which the data point originated.

### 2.3. CONFIRMATION BIAS AND BELIEF BIAS

Since the 1970s, researchers have worked to identify biases and fallacies that commonly arise when people reason about basic ideas in statistics and probability (e.g., Kahneman, Slovic & Tversky, 1982; Konold, 1989; Tversky & Kahneman, 1973). For example, in their 1973 paper, Tversky and Kahneman found that people tend to assign probabilities to events based on their personal experience with them—i.e., how frequently the event appears in that person’s own life.

Additional research on the effect of personal experiences and beliefs on one’s reasoning has led to the identification of two related psychological phenomena that statisticians and particularly statistical educators need to be aware of—confirmation bias and belief bias. These biases manifest as the selective interpretation of evidence that pertains to beliefs—specifically, people tend to seek out or

interpret evidence in a way that confirms existing expectations (Allport, 1954; Nickerson, 1998) or rate the strength of arguments based on their believability (e.g., Munro, 2010).

For example, in a study involving capital punishment, Lord, Ross, and Lepper (1979) asked students, some who supported capital punishment and some who did not, to rate the quality of two studies, which presented opposite findings on the deterrent effects of the death penalty. They found that students provided the more favorable rating to whichever study supported their initial beliefs. In fact, Lord and his colleagues also found that participating in this exercise resulted in the students having more confidence in their initial beliefs.

As another example, Kaplan (2009) found evidence of such bias when statistics students were presented with results of an experiment on extra-sensory perception (ESP): the students questioned the design of the experiment only when they did not believe its conclusions. People typically do not want to hear the results of a study that conflict with their beliefs and thus find flaws with the study, since information that is contradictory to a strongly held belief “creates an inconsistency in the cognitive system that threatens one’s self image as a smart person” (Munro, 2010, p. 582).

Evidence of confirmation bias and belief bias has been found at all ages. For example, Jacobs (1999) found evidence in upper-elementary school students when considering various sampling methods. Students tended to rate a sampling method more favorably when the results of a survey based on that sample agreed with their existing expectations, and less favorably when the results did not agree with their expectations. Kaplan’s 2009 ESP study was conducted with college-age students. She noted three reactions when students were faced with what they considered to be unbelievable conclusions: 1) to look for problems in the design of the experiment, 2) to ask for more information, particularly additional studies that replicated the results presented, and 3) to formulate an alternative, rational explanation that explained (or explained away) the results of the study. Finally, Koehler (1993) found that both advanced graduate students and practicing scientists rated the quality of evidence higher for studies that agree with their prior beliefs, and that this effect was larger among those who had stronger prior beliefs.

***Confirmation bias, belief bias and sampling*** The Jacobs (1999) study found evidence of confirmation bias and belief bias explicitly affecting the ability to reason statistically about sampling. The Koehler (1993) study also hints at the presence of these biases when evaluating sampling, as they appeared to override consideration of the quality of sampling methods, which is critically important when evaluating the quality of evidence provided by a study. How widespread the effects of these biases are on statistical reasoning should be of great concern to statistics educators. As such, the current research study searches for evidence of confirmation bias and belief bias in a population of great concern to statistics educators: students currently taking an introductory statistics course. Due to the importance of understanding the need for using unbiased sampling methods, we focus specifically on undergraduate students’ reasoning about sampling.

### 3. RESEARCH METHODOLOGY

Students at the authors’ academic institutions took part in this research. The varied institutions—a small liberal arts college, a medium sized midwestern university, a large research-focused institution, and a community college—yielded a diverse group of participants and perspectives (see Section 4.1 for more detailed information about the sample). Each course in which this research was implemented was an algebra-based introductory statistics course.

#### 3.1. PROCEDURE, TASKS, AND INSTRUMENTS

Data collection took place from January to May 2011 (see Section 3.2). Students completed the tasks and instruments as part of their normal course activities and provided or refused consent for their results to be used in the study. The study consisted primarily of two instruments: the opinion survey and the research instrument, each described in this section. Additionally, a limited number of interviews were conducted with students at two of the research sites.

**The opinion survey** At the beginning of the semester, students completed a short survey that included some demographic information (sex, age, major) and assessed the strength of their opinions about a variety of topics (see Appendix A). There were twelve items on this survey, which were rated on a Likert scale, ranging from 1=*Strongly Disagree* to 7=*Strongly Agree*, where 4 represented a neutral opinion and 0 represented unfamiliarity with the topic.

A longer version of this survey, with over 30 items, was piloted in the Spring 2010 semester. As in the final survey, the pilot survey contained items that ranged from serious (e.g., “Marijuana should be legalized for medical purposes”) to ‘fun’ (e.g., “In the dog versus cat debate, I choose dogs”).

The results from the pilot survey were used to select particular topics that students:

- Had strong opinions about [defined as ratings of 1, 2, 6 or 7];
- Had neutral opinions about [defined as ratings of 0 or 4]; and
- Had split opinions about [roughly equal percentages of strong and neutral opinions].

Based on the pilot survey results, two topics that fell into each of these three categories were selected to provide the context for examples to be examined during the research portion of the study. For the strong opinion questions, the topics dealt with texting while driving and college tuition costs. For the neutral questions, the topics dealt with food irradiation and the amount of luggage that can be carried on a plane. For the split opinion questions, the topics dealt with the use of Mac versus PC computers and medicinal marijuana. Six additional topics from the pilot survey were retained on the final opinion survey in order to provide potential contexts for additional examples, if needed. The rest of the items on the pilot survey were dropped to reduce completion time.

**The research instrument** The six topics selected from the pilot opinion survey were developed into scenarios that served as the research instrument for this project (see Appendix B). Whenever possible, the scenarios for the research instrument were based on real studies or real data related to these topics; when this was not possible, a “realistic” scenario was created. Sometimes, details of a real study were altered for research purposes. Since the focus of this study was on the role that sample selection plays in the appropriateness of inferential conclusions and how students’ prior opinions affect their ability to judge these features, several versions of each scenario were created. These versions varied in the quality of the selection (biased versus random sampling) and, when possible, in the nature of the conclusions. For example, four versions of the ‘Mac/PC’ scenario were prepared based on real studies:

- Using a reliable sampling method, with a conclusion stating PC was preferred;
- Using a reliable sampling method, with a conclusion stating Mac was preferred;
- Using a biased sampling method, with a conclusion stating PC was preferred; and
- Using a biased sampling method, with a conclusion stating Mac was preferred.

There were times when it was not appropriate to vary the conclusion of a study, so only two scenario versions were created. For example, there were only two versions of the ‘texting while driving’ scenario, which were based on real studies. Each study found that texting while driving was dangerous, and the researchers felt it would be irresponsible to present the students with “evidence” that texting while driving was not dangerous.

Three of the six topics on the research instrument were selected for use in a testing environment, in which the scenarios were completed as part of a course exam. As the exams were given at different institutions, the structure of the exams was a bit different, ranging from 11 questions (including three that were multiple-choice) to 40 questions (including 30 that were multiple choice). However, at all institutions, the research scenarios were graded for credit. The remaining topics were reserved for one-on-one interviews with the students at two of the authors’ universities. Due to time constraints, no interviews were conducted at the other two authors’ universities. Students self-selected to participate in the interviews by responding to an email sent by the instructor of their course. Not surprisingly, few students (16, about 7%) volunteered, resulting in a much smaller sample size for the interview environment compared to the testing environment.

In the testing environment, different versions of each scenario were assigned pseudo-randomly by mixing the printed copies before passing them out to students. In the interviews, the different versions were randomly selected for each interviewee. Random assignment of topics to students, rather than

purposeful assignment, was used for two reasons. First, it would have been logistically problematic to pass out particular questions to particular students, especially during an exam. Second, the hypotheses of this research are complex, considering both how those with strong opinions and those with neutral opinions would react to each of the various contexts (see Section 3.3 for the detailed hypotheses). Random assignment of the contexts ensured receiving a broad range of opinions on each, so that the hypotheses could be fully explored, and also allowed for the discovery of results that were not expected under the hypotheses.

In both the testing and interview environments, students were asked to perform the same task: read the scenario, state whether or not the conclusion was valid, and explain their reasoning. More details on data collection are presented in the next section; results from both the testing environment and the interviews are presented in Section 4.

### 3.2. DATA COLLECTION AND CODING

As stated previously, data collection took place from January to May 2011. Since the four academic institutions differed in terms of the specific schedule for this semester, data were collected during different weeks, but at similar progression in the course material, at each institution. Table 1 summarizes this timeline.

*Table 1. Timeline of data collection*

Time Span	Instrument	When/how administered
January 2011	Opinion survey	during the first week of classes
January-March 2011	Research instrument (RI): luggage scenario (Appendix B, Topic 1)	on first midterm exam; for most students this was after sampling, but before inference, had been taught
March-April 2011	RI: tuition costs scenario (Appendix B, Topic 2)	on second midterm exam (except at one university, where it was administered in class); for most students, this was after formal inference procedures had been introduced
May 2011	RI: Mac vs. PC scenario (Appendix B, Topic 3)	on the final exam
April-May 2011	RI: Additional scenarios (Appendix B, Topics 4-6)	during individual student interviews

Ratings on the opinion survey were grouped based on the strength of a student's opinion for each statement. Ratings of 1, 2, 6 or 7 were considered strong opinions; ratings of 0 or 4 were considered neutral opinions. To keep a clear distinction between neutral and strong opinions, ratings of 3 and 5 were not included in either of these categories. The authors felt that it was not clear if these "in between" ratings were closer to neutral or to strong opinions. As such, these ratings were considered non-informative for the purposes of this research and the responses of students who gave these ratings were not included in any analyses of the testing environment scenarios.

Several grounded theory techniques were used in coding the research instrument data, including: "(a) simultaneous collection and analysis of data, (b) a two-step data coding process ... (e) sampling to refine the researcher's emerging theoretical ideas, and (f) integration of the theoretical framework." (Charmaz, 2000, p. 510). Specifically, the data coding process involved an iterative cycle of meeting to discuss codes, coding individually, and then meeting again to discuss discrepancies and different perspectives until an acceptable agreement rate was achieved.

As a result of this process, two sets of codes were developed to analyze student responses to each scenario on the research instrument. One set of codes captured the strength of a student's understanding of the issues relevant to sampling (e.g. recognizing the benefit of random sampling or

identifying possible sources of bias when a sample is not random). The second set of codes captured when a student focused more on irrelevant contextual issues or personal beliefs (e.g., “I believe people are carrying heavier bags onto airplanes”), or even on irrelevant statistical issues or misconceptions (e.g. that the sample size needs to be a large enough percentage of the population size). Both coding schemes were based on a modified Structural Observation of Learning Outcomes (SOLO) taxonomy (Biggs & Collis, 1982). Like SOLO, the coding schemes used in this research capture both the quantity and correctness (or incorrectness) of ideas in a student’s response. However, these coding schemes were modified also to capture specifically the varying degrees of statistical and contextual reasoning used by students. This modification was necessary to address the complex research hypotheses of the current study. The full coding scheme is presented in Tables 2 and 3.

*Table 2. Codes used to analyze student responses for matters relevant to sample selection (Coding set 1)*

Classification for Coding Set 1	Code Assigned
No ideas relevant to sample selection were included	0 (zero)
At least one relevant idea was included, however <u>all</u> ideas were used incorrectly (e.g., student says sample is “random” when it is not)	P
A single correct and relevant idea was included (e.g., student provided a possible source of bias but didn’t overelaborate)	U
At least two correct and relevant ideas were included	MC
At least two relevant ideas were included, some of which were correct but some of which were not	MM
The student’s entire response demonstrated clearly correct reasoning, and did not include any irrelevant information.	R

*Table 3. Codes used to analyze student responses for matters not relevant to sample selection (Coding set 2)*

Classification for Coding Set 2	Code Assigned
No irrelevant ideas were included	0 (zero)
A single irrelevant idea, based on context or opinion, was included	UC
A single irrelevant idea, based on (often incorrect) statistical reasoning, was included (e.g., student says that the sample size is too small to accurately represent a large population)	US
At least two irrelevant ideas, all based on context or opinion, were included	MC
At least two irrelevant ideas, all based on (often incorrect) statistical reasoning, were included	MS
At least two irrelevant ideas, representing a mixture of context/opinion and (often incorrect) statistical reasoning, were included	MB

*Table 4. Expected results under the research hypotheses*

	Study used random sampling methods	Study used biased sampling methods
Student had a strong opinion with which study conclusion agreed	Student believed the conclusion was valid because they agreed with it. Thus they correctly believed the study was sound.	Student believed the conclusion was valid because they agreed with it. Thus they believed the study was sound even though it used biased sampling methods.
Student had a strong opinion with which study conclusion disagreed	Student did not believe the conclusion was valid because they did not agree with it. Thus they believed the study was poorly designed even though it used random sampling methods.	Student did not believe the conclusion was valid because they did not agree with it. Thus they correctly believed the study was poorly designed.
Student had a neutral opinion	Student was not influenced by personal beliefs or opinions and therefore recognized good aspects of the sampling method. Thus they believed the study was sound and the conclusion was valid.	Student was not influenced by personal beliefs or opinions and therefore recognized biased aspects of the sampling method. They believed the study was poorly designed and the conclusion was not valid.

### 3.3. EXPECTED RESULTS

The researchers believe that, due to the influence of confirmation bias and belief bias, students who have strong opinions about a topic will evaluate the validity of a study's conclusions based on whether or not that conclusion agrees with their opinion, rather than the actual quality of the sampling methods used. Table 4 presents specific expectations for student responses based on these factors.

### 3.4. DATA ANALYSIS

After coding, responses from the testing environment and the interviews were grouped according to the strength of the student's opinion, whether or not the study conclusion agreed with that opinion, and the type of sampling method used (as in Table 4). Within these groupings, the raw codes were examined to explore students' reasoning about each scenario. The number of responses that included irrelevant contextual factors was also recorded, and served as the primary response variable for most analyses.

When sample sizes were adequate, several methods were used to determine if there was a statistically significant difference in the use of irrelevant contextual information between the groups. Logistic regression was chosen due to the nature of the data and the hypotheses; that is, the response variable (inclusion of context) is nominal and the hypotheses lead to a test for a statistically significant interaction between sampling method and strength of opinion. Once it was determined if an interaction exists, it was deemed reasonable to explore confidence intervals for relative risk to determine how the inclusion of context was different, and if so, by how much. More details on each of these procedures can be found in Section 4, where they are presented with the specific results for reader convenience.



## 4. RESULTS

### 4.1. DEMOGRAPHICS

Table 5 presents the demographics of the sample, by school and overall. Some differences between the students at the various schools can be seen. For example, the students at School 4 had a greater variation in age and were slightly older on average than the students at the other schools. This reflects the wide variety of students that attend community college. Additionally, the sample at School 2 had a much lower percentage of female students compared to the other schools. This difference was attributed to the data being collected in a business statistics class.

*Table 5. Sample demographics*

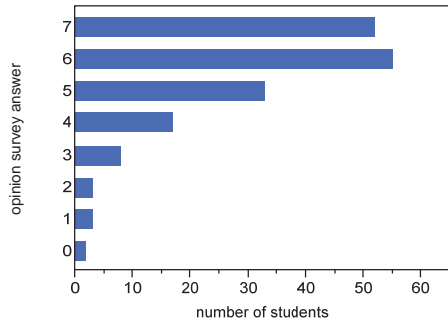
School	<i>n</i>	Mean (SD) Age in years	% Female
School 1	173	21 (3.8)	67.0%
School 2	63	20 (3.9)	36.5%
School 3	8	20 (1.2)	50.0%
School 4	33	23 (5.8)	60.6%
Entire Study	277	21 (4.1)	58.8%

### 4.2. OPINION SURVEY

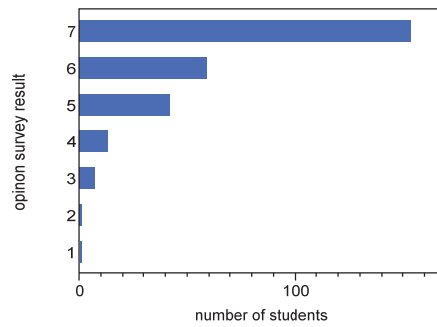
Figure 1 shows student opinion ratings for each of the six statements that provided context for the research instrument. The first row of the figure shows those contexts about which students tended to have strong opinions (i.e., providing ratings of 1, 2, 6, or 7); the second row shows those contexts about which students had split opinions (roughly equal percentages of students with strong and with neutral opinions); and the third row shows those contexts about which students tended to have neutral opinions (providing ratings of 0 or 4). The first column of the figure shows those contexts which were administered in the testing environment, and the second column shows those contexts that were administered during the interviews.

The results are consistent with the researchers' expectations about the contexts for which students would or would not have strong opinions. The sample size for the statement about college tuition is smaller than for the other topics because this question was inadvertently left off of the opinion survey for some students. The opinion survey results for those students who participated in one-on-one interviews were consistent with the results of the overall sample with the exception of the question about irradiation, which showed split results in the interview group and neutral results in the overall sample.

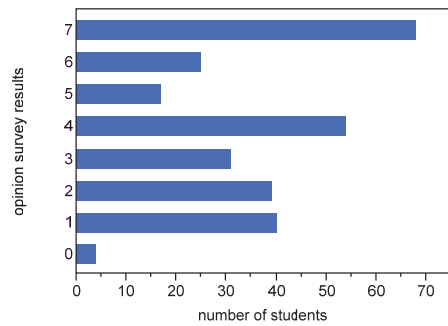
Tuition costs in the US are too expensive at both public and private institutions. ( $n = 173$ )



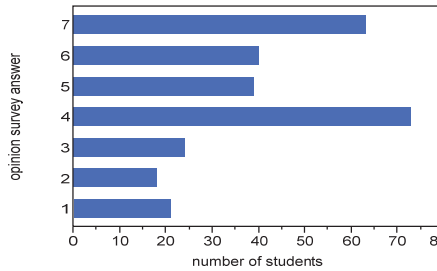
Texting while driving is dangerous. ( $n = 276$ )



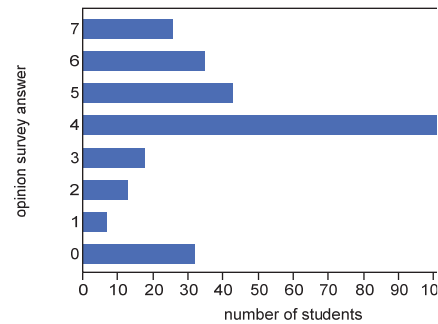
In the Mac versus PC debate, I choose Mac. ( $n = 277$ )



Marijuana should be legalized for medicinal purposes in the US. ( $n = 277$ )



The amount of luggage that can be carried by a plane needs to be reassessed. ( $n = 277$ )



Foods that I eat should be irradiated for health reasons. ( $n = 276$ )

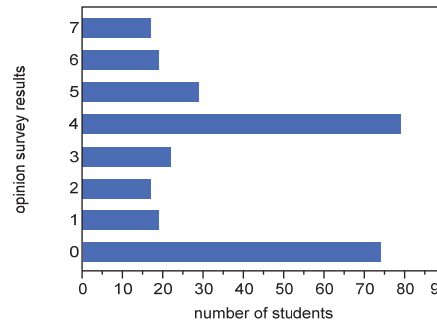


Figure 1. Results from spring 2011 opinion survey items

### 4.3. RESEARCH INSTRUMENT

**Testing environment results** The scenarios that were selected for use in the testing environment included the luggage, tuition, and Mac/PC scenarios. Each student completed one randomly selected version of the scenarios for each of the selected topics. To explore the expectations expressed in Section 3.3, the student responses were grouped according to the strength of the student’s opinion, whether or not the study conclusion agreed with that opinion, and the type of sampling method, as in Table 4. As previously indicated, the number of responses within each of these groups that included irrelevant contextual factors was recorded. Tables 6, 7, and 8 present the results of these analyses for the luggage, tuition, and Mac/PC scenarios, respectively. Examples from the testing environment results that were considered irrelevant contextual information are:

- (Luggage scenario with random sampling and conclusion that excess weight on airplanes is a problem) "...they may be getting more samples from a preppy woman going to France who tries to stuff a lot of useless things in her bag than a business man who is only carrying his laptop."
- (Luggage scenario with biased sampling and conclusion that excess weight on airplanes is a problem) "There are rich people in Orlando meaning they will want more things."
- (Tuition scenario with biased sampling and conclusions that tuition has not gone up) "...because schools raise prices every year, it's a given fact. Also, it has only been one year, if they would have done a larger sample such as 3 or 4 years, the numbers would have been different."
- (Mac/PC scenario with random sampling and conclusion that Mac is preferred) "It is common knowledge that while Macs are steadily gaining popularity, PCs still dominate the market as a whole. Also, something to consider is most college students don't have a lot of money and Macs are considerably more expensive."
- (Mac/PC scenario with random sampling and conclusion that PC is preferred) "... but more than likely it is because of cost that swayed their decision."

*Table 6. Inclusion of irrelevant context on testing environment responses when strength of opinion and study conclusions are considered (luggage scenario)*

Luggage scenario	Study uses random sampling methods	Study uses biased sampling methods	Total
Student has a strong opinion with which study conclusion agrees	3 out of 18 (16.7%)	5 out of 23 (21.7%)	8 out of 41 (19.5%)
Student has a strong opinion with which study conclusion disagrees	2 out of 15 (13.3%)	3 out of 16 (18.8%)	5 out of 31 (16.1%)
Student has a neutral opinion	6 out of 65 (9.2%)	13 out of 59 (22%)	19 out of 124 (15.3%)
Total	11 out of 98 (11.2%)	21 out of 98 (21.4%)	32 out of 196 (16.3%)

*Table 7. Inclusion of irrelevant context on testing environment responses when strength of opinion and study conclusions are considered (tuition scenario)*

Tuition scenario	Study uses random sampling methods	Study uses biased sampling methods	Total
Student has a strong opinion with which study conclusion agrees	1 out of 22 (4.5%)	1 out of 25 (4%)	2 out of 47 (4.3%)
Student has a strong opinion with which study conclusion disagrees	5 out of 28 (17.9%)	4 out of 28 (14.3%)	9 out of 56 (16.1%)
Student has a neutral opinion	0 out of 4 (0%)	0 out of 12 (0%)	0 out of 16 (0%)
Total	6 out of 54 (11.1%)	5 out of 65 (7.7%)	11 out of 119 (9.2%)

Table 8. Inclusion of irrelevant context on testing environment responses when strength of opinion and study conclusions are considered (Mac/PC scenario)

Mac/PC scenario	Study uses random sampling methods	Study uses biased sampling methods	Total
Student has a strong opinion with which study conclusion agrees	8 out of 34 (23.5%)	5 out of 42 (12%)	13 out of 76 (17.1%)
Student has a strong opinion with which study conclusion disagrees	14 out of 43 (32.6%)	5 out of 37 (13.5%)	19 out of 80 (23.8%)
Student has a neutral opinion	5 out of 28 (17.9%)	3 out of 23 (13%)	8 out of 51 (15.7%)
Total	27 out of 105 (25.7%)	13 out of 102 (12.7%)	40 out of 207 (19.3%)

Aggregating across the rows and columns of Tables 6, 7 and 8, respectively, it can be seen that students used irrelevant context in their responses 16.3% of the time in the luggage scenario, 9.2% of the time in the tuition scenario, and 19.3% of the time in the Mac/PC scenario. It may seem surprising that students used the least amount of irrelevant context for the tuition scenario because they tended to feel strongly about the topic. However, this was the only question that included some “technical” information, namely a  $p$ -value. As a result, many students stated that since the conclusion was consistent with the  $p$ -value, it must be valid. In fact, for 69 students (30%), this was the only comment in their response. The exam including the tuition scenario was given soon after the students learned about  $p$ -values, so this may have been the result of  $p$ -values being the statistical concept on which the students were focused at the time of the exam.

For each of Tables 6-8, logistic regression models were run with inclusion of irrelevant context (yes/no) as the response and sampling method (biased/random) and strength of opinion (using the row headings of each table as the categories) as predictors. For each model, an interaction between sampling method and strength of opinion was included, since the hypotheses presented in Section 3.3 suggest one might exist. However, for each of the three testing environment scenarios, the interaction term was not found to be statistically significant at the 10% level (results not shown) and so it was dropped and the models refit with main effects only. Significance of the main effects varied, with at least one of the factors being not statistically significant for each scenario. In fact, neither sampling method nor strength of opinion was statistically significant for the luggage scenario. The final logistic models for the tuition and Mac/PC scenarios, after dropping the predictors that were not statistically significant, are presented in Table 9.

Table 9: Final logistic models for the tuition and Mac/PC scenarios

	Estimate	Std. error	df	$t$ -statistic	$p$ -value
<i>Model for tuition scenario (referencing Table 7)</i>					
Intercept	<0.0001	0.07143	116	0.00	1.0000
Strong opinion, conclusion agrees	0.0426	0.08269	116	0.51	0.6078
Strong opinion, conclusion disagrees	0.1607	0.08099	116	1.98	0.0496
<i>Model for Mac/PC scenario (referencing Table 8)</i>					
Intercept	0.2571	0.03819	205	6.73	<0.0001
Biased sampling methods used	-0.1297	0.05441	205	-2.38	0.0181

For the tuition scenario, sampling method did not appear to influence students’ inclusion of irrelevant context in their responses. To explore further the role of strength of opinion, relative risks

were calculated. Table 7 indicates that no students who had a neutral opinion included irrelevant context in their responses. Thus the risks of *not* including context were calculated to compare the rows of this table. These relative risks (and 95% confidence intervals) are:

- Comparing those who had a strong opinion with which the study conclusion agreed to those for whom the conclusion disagreed: 1.14 (1.002, 1.299)
- Comparing those who had a strong opinion with which the study conclusion agreed to those with neutral opinions: 0.96 (0.901, 1.017)
- Comparing those who had a strong opinion with which the study conclusion disagreed to those with neutral opinions: 0.84 (0.741, 0.941)

The first and last relative risks are significant at the 5% level, though the first only barely so. The second relative risk is marginally not statistically significant. From this, it can be seen that those who had a strong opinion with which the study conclusion disagreed were *less* likely to include a response *without* irrelevant context than either those for whom the study conclusion agreed with their strong opinion or those with neutral opinions.

For the Mac/PC scenario, strength of opinion did not appear to influence students' inclusion of irrelevant context in their responses. Again, relative risks were calculated to explore further the role of the significant predictor—sampling method, in this case. Unlike the tuition scenario, the risk of actually including irrelevant contextual information could be calculated. The relative risk (95% confidence interval) of including irrelevant contextual information when the sample was biased compared to when the sample was random was estimated to be 0.50 (0.27, 0.91). Thus, students who received a random sample scenario were between 1.10 and 3.70 times as likely to give a response with irrelevant context as those with a biased sample scenario.

In addition to looking at each scenario separately, the data were aggregated and the students were grouped according to the total number of strong opinions about the research scenario topics that they indicated on the opinion survey. The goal was to determine if students who tended to be opinionated used more irrelevant contextual information as compared to students who tended to have neutral opinions. Table 10 shows the percent of students in each of these groups who included irrelevant contextual information in any of their responses, as well as the total number of students in each group.

Within the groups presented in Table 10, the results of those students who provided mixed (some strong, some neutral) ratings are explored further to see if they included more irrelevant contextual information for those contexts which they rated strongly (e.g., did a larger percentage of those students who provided a strong opinion rating only for the tuition context include irrelevant contex-

*Table 10. Inclusion of irrelevant context on testing environment responses when only strength of opinion is considered*

Students' opinion ratings	Study uses random sampling methods	Study uses biased sampling methods	Total
Students who provided strong ratings for all three contexts	1 out of 18 (5.6%)	4 out of 28 (14.3%)	5 out of 46 (10.9%)
Students who provided 1 neutral, 2 strong ratings	15 out of 70 (21.4%)	11 out of 54 (20.4%)	26 out of 124 (21.0%)
Students who provided 1 strong, 2 neutral ratings	4 out of 27 (14.8%)	3 out of 28 (10.7%)	7 out of 55 (12.7%)
Students who provided all neutral ratings	1 out of 4 (25.0%)	0 out of 6 (0%)	1 out of 10 (10.0%)
Total	21 out of 119 (17.6%)	18 out of 116 (15.5%)	39 out of 235 (16.6%)

tual information on the tuition scenario but not the other two scenarios?). Due to small sample sizes in the resulting cells, results (not shown) showed little use of irrelevant contextual information across any scenario.

Finally, the students' patterns of reasoning were explored as indicated by their coded responses for each scenario, as shown in Tables 11, 12, and 13, to see if they revealed any additional insight into the students' reasoning about sampling for each scenario. In each of Tables 11-13, the codes are presented in the following format: [Code from set 1, Code from set 2]. For example, the notation [MC, 0] indicates a response for which the student included multiple correct ideas that are relevant to sampling (coding set 1) and no information that is irrelevant to sampling (coding set 2). Similarly, the notation [U, UC] indicates a response for which the student included a single correct idea that was relevant to sampling and also a single irrelevant, contextual idea. Tables 2 and 3 provide the full list of codes. Please note that in each of Table 11, 12, and 13, the sample sizes apply to the entire cell in which they are provided (for example, there were 18 students who belong in the first cell of Table 11, 22% of which had their response coded as [MC, 0]). Also note that coding combinations that were given for fewer than 10% of responses are not shown.

Table 11. Common codes for the luggage scenario

	Study uses random sampling		Study uses biased sampling	
	Reasoning Code	% of responses	Reasoning Code	% of responses
	<i>n</i> =18		<i>n</i> =23	
Student has a strong opinion with which study conclusion agrees	[MC, 0]	22.2%	[MM, 0]	30.4%
	[P, 0]	22.2%	[MC, 0]	13.0%
	[P, US]	11.1%	[P, 0]	13.0%
	[MM, 0]	11.1%	[U, 0]	13.0%
	<i>n</i> =15		<i>n</i> =16	
Student has a strong opinion with which study conclusion disagrees	[MC, 0]	40.0%	[MM, 0]	31.3%
	[P, 0]	20.0%	[MC, 0]	18.8%
	[R, 0]	13.3%	[MM, US]	12.5%
	<i>n</i> =65		<i>n</i> =59	
Student has a neutral opinion	[P, 0]	30.8%	[MM, 0]	25.4%
	[MC, 0]	21.5%	[MC, 0]	22.0%
	[MM, 0]	10.8%	[MM, UC]	10.2%

Table 12. Common codes for the tuition scenario

	Study uses random sampling		Study uses biased sampling	
	Reasoning Code	% of responses	Reasoning Code	% of responses
	<i>n</i> =22		<i>n</i> =25	
Student has a strong opinion with which study conclusion agrees	[U, US]	27.3%	[0, US]	32.0%
	[0, US]	22.7%	[U, US]	20.0%
	<i>n</i> =28		<i>n</i> =28	
Student has a strong opinion, with which study conclusion disagrees	[0, US]	42.9%	[U, 0]	28.6%
	[0, MB]	14.3%	[MC, 0]	10.7%
		<i>n</i> =4		<i>n</i> =12
Student has a neutral opinion			[U, 0]	25.0%
	[0, US]	75.0%	[MM, 0]	16.7%
	[P, 0]	25.0%	[0, MS]	16.7%
			[0, US]	16.7%
			[MC, US]	16.7%

Table 13. Common codes for the Mac/PC scenario

	Study uses random sampling		Study uses biased sampling	
	Reasoning Code	% of responses	Reasoning Code	% of responses
Student has a strong opinion with which study conclusion agrees	<i>n</i> =34		<i>n</i> =42	
	[MM, 0]	14.7%	[MC, 0]	35.7%
	[P, 0]	14.7%	[MM, 0]	16.7%
	[U, 0]	11.8%	[U, 0]	14.3%
Student has a strong opinion with which study conclusion disagrees	<i>n</i> =43		<i>n</i> =37	
	[P, 0]	23.3%		
	[P, UC]	11.6%	[MM, 0]	29.7%
	[MC, 0]	11.6%	[MC, 0]	27.0%
	[0, MC]	11.6%		
Student has a neutral opinion	<i>n</i> =28		<i>n</i> =23	
	[P, 0]	32.1%	[MM, 0]	30.4%
	[U, 0]	14.3%	[MC, 0]	26.1%
			[U, 0]	17.4%

**Results by school** The testing environment results were also examined by school. The institutions at which this research was conducted tend to attract different populations of students. The authors were curious to see if this would in turn translate to differences in reasoning among the study participants. Unfortunately, sample sizes for Schools 3 and 4 were too small to interpret; thus, school-specific comparisons were restricted to Schools 1 (a statistics PhD granting institution) and 2 (a regional statistics BS granting institution). Examining Tables 6-8 between these schools revealed a few differences in the percentage of students who included irrelevant context in their responses (results not shown). In particular, students at School 2 tended to use more irrelevant context for the Mac/PC scenario when they had a strong opinion, regardless of whether or not the study conclusion agreed with that opinion. For the luggage scenario, students at School 1 tended to use less irrelevant context when they had a strong opinion with which the conclusion agreed, a departure from the overall pattern observed in Table 6. Correspondingly, students at School 2 tended to include more irrelevant context in this situation. These differences could be explained by differences between the samples: there were more male students at School 2 than at any of the other schools, and the course focused on business statistics whereas the other courses were more general.

Unfortunately, it was difficult to examine Tables 11-13 by school. The numbers of students in each cell were small and many different codes were observed. As a result, the percentages for each code for each school were greatly influenced by only one or two students.

**Interview results** A total of 16 students were interviewed at Schools 2 and 4. Each interviewee was asked to respond to two scenarios randomly selected from the following contexts: legalizing marijuana, irradiating food, and texting while driving. Additionally, two students received the Mac/PC scenario in an interview setting as well as the test setting. Due to the comparatively small sample sizes for the interviews, the analyses presented here do not exclude the responses of students who gave a rating of 3 or 5 on the corresponding question on the opinion survey.

The student responses to the interview scenarios were classified using the same modified SOLO scheme as presented in Tables 2 and 3. Again, any students who had codes of UC, MC, or MB for coding set 2 were considered as having included irrelevant context in their response. Examples of irrelevant context that arose in the interview setting are:

- (Food irradiation scenario with biased sampling and conclusion that food should be irradiated) “There are health issues...just eat them fresh.”
- (Food irradiation scenario with biased sampling and conclusion that food should be irradiated) “I think they should have tested more and tested them years down the road to see if anyone who ate them got anything from it. I don’t know that I’d eat strawberries with gamma rays.”

- (Texting while driving scenario with biased sampling and conclusion that it is dangerous) “You text and are distracted and don’t see the road, so definitely I could see why that would be like that. And, you know, some of my friends text and drive. Sometimes I’m worried they miss something. It’s like ‘Dude, you see that? Put your phone down.’ It’s bad enough when you call people but texting you got to look. You look at what you are texting and look down and up and down and up. That’s just not safe.”
- (Texting while driving scenario with biased sampling and conclusion that it is dangerous) “...more believable plus I’ve done it myself. I have never crashed but it is very hard to text and drive. I tell people ‘don’t text me because I can’t do it; it takes my attention away’.”
- (Texting while driving scenario with biased sampling and conclusion that it is dangerous) “I agree with it, that texting while driving definitely increases accidents. But I feel like that it doesn’t say exactly who is driving these U-Hauls or the age limit because it can change depending on the age. If it’s younger they are more likely to be texting and especially in long haul trucks it’s going to be tough. If you did it maybe in rental cars depending on which age limit they allow it could give you a little more of a verified answer, but I would agree with the conclusion.”
- (Texting while driving scenario with biased sampling and conclusion that it is dangerous) “...texting while driving is not good. My daughter was in the car the other day and said ‘Mom’ and I wasn’t texting but she thought I just picked up my phone and she was like ‘don’t do anything while driving.’ But yeah it takes your attention away a lot from driving.”
- (Medical marijuana scenario with biased sampling and conclusion that it should be legalized) “People could be supporting it not for medicinal purposes. Or people just felt like when they do the drug test in high school and they just put whatever answer.”
- (Medical marijuana scenario with random sampling and conclusion that it should be legalized) “I know a lot of people will just say that they want marijuana legal for medical purposes but have false answers I think.”

Table 14 shows a comparison between the results for the testing and interview scenario for the 16 students who were interviewed. The table includes a summary of student opinion, sampling method, and the percentage of testing versus interview scenarios that used irrelevant context.

*Table 14. Summary of student opinion, sampling method, environment and inclusion of irrelevant context*

	Testing Environment			Interview Setting		
	Random Sampling	Biased Sampling	Total	Random Sampling	Biased Sampling	Total
Student has a strong opinion with which study conclusion agrees	0 of 8 (0%)	0 of 3 (0%)	0 of 11 (0%)	8 of 9 (89%)	6 of 11 (55%)	14 of 20 (70%)
Student has a strong opinion with which study conclusion disagrees	2 of 9 (22%)	1 of 5 (20%)	3 of 14 (21%)	1 of 1 (100%)	1 of 2 (50%)	2 of 3 (67%)
Student has a neutral opinion	1 of 7 (14%)	0 of 7 (0%)	1 of 14 (7%)	2 of 2 (100%)	2 of 3 (67%)	4 of 5 (80%)
Student has “moderate” opinion (neither neutral nor strong)	1 of 5 (20%)	0 of 2 (0%)	1 of 7 (14%)	0 of 0	3 of 4 (75%)	3 of 4 (75%)
Total	4 of 29 (14%)	1 of 17 (6%)	5 of 46 (11%)	11 of 12 (92%)	12 of 20 (60%)	23 of 32 (72%)



The remainder of this section focuses on the subset of 16 students who have both testing and interview results (no student participated in an interview without also participating in the testing environment research). These students completed a total of 46 test scenarios and 32 interview scenarios. Of the 46 test environment scenarios, 5 (11%) responses contained irrelevant context. Of the 32 interview scenarios, 23 (72%) responses contained irrelevant context—a significantly higher percentage for the testing scenarios ( $p$ -value < 0.0001). In fact, a 95% confidence interval of the relative risk finds that students were between 4.56 to 9.39 times as likely to use irrelevant context during interviews compared to the testing environment.

Responses from the testing and interview scenarios were also compared on the individual level. Thirteen of the sixteen students (81%) used more irrelevant context in their responses to the interview scenarios than in their responses to the testing scenarios. A paired analysis of the data for each student using the non-parametric sign test was statistically significant ( $p$ -value = 0.0037).

To take the strength of the students' opinions, agreement of the conclusion with those opinions, and the type of sampling method used into consideration, each student's test and interview data were compared to determine whether the student had been assigned scenarios with matching opinion (using the row headings of Table 14 as the four categories) and sampling methods (random/biased). Nine of the sixteen students were given scenarios that matched in terms of these variables between the test and interview scenarios. Out of these nine students, five (56%) used irrelevant context in the interview setting but not in the testing environment.

As an example of how students appeared to use more irrelevant context during the interview, responses for a particular student (Student 1) are presented in detail below. This student completed the Mac/PC question in both the testing and interview settings. (Note: this student did not consent to audiotape recording during the interviews, so his responses were summarized by the interviewer.) This student had a very strong opinion on the Mac/PC context – he strongly disagreed with preferring a Mac, thus he strongly preferred a PC. In the testing scenario, his question had a biased sample that resulted in the conclusion that students prefer Macs over PCs. His response was: “I think it is a valid conclusion because the sample size is pretty big, giving us a better answer.” (Coded as [P, 0])

In the interview setting, Student 1 had a biased sample that resulted in the conclusion that students prefer PCs over Macs and his response was:

“62.5% is the largest number ... Yes – it is valid...most businesses use it...it is preferred. ... Mac is mostly for college ‘artsy’ students. ... PC has better programs—I’m a PC user!” (Coded as [0, MC])

This student responded to the same context very differently in a test setting as compared to an interview setting. During the testing scenario, he focused on sample size and did not respond based on his opinion or prior beliefs. However, in the interview setting he explained his reasoning in much more detail, and it can be seen that this reasoning was motivated by his context-based opinions.

Another example to note was in response to the food irradiation scenario (with random sampling and conclusion that food should be irradiated). During the interview, a different student seemed to address explicitly the division between statistical and contextual reasoning by asking, “Do you want me to just say that it is a valid conclusion or would you want my opinion with the gamma rays?”

## 5. DISCUSSION

### 5.1. THE TESTING ENVIRONMENT

For the luggage scenario (Table 6), students tended to include more context when they had a strong opinion with which the study agreed, and among those, when biased sampling methods were used. However, results of a logistic regression model (not shown) indicate that these differences were not statistically significant. Interestingly, the codes presented in Table 11 indicate that students generally wrote longer answers for this scenario, as there was a higher proportion of responses with an “M” designation (indicating multiple responses) for coding sets 1 and 2. It may be that the bias was so obvious that students were able to write a lot about it, or perhaps they were “bored” by the context and not sure what specifically to address. This scenario also seemed to elicit more incorrect responses: having a higher proportion of “P” or “MM” codes for coding set 1, as well as a higher proportion of

non-zero codes for coding set 2. Not surprisingly, it seems that the more students wrote, the more likely they were to include something wrong.

In contrast to the luggage scenario, for both the tuition and Mac/PC scenarios students were more likely to use irrelevant context when they had a strong opinion with which the study conclusion disagreed. Among these students (the middle rows in Tables 7 and 8), there also tended to be more irrelevant context included when the sample was random, which offers some support for our hypotheses. The results found from the Mac/PC scenario are consistent with those found on the cars question posed by Watson and Moritz (2000a), as students had a tendency to base their answer on their own opinion about the brand of computer rather than the sampling technique used. In addition, the results of the tuition scenario are consistent with Jacobs (1999) and Konold and Higgins (2003) in that students relied more on their own personal experiences with tuition than on the sampling methods employed. However, evidence of statistical significance for these differences was mixed, weakening this support. Similarly, the fact that the hypothesized interaction between sampling method and opinion was not statistically significant for each of the scenarios seems to indicate that there was not overwhelming evidence of confirmation bias or belief bias based on the brand of computer data.

Focusing just on strength of opinion, Table 10 indicates that students who provided two or three strong opinion ratings included a higher percentage of irrelevant context in their responses overall than did their more “neutral” classmates (i.e., students who provided no more than one strong opinion rating for the testing environment scenarios). This can also be seen by looking at the codes in the last row of Tables 11-13: nearly all of the common codes have zero or “S” designations for coding set 2, indicating little use of irrelevant contextual information among students with neutral opinions. This does provide some support for the hypotheses.

Several other interesting patterns were observed when looking at Tables 11-13. First, no single pair of codes seemed to dominate any of the tables. The lone exception to this was for the tuition scenario, where the coding pair [0, US] was frequently used. This was the code assigned to responses that focused solely on the *p*-value, as discussed previously. Second, there was little inclusion of irrelevant information (contextual or statistical) overall, as indicated by the fact that most codes had a zero for coding set 2 (again, the tuition scenario was an exception to this). On one hand, this is encouraging: the luggage scenario did not elicit much irrelevant context, which is consistent with the hypothesis that students would not be distracted by the neutral context. However, the Mac/PC scenario, about which a reasonable proportion of students did hold strong opinions, did not elicit much irrelevant contextual information either.

While certain results are consistent with the biasing effects of context seen in previous research, the overall support for the hypotheses explored in this study is somewhat weak, as there was simply a general lack of irrelevant contextual information contained in student responses to the testing environment scenarios. However, this may be due to a limitation in the study design (see Section 6.1).

## 5.2. THE INTERVIEWS

In contrast with the testing environment results, student responses to the interview scenarios contained much more irrelevant contextual information. Moreover, the increased use of irrelevant context occurred regardless of strength of opinion. This is consistent with the general body of literature on confirmation bias and belief bias—that people have difficulty interpreting empirical evidence based on an objective evaluation of the statistical strength of the study design—but again, it is not consistent with the hypothesized expectation that the effect of this bias would depend on strength of opinion. Rather, the effect of this bias appeared to depend on the environment in which reasoning about empirical evidence took place. The quote from Student 2 (presented at the end of Section 4) points explicitly to the division between statistical and contextual reasoning. This study provides evidence that students appear to reason statistically in their statistics courses, but appear to forget or ignore statistical reasoning once outside of the classroom.

## 6. CONCLUSIONS

### 6.1. LIMITATIONS

Through the process of analyzing both the testing environment responses and the interview responses, a possible limitation of the study design was noted. Specifically, there was limited depth to students' responses in the testing environment; therefore, it was very difficult to understand students' reasoning entirely. Since it was not possible to ask the students to clarify their reasoning during the exam, the researchers were often left wondering why a student responded in a certain way.

The decision was made to use exams to collect data for several reasons. First, as described above, the researchers wanted to investigate whether the students would respond differently, depending on the school at which they were enrolled. As such, it was necessary to obtain data from as many students as possible at each institution. Collecting data through the use of course exams resulted in many more students agreeing to participate in the study than if the study had consisted solely of interviews that would require students to participate outside of class. In addition to anticipating lower response rates for interviews, the researchers were also concerned that those students who did volunteer to be interviewed might be more opinionated than those who did not volunteer. Thus, the use of data collection in the testing environment allowed for a broader range of student opinions to be sampled. Finally, one of the authors had used this type of data collection previously with reasonable success in getting "rich" responses from students. It is possible that the nature of this study—looking for evidence of confirmation bias and belief bias—was so subtle that this form of data collection did not provide data that were as revealing of student thought and motivations as the researchers might have hoped. In contrast, the interviews gave the researchers the benefit of being able to ask students for clarification of their thought processes, resulting in stronger evidence of confirmation bias and belief bias.

### 6.2. IMPLICATIONS

Statistics educators are encouraged to use real-life examples and data sets in their courses to provide students with motivation for understanding how statistical concepts are applied in the real world (ASA, 2005). However, the use of data in context has caused concern about students' abilities to separate statistical reasoning from their personal opinions and prior beliefs as a result of confirmation bias and belief bias. Despite this study's potential limitation (described above), the data collected in the testing environment did identify an important "lower bound" of sorts to the effects of confirmation bias and belief bias on statistical reasoning about sampling. These biases did not appear to have a severe effect on reasoning when students are in a "high stakes" educational setting. This result was somewhat surprising to the authors, who, based on the literature about these biases, expected their effects to overpower statistical reasoning regardless of the environment in which the reasoning took place. This result is encouraging, though, as it would have been a "worst-case scenario" if students had shown strong evidence of confirmation bias and belief bias overriding statistical reasoning on a test in a statistics course.

Unfortunately, the effects of confirmation bias and belief bias were observed in the less formal interview setting. This is troubling, as many statistics educators recognize the widespread need for developing statistical literacy and reasoning (e.g., Moreno, 2002; Utts, 2003) that lasts beyond the statistics classroom. Educators hope that students will take what they have learned in their course and apply it in the "real world." However, this paper adds to the body of evidence that students struggle with reasoning about sampling, relying on personal experience and belief instead of the principles they have learned in their course. The fact that the results presented in this paper showed evidence of increased use of contextual reasoning in the interview environment is especially concerning, as the students were interviewed by their statistics professor either during or shortly after completion of their statistics course. If they are already "forgetting" what they learned in the classroom, what will happen to these students once they are farther removed from their learning experience?

### 6.3. SUGGESTIONS FOR FUTURE RESEARCH

Future research on this topic would benefit from greater use of a data collection method in which the participants provide more in-depth responses or where researchers can probe further into students' reasoning, such as during an interview or an in-class discussion or essay. In addition, having students investigate two scenarios on each topic (one with random sampling, the other with biased sampling) may help the researchers discern what students are thinking when they have to compare and contrast the two. When choosing scenarios, it may be better to select only those options for which it is likely for students to include many "for" and "against" opinions, rather than one-sided contexts such as the texting while driving scenario. This could better allow the researcher to contrast students' reasoning across the range of strong opinions within a particular topic.

Future research should also further explore the effect of environment on student reasoning about sampling. In the testing environment, students' reasoning was based more on statistics (albeit sometimes incorrectly applied) than on confirmation bias or belief bias. Would this still be true if students' responses were collected during an in-class discussion or activity rather than on an exam? Can we explore further the reasoning process used by students on an exam and, if so, can we replicate this process in less formal settings? Is incorrect statistical reasoning motivated by poor understanding of statistical principles, or may it in fact be motivated by confirmation bias and belief bias? Finding answers to these questions would help statistics educators better understand the impact of these biases, and may help identify ways to minimize their impact on students' reasoning in the real world.

### ACKNOWLEDGEMENTS

This project was sponsored by the Consortium for the Advancement of Undergraduate Statistics Education Making Outreach Sustainable (CAUSEmos) program (NSF DUE Award No. 0618790). Special thanks to CAUSE Research Advisory Board members who guided us: Felicity Enders (Mayo Clinic), Randy Groth (Salisbury University), and Hollylynn Lee (North Carolina State University). The authors also thank the anonymous reviewers for helpful feedback in revising this manuscript.

### REFERENCES

- Allport, G. (1954). *The nature of prejudice*. Reading, MA: Addison Wesley.
- American Statistical Association (2005). *Guidelines for assessment and instruction in statistics education (GAISE) college report*. Washington, DC: ASA.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcome)*. New York: Academic Press.
- Charmaz, K. (2000). Grounded theory: Objectivist and constructivist methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research Vol. 2* (pp. 509–535). Thousand Oaks, CA: Sage Publications.
- delMas, R., Garfield, J., & Zieffler, A. (2009). The tyranny of context. Paper presented at The International Collaboration for Research on Statistical Reasoning, Thinking, and Literacy, Brisbane, Australia.
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Jacobs, V. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School*, 5(4), 240–263.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
- Kaplan, J. (2009). Effect of belief bias on the development of undergraduate students' reasoning about inference. *Journal of Statistics Education*, 17(1), 1–12.
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, 56(4), 28–55.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.

- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59–98.
- Lampen, E. (2010). Structuring contexts for statistical treatment: Initializing statistical reasoning. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference of Teachers of Statistics*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute.  
[ Online: [http://iase-web.org/documents/papers/icots8/ICOTS8\\_3D2\\_LAMPEN.pdf](http://iase-web.org/documents/papers/icots8/ICOTS8_3D2_LAMPEN.pdf) ]
- Langrall, C., Nisbet, S., & Mooney, E. (2006). The interplay between students' statistical knowledge and context knowledge in analyzing data. Paper presented at the Seventh International Conference on Teaching Statistics, Salvadore, Brazil.
- Lord, C, Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, 4(1), 23–63.
- Moore, D., & Cobb, G. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly*, 107(7), 615–630.
- Moreno, J. (2002). Toward a statistically literate citizenry: What statistics everyone should know. In *Proceedings of the 6<sup>th</sup> International Conference on Teaching Statistics*. Capetown, South Africa: International Statistical Institute.  
[ Online: [http://iase-web.org/documents/papers/icots6/1b6\\_more.pdf](http://iase-web.org/documents/papers/icots6/1b6_more.pdf) ]
- Munro, G. D. (2010). The scientific impotence excuse: Discounting belief-threatening scientific abstracts. *Journal of Applied Psychology*, 40(3), 579–600.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Rossman, A. J. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5–19.  
[ Online: [http://iase-web.org/documents/SERJ/SERJ7\(2\)\\_Rossman.pdf](http://iase-web.org/documents/SERJ/SERJ7(2)_Rossman.pdf) ]
- Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics Vol. I* (pp. 314–319). Voorburg, Netherlands: International Statistical Institute.  
[ Online: <http://iase-web.org/documents/papers/icots3/BOOK1/A9-4.pdf> ]
- Russell, S. J., Schifter, D., & Bastable, V. (2002). *Developing mathematical ideas: Working with data*. Parsippany, NJ: Dale Seymour.
- Schwartz, D., Goldman, S., Vye, N., Barron, B., & Cognition and Technology Group at Vanderbilt (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching and assessment in grades K-12* (pp. 233–274). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74–79.
- Watson, J., & Moritz, J. (2000a). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19(1), 109–136.
- Watson, J., & Moritz, J. (2000b). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44–70.
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.  
[ Online: [http://iase-web.org/documents/SERJ/SERJ7\(2\)\\_Zieffler.pdf](http://iase-web.org/documents/SERJ/SERJ7(2)_Zieffler.pdf) ]

JACQUELINE WROUGHTON  
Northern Kentucky University  
Department of Mathematics & Statistics  
Nunn Drive  
Highland Heights, KY 41099

## APPENDIX A

The opinion survey that was administered to all students in the sample during spring 2011 is shown in its entirety below.

*The following is a short survey designed to elicit your opinion on a variety of topics. Your responses will be completely confidential.*

Sex (circle one):      Female              Male

What is your age (in years)? \_\_\_\_\_

What is your major? \_\_\_\_\_

Opinion Statements: Each item has 7 possible responses, ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). If you have no opinion, choose response 4 (Neutral). If you are completely unfamiliar with a topic, please choose response 0 (at the far right). *Please read each statement, and mark the one response that most clearly represents your degree of agreement or disagreement with a statement. Do not spend too much time thinking about each statement. Record your answer and move quickly to the next statement. Please respond to all statements.*

Topic	Strongly Disagree			Neutral			Strongly Agree	Not familiar with Topic
1. In the Pepsi versus Coke debate, I choose Coke.	1	2	3	4	5	6	7	0
2. Marijuana should be legalized for medicinal purposes in the United States.	1	2	3	4	5	6	7	0
3. The amount of luggage that can be carried by a plane needs to be reassessed.	1	2	3	4	5	6	7	0
4. In the Mac vs. PC debate, I choose Mac.	1	2	3	4	5	6	7	0
5. Texting while driving is dangerous.	1	2	3	4	5	6	7	0
6. Foods that I eat should be irradiated for health reasons.	1	2	3	4	5	6	7	0
7. The legal drinking age should be lowered in the U.S.	1	2	3	4	5	6	7	0
8. Violent video games encourage violent behavior.	1	2	3	4	5	6	7	0
9. The US government should ensure that all people have health insurance.	1	2	3	4	5	6	7	0
10. Boys are better at math than girls.	1	2	3	4	5	6	7	0
11. Tuition costs in the US are too expensive at both public and private institutions.	1	2	3	4	5	6	7	0
12. In the cat versus dog debate, I choose dogs.	1	2	3	4	5	6	7	0

## APPENDIX B

The research instruments are shown in their entirety below. The first set of scenarios was used as the research instrument in the testing environment and the second set of scenarios was used for interviews.

### *Testing Environment Scenarios*

*Topic 1: Luggage [“The amount of luggage that can be carried by a plane needs to be reassessed.” Based on previous results, students tend to have neutral opinions on this topic]*

#### *Problem setup for all scenarios:*

Most airlines limit the number and weight of carry-on bags that can be brought on a plane by a traveler. For example, American Airlines allows one free carry-on bag and one free personal item (such as a purse or laptop bag), for a maximum weight of 40 pounds per ticketed passenger. However, it seems lately like people are trying to bring larger (and heavier) bags onto a plane. An airline safety advocacy group is concerned that additional baggage could lead to excess weight. [see scenarios for rest of problem]

#### *Scenario 1 (Random sampling; No problem with weight):*

They randomly select 20 airports across the country, and at each of these airports, they systematically select every 10th passenger passing through American Airlines domestic security check-points and weigh their carry-on items. They select passengers until they have a total sample size of 1,000 people (50 from each airport). From this survey, the average weight of carry-on items is found to be 35 pounds per passenger, and the group concludes that there is no problem with excess weight on domestic American Airlines flights. Do you think this is a reasonable conclusion? Why or why not?

#### *Scenario 2 (Random sampling; Problem with weight):*

They randomly select 20 airports across the country, and at each of these airports, they systematically select every 10th passenger passing through American Airlines domestic security check-points and weigh their carry-on items. They select passengers until they have a total sample size of 1,000 people (50 from each airport). From this survey, the average weight of carry-on items is found to be 45 pounds per passenger, and the group concludes that there is a problem with excess weight on domestic American Airlines flights. Do you think this is a reasonable conclusion? Why or why not?

#### *Scenario 3 (Biased sampling; No problem with weight):*

They go to airports in New York, Chicago, and Los Angeles and select 500 passengers getting off of American Airlines flights arriving from Orlando and Miami, FL (two popular tourist spots) and weigh their carry-on items. From this survey, the average weight of carry-on items is found to be 35 pounds per passenger, and the group concludes that there is no problem with excess weight on domestic American Airlines flights. Do you think this is a reasonable conclusion? Why or why not?

#### *Scenario 4 (Biased sampling; Problem with weight):*

They go to airports in New York, Chicago, and Los Angeles and select 500 passengers getting off of American Airlines flights arriving from Orlando and Miami, FL (two popular tourist spots) and weigh their carry-on items. From this survey, the average weight of carry-on items is found to be 45 pounds per passenger, and the group concludes that there is a problem with excess weight on domestic American Airlines flights. Do you think this is a reasonable conclusion? Why or why not?

*Topic 2: Tuition Costs [“tuition costs in the US are too expensive at both public and private institutions.” Based on previous results, students tend to have strong opinions on this topic]*

*Scenario 1 (Biased sampling; Costs increased):*

Each year, the College Board ETS produces a “Trends in College Pricing” Report. It is reported that the average annual cost of tuition and fees at 4-year in-state public colleges in the United States during the 2009-2010 academic year was \$7020. In order to see if the average annual cost has increased for the 2010-2011 academic year, we collect the annual cost of in-state tuition and fees at all colleges in California. We find that the average cost was \$7395.20. When a one-sided hypothesis test was conducted, we found the  $p$ -value was statistically significant (0.0134) and were able to conclude that the average annual cost of tuition and fees at all 4-year-in-state public colleges in the United States has increased. Do you think that this is a valid conclusion? Why or why not?

*Scenario 2 (Biased sampling; Costs didn't change):*

Each year, the College Board ETS produces a “Trends in College Pricing” Report. It is reported that the average annual cost of tuition and fees at 4-year in-state public colleges in the United States during the 2009-2010 academic year was \$7020. In order to see if the average annual cost has increased for the 2010-2011 academic year, we collect the annual cost of in-state tuition and fees at all colleges in California. We find that the average cost was \$7100.18. When a one-sided hypothesis test was conducted, we found the  $p$ -value was not statistically significant (0.3183) and were unable to conclude that the average annual cost of all tuition and fees at 4-year-in-state public colleges in the United States has increased. Do you think that this is a valid conclusion? Why or why not?

*Scenario 3 (Random sampling; Costs didn't change):*

Each year, the College Board ETS produces a “Trends in College Pricing” Report. It is reported that the average annual cost of tuition and fees at 4-year in-state public colleges in the United States during the 2009-2010 academic year was \$7020. In order to see if the average annual cost has increased for the 2010-2011 academic year, we collect the annual cost of in-state tuition and fees at 50 colleges randomly selected from a list of all public 4-year colleges in the United States. We found that the average cost was \$7100.18. When a one-sided hypothesis test was conducted, we found the  $p$ -value was not statistically significant (0.3183) and were unable to conclude that the average annual cost of tuition and fees at all 4-year-in-state public colleges in the United States has increased. Do you think that this is a valid conclusion? Why or why not?

*Scenario 4 (Random sampling; Costs increased):*

Each year, the College Board ETS produces a “Trends in College Pricing” Report. It is reported that the average annual cost of tuition and fees at 4-year in-state public colleges in the United States during the 2009-2010 academic year was \$7020. In order to see if the average annual cost has increased for the 2010-2011 academic year, we collect the annual cost of in-state tuition and fees at 50 colleges randomly selected from a list of all public 4-year colleges in the United States. We found that the average cost was \$7395.20. When a one-sided hypothesis test was conducted, we found the  $p$ -value was statistically significant (0.0134) and were able to conclude that the average annual cost of tuition and fees at all 4-year-in-state public colleges in the United States has increased. Do you think that this is a valid conclusion? Why or why not?

*Topic 3: Mac vs. PC [“In the Mac vs. PC debate, I choose Mac.” Based on previous results, students tended to be split in their opinions on this topic—some felt strongly, others were more neutral]*

*Scenario 1 (Biased sampling; PC preferred):*

One of the more popular topics among computer users to debate is whether they prefer using a Mac or a PC. TechEBlog, a blog dedicated to the latest tech and gadget news, asked that exact question in an online poll in 2007. They asked online users to vote for which system they preferred: Mac, PC, or Other. The results of the online poll, which included over 15,000 respondents, showed



that 35.7% of users preferred a Mac, 58.5% preferred a PC and 5.8% preferred some other system. TechEBlog concluded that PC is the preferred system. Do you think this is a valid conclusion? Why or why not?

*Scenario 2 (Biased sampling; Mac preferred):*

One of the more popular topics among computer users to debate is whether they prefer using a Mac or a PC. TechEBlog, a blog dedicated to the latest tech and gadget news, asked that exact question in an online poll in 2007. They asked online users to vote for which system they preferred: Mac, PC, or Other. The results of the online poll, which included over 15,000 respondents, showed that 58.5% of users preferred a Mac, 35.7% preferred a PC and 5.8% preferred some other system. TechEBlog concluded that Mac is the preferred system. Do you think this is a valid conclusion? Why or why not?

*Scenario 3 (Random sampling; Mac preferred):*

One of the more popular topics among computer users to debate is whether they prefer using a Mac or a PC. A 2010 survey of 1200 randomly selected incoming freshman at five different universities showed that 70% of them had chosen to purchase a Mac as their personal laptop computer. Do you think it is a valid conclusion that Macs are more popular among college freshman than PC's? Why or why not?

*Scenario 4 (Random sampling; PC preferred):*

One of the more popular topics among computer users to debate is whether they prefer using a Mac or a PC. A 2010 survey of randomly selected incoming freshman at five different universities randomly selected to participate in a survey, showed that 53% of them had chosen to purchase a PC as their personal laptop computer. Do you think it is a valid conclusion that PCs are more popular among college freshman than Macs? Why or why not?

***Interview scenarios***

*Topic 4: Food Irradiation [“Foods that I eat should be irradiated for health reasons.” Based on previous results, students tend to have neutral opinions on this topic]*

*Scenario 1: Recommendation okay based on sample (Random sampling)*

In the 1980s, experimenters wanted to test if strawberries exposed to gamma rays (irradiated fruit) had a longer shelf life compared with non-irradiated strawberries. They randomly selected one thousand strawberries from 100 different farms across the US, exposed them to gamma rays, and measured how long it took for mold to grow. As a control group, they randomly selected one thousand other strawberries from the same 100 farms across the US, did not expose them to gamma rays, and measured how long it took for mold to grow. The results were statistically significant and the researchers concluded that, based on the data collected, irradiated strawberries had a delayed mold growth, and thus an increased shelf life, compared with the non-irradiated strawberries. Based on the results of this study, the American Council on Health and Science recommended that strawberries should be irradiated before being shipped to markets to ensure longer shelf life. Do you think this is a valid conclusion? Why or why not?

*Scenario 2: Recommendation not okay based on sample used (Biased sampling)*

In the 1980s, experimenters wanted to test if strawberries exposed to gamma rays (irradiated fruit) had a longer shelf life compared with non-irradiated strawberries. They randomly selected 10 strawberries that were grown by one farmer, exposed them to gamma rays, and measured how long it took for mold to grow. As a control group, they randomly selected 10 strawberries from a different farmer, did not expose them to gamma rays, and measured how long it took for mold to grow. The results were statistically significant and the researchers concluded that, based on the data collected, irradiated strawberries had a delayed mold growth, and thus an increased shelf life, compared with the non-irradiated strawberries. Based on the results of this study, the American Council on Health and

Science recommended that strawberries should be irradiated before being shipped to markets to ensure longer shelf life. Do you think this is a valid conclusion? Why or why not?

*Topic 5: Texting [“Texting while driving is dangerous.” Based on previous results, students tend to have strong opinions on this topic]*

*Scenario 1: Evidence is not convincing (Biased sampling)*

Recently, there has been concern about the dangers of texting while driving. To investigate this concern a study was conducted which involved placing video cameras in the cabs of more than 100 long-haul trucks for 18 months and observing the behavior of drivers. This study found that the risk of crashing when the drivers were texting was 23 times greater than when they were not texting. Suppose you are a state legislator and need to vote on a bill banning texting while driving. Does this study provide convincing evidence that texting while driving should be outlawed? Why or why not?

*Scenario 2: Evidence is convincing (Random sampling)*

Recently, there has been concern about the dangers of texting while driving. To investigate this concern a study was conducted in which 17 participants completed three drives in a driving simulator: one to familiarize themselves with the simulator, one while performing a number of text message functions (e.g. read, write) and one while not texting. To account for any learning effect caused by repeating the same route during the texting and non-texting drives, the order in which these drives were completed was alternated between participants. This study found that texting while driving resulted in a 35 percent increase in reaction time compared to not texting. Suppose you are a state legislator and need to vote on a bill banning texting while driving. Does this study provide convincing evidence that texting while driving should be outlawed? Why or why not?

*Topic 6: Medical Marijuana [Marijuana should be legalized for medicinal purposes in the United States.” Based on previous results, students tended to be split in their opinions on this topic—some felt strongly, others were more neutral]*

*Scenario 1 (Random sampling; Majority Support Legalization):*

The issues surrounding legalizing marijuana for medical purposes are being debated in the United States. Currently 14 states allow the use of marijuana solely for medical purposes. In April of 2010, AP and CNBC pollsters randomly selected 1001 adults nationally and found that 60% favored legalization of marijuana solely for medical purposes. Do you think it is valid to conclude that a majority (greater than half) of Americans think marijuana should be legalized for medical purposes? Why or why not?

*Scenario 2 (Random sampling; Minority Support Legalization):*

The issues surrounding legalizing marijuana for medical purposes are being debated in the United States. Currently 14 states allow the use of marijuana solely for medical purposes. In April of 2010, AP and CNBC pollsters randomly selected 1001 adults nationally and found that 45% favored legalization of marijuana solely for medical purposes. Do you think it is valid to conclude less than a majority (less than half) of Americans think marijuana should be legalized for medical purposes? Why or why not?

*Scenario 3 (Biased sampling; Majority Support Legalization):*

The issues surrounding legalizing marijuana for medical purposes are being debated in the United States. Currently 14 states allow the use of marijuana solely for medical purposes. In March of 2010, MSNBC posted an online poll on its website and found 60% favored legalization of marijuana solely for medical purposes. These results were based on over 15,000 responses. Do you think it is valid to conclude that a majority (greater than half) of Americans think marijuana should be legalized for medical purposes? Why or why not?

*Scenario 4 (Biased sampling; Minority Support Legalization):*

The issues surrounding legalizing marijuana for medical purposes are being debated in the United States. Currently 14 states allow the use of marijuana solely for medical purposes. In March of 2010, MSNBC posted an online poll on its website and found 45% favored legalization of marijuana solely for medical purposes. These results were based on over 15,000 responses. Do you think it is valid to conclude less than a majority (less than half) of Americans think marijuana should be legalized for medical purposes? Why or why not?