Lexical Ambiguity in Statistics: How students use and define the words: association, average, confidence, random and spread

Jennifer Kaplan
Michigan State University

Diane G. Fisher
The University of Louisiana at Lafayette

Neal T. Rogness
Grand Valley State University

Journal of Statistics Education Volume 18, Number 2 (2010),
www.amstat.org/publications/jse/v18n2/kaplan.pdf

Key Words: Statistics education, Lexical ambiguity, Language, Word usage.

Abstract

Language plays a crucial role in the classroom. The use of specialized language in a domain can cause a subject to seem more difficult to students than it actually is. When words that are part of everyday English are used differently in a domain, these words are said to have lexical ambiguity. Studies in other fields, such as mathematics and chemistry education, suggest that in order to help students learn vocabulary instructors should exploit the lexical ambiguity of the words. The study presented here is the second in a sequence of studies designed to understand the effects of and develop techniques for exploiting lexical ambiguities in statistics classrooms. In particular, this paper looks at five statistical terms and the meanings of these terms most commonly expressed by students at the end of an undergraduate statistics course.

1. Introduction

Language plays a crucial role in the classroom. It is a major means of communication of new ideas, the way in which students build understanding and process ideas and the method by which student learning is assessed (Thompson and Rubenstein, 2000). Language acquisition, the learning of language, is not a trivial process (Leung, 2005). Some words may have "core" meanings, where the word brings to mind a mental image, but even words that have core meanings, such as *cat*, may have associated characteristics that are not part of the core meaning. For example, *black cat* and *cattiness* have connotations that are not necessarily included in the core meaning of cat. When a word does not have a core meaning, it is even more difficult to learn to use it properly. Of the words that will be discussed in this paper, *average* is one that may not have a core meaning, since there is no mental image associated with the word *average* or perhaps multiple images depending on the context in which the word is used. For example, *average weight* and *average height* lead to distinct and different mental images.

In addition to the general issue of language acquisition, it is the case that as students begin to take specialized subjects in middle or high school and become exposed to each subject's specialized vocabulary they do not yet speak the language of the domain (Lemke, 1990). According to Lemke (1990), the use of specialized language that is unfamiliar to the student portrays a subject as more difficult than it is, a subject that can only be mastered by geniuses. Lemke (1990) further observed that people connect what they hear to what they have heard and experienced in the past. If a commonly used English word is co-opted by a technical domain, the first time students hear the word used in that domain they may incorporate the technical usage as a new facet of the features of the word they had learned previously. The use of domain-specific words that are similar to commonly used English words, therefore, may encourage students to make incorrect associations between words they know and words that sound similar but have specific meanings in statistics that are different from the common usage definitions. These words are said to have lexical ambiguity (Barwell, 2005).

Within the domain of statistics, Konold (1995) has found that students enter statistics classes with strongly-held, but incorrect, intuitions that are highly resistant to change. Coupled with the notion that students attach what they learn to previously held knowledge, this suggests a possible interference with statistics learning when statistics terms have lexical ambiguities in comparison to the words' everyday meanings. To date there has not been a methodical, large-scale study of language use in statistics classrooms, but statistics instructors have anecdotal evidence of students' misunderstandings and misinterpretation of words such as correlation, spread, and outlier, just to name a few.

Research done with elementary school children provides "evidence that awareness of linguistic ambiguity is a late developing capacity which progresses through the school years" (Durkin and Shire, 1991, pg. 48). Shultz and Pilon (1973) conducted a study on the development of the ability to detect linguistic ambiguity and found a steady, almost linear improvement across students in grade one, four, seven and ten. We can therefore conclude that college students, once made aware of the ambiguities, should be able to correctly process the statistics meaning of the ambiguous words. Helping students to become aware of and overcome the effects of the ambiguity is not hypothesized by the authors to be a trivial task. This paper describes an early stage of a research program that is designed to (1) highlight specific words and document obstacles to students' comprehension that are associated with misunderstandings of those words; (2) design and implement an intervention to investigate whether the explicit examination of the lexical ambiguity of certain words during instruction promotes deeper understanding of statistics; and (3) assess the success of the intervention on student learning outcomes using data. In particular, this paper will illustrate the ways in which students use the words association, average, confidence, random and spread when asked to write sentences and definitions for the statistical meanings of the words.

2. The study

2.1 Research Question

The paper describes the second stage of a pilot study of five words identified by the research team as possibly having lexical ambiguity: association, average, confidence, random and spread. Also included is the validation of the coding rubrics for each of the five words. For a detailed discussion of the choice of the five words as well as a complete literature review, see Kaplan, Fisher & Rogness (2009). In order to establish that these words have lexical ambiguities for students, we must first uncover what statistical meanings the students in an introductory statistics class have attached to the target words. The research question for the study presented here is: For the five target words, what are the statistical meanings most commonly developed and expressed by students at the end of an undergraduate statistics course?

2.2 Research Design

2.2.1 Pilot Study

The pilot study was conducted in the spring semester of 2008 at a university in the Southeastern United States. The university is classified as a research university with high research activity and has a total enrollment of approximately 16,000 students. The subjects were students in two sections of Elementary Statistics, a semester-long, three-hour course. There were approximately forty students enrolled in each section. This course is a service course for students in a variety of majors including nursing and the social sciences. The topics covered include descriptive statistics, confidence intervals, hypothesis testing, introduction to correlation and regression, and Chi Square Test of Independence.

Forty-nine students completed a questionnaire during the last week of the course, 31 women, 15 men and 3 students who did not provide information on gender. The questionnaire was administered during a class meeting so the students represent a convenience sample of those students who attended class on that day. Fourteen of the students (29%) were nursing majors; there were 21 other majors reported, such as psychology, advertising, and public relations, but no other major had more than 3 students. The distribution of the self reported GPAs of the students was unimodal with slight left skew, mean GPA of 2.98 and standard deviation of 0.52. The distribution of self reported ages of the subjects was unimodal with right skew; the median age was 19 years and the middle 50% of the ages between 19 and 21. No students under 18 years of age were surveyed.

The questionnaire asked:

a) Define or give a synonym for the word “association” as it is used in everyday English.
b) Define or give a synonym for the word “association” as it is used in statistics.

The same questions were repeated for each of the other four words. Explaining the study, obtaining consent and administering the instrument took approximately 15 minutes.

2.2.2 Validation Sample

A larger-scale study was conducted during the fall semester of 2008. In addition to the institution described above, two institutions in the Midwestern United States were included in the data collection. One is a large research university at which introduction to statistics courses are taught in lecture format. For three hours each week, the students meet in lecture halls with approximately 120 students per lecture. The students attend an additional hour of recitation with a graduate teaching assistant once per week in classes of 30 students. The other institution is a medium-sized comprehensive university which offers roughly 50 sections of a three-credit-hour introductory statistics class each semester. Enrollments across sections are approximately 30 students and all sections are taught by faculty members. In addition to meeting in a traditional classroom, each section also meets once per week in a computer lab. The topics covered in the classes at the two Midwestern institutions are comparable to those covered at the pilot study institution. The total number of subjects for the large-scale study was 777, with 14 different instructors across the three institutions.

Different from the pilot study, each subject in the large-scale study was asked to use each word in a sentence *and* give a definition or synonym for each word. This change from the pilot study was made because the researchers found that grouping or categorizing responses with definitions only was much more difficult than for responses that contained both definitions and sentences. Because asking students to give two sentences and two definitions for each word was more time consuming than the original version, some subjects were only asked to complete the task for three of the five target words. There were 35 versions of the questionnaire so the words could appear in different orders for the subjects. An example of one instrument is given in Appendix A. Most students completed the instrument within 10 minutes.

2.3 Analysis

The research team used the pilot study data to create coding categories for the students' definitions. Responses were grouped as being similar and then the groups created were described based on the similarities of the responses. Complete coding rubrics for each word are given in the next section. As an example, some of the coding categories for the word *average* are: mean, median, and representative number. One researcher read all the responses to one word and used the responses to create categories. Once the first researcher had finished creating coding categories for the responses and had coded all the responses, draft versions of coding categories and the instruments were then sent to another researcher. The second coder used the draft versions of coding categories from the first coder, but s/he did not have the results of the coding of the first researcher when making his/her own determinations. The initial agreement between the two coders appears in Table 1. After two researchers had coded the same instruments using the draft rubrics, modifications and edits to the coding categories were made. The three researchers discussed the responses on which the two independent coders disagreed and modified the coding rubric further as necessary. After this discussion there was 100% agreement between the three researchers as to the coding of each response.

Table 1: Initial agreement of 2 independent coders (total number of responses in parentheses)

Word	Association	Average	Confidence	Random	Spread
Pilot Study	96% (49)	96% (48)	89% (47)	67% (48)	98% (49)
Validation Sample	75% (63)	85% (70)	96% (65)	72% (66)	81% (74)

The research team then selected a random sample of 100 subjects from the large-scale study to validate the rubrics created with the pilot study data. These data will be referred to as the *validation sample.* Each definition and sentence pair was independently coded by two researchers. The initial agreement figures are generally lower for the validation sample, but that is largely due to the variability introduced by the diversity in the sample. The pilot study data were collected from students of the same instructor and the validation sample represents a random sample of students from a population at 3 institutions with 14 different instructors. All disagreements were discussed by the three researchers and the coding rubric was amended as necessary until there was 100% agreement for each response given by each student.

Some of the subjects provided definitions that the research team could not classify. This occurred when the researchers could not infer meaning from what the subject had written. Unlike grading a test, when an instructor attempts to find meaning in an incorrect response to award partial credit, the coding was done without inference into the subjects attempted meaning. Recall that this study is a preliminary stage of a research program. The results will first serve as a basis for uncovering the most common meanings that are expressed by students at the end of a course to assess whether the word exhibits lexical ambiguity. Those words for which the preliminary study finds evidence of lexical ambiguity will then be studied in more detail, using interviews to gain more insight into the meanings students hold for certain words in order to develop classroom activities and interventions. At this stage of the research, it is not necessary to have a detailed understanding of each misconception or misunderstanding held by individual students. More detailed results at the student level will be presented in future papers. Table 2 gives the number and percent of responses for each of the target words that could not be classified using the rubric. Examples of responses that could not be classified are given later for each of the target words.

Table 2: Number of responses unable to be coded (Percent in parentheses)

Word	Association	Average	Confidence	Random	Spread
Pilot Study	1 (2%)	4 (8%)	12 (26%)	4 (8%)	4 (8%)
Validation Sample	1 (2%)	6 (9%)	17 (26%)	6 (9%)	11 (15%)

2.4 Results

The results on inter-rater reliability and the percent of answers that can be coded, discussed in the previous section, provide evidence that meaningful results about student expressions of definitions of words can be obtained on a large scale using the coding rubrics and research methods of this study. In the remainder of this section, we provide the coding rubric and examples of student responses for each of the target words. The intent of this section is to identify and illustrate the common meanings for the target words that are expressed by students after taking introduction to statistics classes.

2.4.1 Association

Introductory statistics textbooks use association as a synonym for relationship, specifically, the relationship between two variables. The following are typical sentences using the word association found in Moore (2007):

> There is positive association – more boats goes with more manatees killed (pg. 101).
> A strong association between two variables is not enough to draw conclusions about cause and effect (pg. 144).

The coding rubric for *association* was designed with a hierarchy in mind so that the definitions at the top of Table 3 represent student responses that are closer to those considered statistically sophisticated. Student 414 provides an example of a response that the authors consider to be statistically strong, indicating that statistical association is a relationship between variables.

Example of association as relationship between variables – Student 414
Sentence: The birth rate had an association to mother's age.
Definition: A relationship or interaction between two variables

Table 3: Student statistical definitions of Association

Definition	Number of Subjects	
	Pilot Study	Validation Sample
Relationships between variables	9 (19%)	16 (25%)
Indeterminate relationships or linkages	15 (31%)	23 (37%)
Numerical comparisons	10 (21%)	10 (16%)
Having something in common	3 (6%)	10 (16%)
Incorrect statements: not about relationships or comparing	10 (21%)	3 (5%)
Not classified	1 (2%)	1 (2%)

Most of the students in the study, 50% and 62% in the pilot study and validation sample respectively, gave definitions that implied a relationship or linkage between two things. In both cases, only about 40% of that subgroup specified that the relationship is between variables in particular. The remaining responses that discussed relationships were vague about what things were related, as seen in the response from Student 579, which was coded as an indeterminate relationship.

Example of association as indeterminate relationship or linkages – Student 579
Sentence: There is an association between the to(sic) graphs
Definition: Association is to have some type of relation

A fair number of students (21% pilot; 16% validation) defined *association* using a numerical comparison, such as the correlation coefficient. One example is given by Student 86. Student responses that contained numbers in the sentences and those that referred to the correlation coefficient by name were categorized within this definition.

Example of association as numerical comparison – Student 86
Sentence: The association between the variables is -1.
Definition: strength of the relationship between variables

Finally, 16% of the validation sample defined *association* as a similarity between two objects, variables or groups with the example given by Student 57.

Example of association as similarity – Student 57
Sentence: There is an association between population 1 and population 2.
Definition: Association is similarities between populations.

2.4.2 Average

Introductory statistics textbooks tend to use the word *average* to describe the process of finding the mean of a data set (see for example, Moore, 2007). Triola (2006), however, specifically addresses the concern that many people use *average* interchangeably with the ideas of "median" or even "mode" stating "the term *average* is sometimes used for any measure of center and is sometimes used for the mean" (pg. 81). The coding rubric for *average* does not have a hierarchical structure. Instead, the definitions were grouped according to the statistical measures of center: mean, median and mode. For each of the named statistical measures of center there were three coding subcategories: the use of the word only; an incorrect, incomplete or colloquial definition; and the statistically correct definition. Definitions that did not relate to one of the three traditional measures of center were grouped in the "other definition" category. Responses in the "statistically correct definition" categories for each of the three measures of center (see Student 133) as well as responses in the "representative number" in the "other definition" category (see Student 303) are considered to be statistically more sophisticated than the other definitional categories. In the coding of the validation sample responses for *average*, we found 7 subjects (10%) who gave two distinct and separate meanings for *average*. Student 311 provides an example of this giving both the word mean and the colloquial meaning for mode in the provided definition. The percentages in Table 4, which also contains the descriptions of the coding categories, therefore, do not sum to 100%.

Example of student giving multiple definitions for average – Student 311
Sentence: I found the average of the data set using my calculator
Definition: the mean or most likely to occur

Table 4: Student statistical definitions of average

Definition			Number of Subjects: Pilot Study	Number of Subjects: Validation Sample
Statistical Measures of Center	Mean	statistical: complete and accurate	9 (19%)	17 (24%)
		statistical: incomplete or inaccurate	5 (10%)	11 (15%)
		word only	11(23%)	26(36%)
	Median	statistically correct		
		colloquial or incomplete: normal, standard, in the middle	3 (6%)	6 (8%)
		word only	1 (2%)	1(1%)
	Mode	statistically correct	1 (2%)	
		colloquial: majority, most common	5 (10%)	5 (7%)
		word only		
Other Definitions	Sum		4 (8%)	2 (3%)
	Representative Number		2 (4%)	1 (1%)
	Approximation		2 (4%)	
	Frequency		1 (2%)	
	Number used in inference		1 (2%)	
	Range of numbers			
Not Classified			4 (8%)	6 (8%)

Most of the subjects (72% pilot; 91% validation) gave a definition for *average* categorized as relating to a statistical measure of center. Students 133, 502, 77, 236, and 579 provide examples of responses from each of the categories relating to measures of center. Further, no students provided a statistically correct definition for *median.* The coding in this category was quite strict, in that responses were required to contain language about dividing the data set into two equal parts in order to be considered statistically complete. Responses, such as that given by Student 236, which referenced the middle without specifics, were coded as incomplete or colloquial.

Example of average as mean, statistically complete – Student 133
Sentence: What is the average of the two numbers?
Definition: The number you get from adding a group of numbers and then dividing by how many there were.

Example of average as mean, statistically incomplete – Student 502
Sentence: The average height of girls in our class is 5 ft. 6 in.
Definition: Average is a set of numbers added and divided.

Example of average as mean – word only – Student 77
Sentence: What is your test average?
Definition: mean

Example of average as median, colloquial use – Student 236
Sentence: The average GPA for 2 populations of students is one value between 3.4 and 3.75
Definition: Average is a number that lies in the middle of all the data numbers.

Example of average as mode, colloquial use – Student 579
Sentence: This is the average number from this data.
Definition: average is the mostly likely outcome

The most common definitions not relating to the standard measures of center were the average as sum (8%; 3%) or the average as a representative number (4%; 1%). An example of *average as sum* is given by Student 804 and an example of *average as a representative number* is given by Student 303.

Example of average as sum – Student 804
Sentence: The mean is the average of all the numbers involved.
Definition: average – sum of all the numbers involved in a data set

Example of average as representative number – Student 303
Sentence: What is the average value of the numbers?
Definition: The variable used to describe a larger pool of information.

Also provided below is the response by Student 354, which was not classified by the coding rubric because "compiling various numbers" was too vague for the researchers to have any insight into the student's understanding of the meaning of the word.

Example of unclassified response for average – Student 354
Sentence: I created an average of the data.
Definition: Compiling various numbers

2.4.3 Confidence

Introductory statistics textbooks tend to use confidence in the following types of sentences (DeVeaux, Velleman & Bock, 2009):

> Construct a 95% confidence interval (pg. 504).
> We are 95% confident that between 42.1% and 61.7% of Las Redes sea fans are infected (pg. 488).

The same textbooks define a C% confidence interval as an interval that will capture the true parameter value for C% of all possible random samples. We expected many sentences to be written about confidence intervals, but we had no expectations about possible definitions that students would write.

The largest class of student responses for *confidence* was composed of those that mentioned a level of certainty or surety: 39% in the pilot study and 50% in the validation sample. These responses were further subdivided based on the object of the certainty. Most responses were vague about the object of the certainty or left the object unspecified, as is the case with the example of student 917.

Example of confidence as level of certainty of something vague – Student 917
Sentence: The confidence interval can be calculated on the TI.
Definition: how sure you are

Some other examples of vague objects of surety were the percentage, the outcome of the study, or the claim. In the validation sample, 9% of students specified that the level of certainty is in reference to the location of a value (see Student 354). While not all of the responses were completely correct from a frequentist perspective, these responses were categorized as being very close to the technical meaning of *confidence.*

Example of confidence as level of certainty about the location of a value – Student 354
Sentence: I constructed a confidence interval for the data.
Definition: The level of certainty that a true value is captured in an interval

The last subclass of this category is composed of the responses that had the correctness of something as the object of the certainty. Student 865 provides one such example. Other responses in this category discussed the surety of the *answer* being correct.

Example of confidence as level of certainty about correctness – Student 865
Sentence: The data was found with 95% confidence.
Definition: The percentage that shows how often your data is correct.

Another large category of responses was composed of those in which the students defined *confidence* as a high level of certainty or to be very sure about something. Together with the responses specifying a level of surety, these responses account for about 60% of all the responses in both the pilot study and validation sample. Student 932 provides an example of a response classified having a high level of certainty.

Example of confidence as high level of certainty – Student 932
Sentence: Yesterday we learned how to do confidence intervals.
Definition: Being as sure as possible, that what you state is true.

In both the pilot study and validation sample, 26% of the responses were not classified. This is largely due to the fact that there appeared to be no patterns or commonalities amongst these responses. Had there been commonalities, categories would have been added to the rubric to accommodate the responses. Some examples of definitions that were not classified are that confidence is

- the significance of a statistic
- a possibility that is not quite approximate
- whether we agree with the data or outcome

Table 5 provides the distribution of the student definitions of *confidence* in the pilot study and validation sample.

Table 5: Student definitions of Confidence

Definition		Number of Subjects	
		Pilot Study	Validation Sample
A level of surety or certainty	about the location of a value	1 (2%)	6 (9%)
	of something vague or unspecified	12 (26%)	23 (35%)
	that something is correct	5 (11%)	4 (6%)
Have a high level of certainty, be very sure		9 (19%)	7 (11%)
Accuracy or precision		1 (2%)	5 (8%)
An interval		6 (13%)	1 (2%)
Ability to provide evidence		1 (2%)	1 (2%)
Not classified		12 (26%)	17 (26%)

The responses of Students 312, 799 and 414 provide examples of the three categories not discussed previously, *confidence as accuracy or precision, confidence as an interval* and *confidence as the ability to provide evidence.*

Example of confidence as accuracy or precision – Student 312
Sentence: Create a 95% confidence interval.
Definition: confidence is the amount/% accuracy you can account for in your results.

Example of confidence as an interval – Student 799
Sentence: I found a 95% confidence interval
Definition: a particular type of interval

Example of confidence as the ability to provide evidence – Student 414
Sentence: We have 95% confidence in the data.
Definition: having data that supports the claim

2.4.4 Random

Typical introductory statistics textbooks tend use the word *random* as an adjective that modifies words such as *phenomenon, event, sample,* and *digits* (see for example, DeVeaux et al., 2009, Moore, 2007 and Triola, 2006). Moore (2007) defines *random* writing "We call a phenomenon **random** if individual outcomes are uncertain but there is nonetheless a regular distribution of outcomes in a large number of repetitions" (pg. 248). By the end of a statistics course, the main use of the word *random* is likely to have been in sentences such as "The data were collected using random sampling" where the meaning of *random* is that every person (or same-sized group of people) had an equal chance of being chosen.

The student responses for the adjective form of *random* exhibited variety similar to that found in textbooks. The coding rubric for the word *random*, given in Tables 6 and 7 is a two level rubric.

Responses were first coded as to the usage of the word *random*: a general ***adjective***, the ***outcome*** of random assignment or random selection, or the ***process*** of random assignment or random selection. Few students used *random* as a general adjective not associated with a process or outcome (21% pilot; 13% validation) with an example given by Student 17. Twenty-eight percent of students in the pilot sample and 45% in the validation sample used *random* to describe the outcome of a random process (for example, Students 307, 492, and 275) and 40% of students in the pilot sample and 38% in the validation sample used *random* to describe the process of a random process (for example, Students 938 and 837).

Table 6: Student statistical definitions of random: Pilot Study (43 out of 48 coded)

	Other	By chance (vague)	Without order or reason	Unexpected, not predictable, unplanned	Without bias, representative, fair	Statistical: every element is equally likely
Random as adjective	0	0	5	2	0	2
Outcome of a random action	1	0	5	1	5	0
Process of a random action	0	2	5	1	6	3
Incorrect	5	0	0	0	0	0
Total (% of 43)	6 (13%)	2 (5%)	15 (35%)	4 (9%)	11 (25%)	5 (11%)

Table 7: Student statistical definitions of random: Validation Sample (60 out of 66 coded)

	Other	By chance (vague)	Without order or reason	Unexpected, not predictable, unplanned	Without bias, representative, fair	Statistical: every element is equally likely
Random as adjective	2	0	3	2	1	0
Outcome of a random action	2	1	9	4	9	2
Process of a random action	1	1	13	4	3	1
Incorrect	2	0	0	0	0	0
Total (% of 60)	7 (11.7%)	2 (3%)	25 (41.7%)	10 (16.7%)	13 (21.7%)	3 (5%)

Within each of the usages, there were five levels of definition: by chance, without order or reason, unexpected, without bias and equally likely. As with the definitions for association, these categories are considered to be ordered, with those on the right of Tables 6 and 7 to be closer to a statistically sound understanding of the word *random.* When it was not clear to the researchers from the response what meaning a subject had attached to a word, the response was classified as "unable to be coded." Subjects with understandable answers that did not fit in any other category were coded as "other" such as that given by Student 17.

*Example of a definition for random classified as other (**adjective**) – Student 17*
Sentence: We used a random variable today.
Definition: random: unknown

Other definitions in this category are an outlier, a type of data, and a method of looking at data from different perspectives.

In the first category, the subjects used the phrase "by chance" in a way that was vague and did not specify the method of random choice, as in the example of Subject 938.

*Example of vague definition of random (**process**) – Student 938*
Sentence: For the survey, a random sample was picked.
Definition: by chance that something occurred.

In the second category, subjects discussed random selection or assignment as being without order, reason or pattern, sometimes indicating that there was no known process. While this is not incorrect, per se, it is still not particularly specific in terms of how the random process is enacted, as evidenced by the response of Subject 307.

*Example of random as without order or reason (**outcome**) – Student 307*
Sentence: It was a random sample, which provides independence.
Definition: Random: persons were chosen not based on any reason.

The third category is a more robust description of *random* in that it includes the notion of unpredictability. Responses in this category tended to remain vague and/or include an element of the unexpected or unplanned. Subject 492 provides an example of a response in this category that is vague about the outcome of a random process, but includes the idea of outcomes being unknown prior to the process.

*Example of random unexpected or not predictable (**outcome**) – Student 492*
Sentence: I was picked for a random sample.
Definition: Not pre-determined.

In the fourth category, some subjects referenced the word *unbiased*, while others specifically mentioned that the outcome of a random process would be representative or fair. Since we use random processes to reduce bias, these responses are not technically incorrect. They do not, however, approach the specificity of random as each element or sample having an equal chance of occurring or being chosen. The example provided by Student 275 is one in which representative is specifically mentioned.

*Example of random as without bias or representative (**outcome**) – Student 275*
Sentence: The sample population is a random sample.
Definition: Sample is equally representative of all groups of the population.

The response would have been categorized the same way if the definition had been "the sample is unbiased." Finally, Student 837 provides an example of a response that was coded as a statistically correct definition about the process of random sampling. In order to be considered in this category, a response needed to be specific about the equal chance of selection.

*Example of random as every element being equally likely (**process**) – Student 837*
Sentence: We took a random sample of the students.
Definition: everyone was equally likely to be chosen for the sample.

Notice in Table 6 that at the end of the semester only 11% of subjects gave a correct statistical definition for the word "random." Table 7 shows that the percent of students in the validation sample who gave a correct statistical definition was only 5%.

2.4.5 Spread

The word *spread* in introductory statistics textbooks is used as a synonym for the words *variability* or *dispersion* or the more technical phrase, *scale parameter*. It is quite common to find the instructions, "Be sure to discuss the distribution's shape, center, spread, and unusual features" (DeVeaux et al., 2009, pg. 68). The student responses for *spread* fell into three major categories: definitions relating to dispersion, definitions relating to distribution or layout, and definitions relating to spreadsheets or data.

The results for *spread* are interesting because the results from the two samples, the pilot study and the validation sample, were very different. In the pilot study, 60% of the subjects referred to a spreadsheet in their definition, either discussing the data contained in the spreadsheet (see Student 173 as an example) or the spreadsheet that contains data (see Student 752). In contrast, only 14% of the responses in the validation sample mentioned a spreadsheet. This is especially surprising since none of the classes in which the students were surveyed use a spreadsheet as an analysis tool.

Example of spread as data in a spread sheet – Student 173
Sentence: data
Definition: information in columns

Example of spread as spread sheet – Student 752
Sentence: We used a spread sheet in class.
Definition: table of numbers

In the validation sample, 42% of the responses related to dispersion, two-thirds of which were responses about the range (Student 492) the other one-third giving the synonym of *variability* or mentioning the variance or standard deviation (Student 799).

Example of spread as range – Student 492
Sentence: There was a large spread in the data.
Definition: Range

Example of spread as variability – Student 799 **Sentence:** The spread of the numbers had a very large range. **Definition:** Spread: amount of variability

In the pilot study, only 15% of responses were categorized as being about dispersion. While two-thirds of those were about the range, the remainder contained a reference to a non-specific calculation. Notice that Student 799 mentions both range and variability in his response. This was coded as a response about variability because the definition specified that spread is a measure of variability and the word *range* was considered an adjective modifying *spread* rather than a synonym for *spread.*

Finally, only 14% of the pilot study responses were categorized as about the layout or distribution of the data compared to 35% of the validation sample responses. Examples of each of the subcategories for this major category are given by Students 688, 198 and 770. Note that the percentages in the table may not sum to 100% because some responses were given two codes. The two codes were always from separate major categories.

Example of spread as visual layout – Student 688 **Sentence:** The data was in a uniform spread. **Definition:** Spread – how data is laid out on a graph.

Example of spread as layout or distribution – Student 198 **Sentence:** The group of data were spread out evenly **Definition:** spread is the distribution of the data.

Example of spread as far apart – Student 770 **Sentence:** The points are spread over a huge area in the scatter plot. **Definition:** wide area, not close together.

Table 8 provides the complete distribution of student definitions for *spread* in the pilot study and validation sample.

Table 8: Student definitions of Spread

Definition		Number of Subjects (%)	
		Pilot Study	Validation Sample
Definitions relating to Dispersion	Variability, variance, standard deviation	0	11 (15%)
	Range	5 (11%)	20 (27%)
	A non-specific measure or calculation	2 (4%)	0
Definitions relating to Distribution or Layout of the Data	How the data look when represented visually on a graph	2 (4%)	10 (14%)
	How the data are distributed (no mention of visual representations, includes uniformly distributed)	3 (6%)	10 (14%)
	Data are far apart or scattered (may include reference to visual representations or not)	2 (4%)	5 (7%)
Definitions relating to Spreadsheets	A spread sheet (with data)	14 (30%)	8 (11%)
	Data (in a spread sheet or list)	14 (30%)	2 (3%)
Not classified		4 (8%)	11 (15%)

3. Discussion

3.1 Summary of Findings

Recall the research question of this study: For the five target words, what are the statistical meanings most commonly developed and expressed by students at the end of an undergraduate statistics course? The preliminary findings previously discussed show each of these words to be problematic. With regard to the word *association* it is discouraging that many students believe an association to be a commonality or similarity between two things (6% pilot; 16% validation) rather than a more general relationship. Furthermore, many of the students who have developed a relational understanding of *association* may not have progressed further than describing a numerical relationship (21% pilot; 16% validation). Finally, only 19% of the pilot study students and 25% of the validation sample were able to express the definition of *association* explicitly as a relationship between two variables.

The results for the word *average*, in which students tended to discuss one of the three most common measures of center (72% pilot; 91% validation), may mirror the way in which the instructors use the word *average.* An informal poll of 65 statisticians, statistics educators and instructors at a session at the U.S. Conference on Teaching Statistics (USCOTS) 2009 indicated that about half of the respondents use *average* as a synonym for the mean and the other half use it as a general term implying any measure of center. Of the pilot study students who mentioned a measure of center in their definition of *average*, 72% mentioned the mean specifically. The corresponding figure for the validation sample is 82%.

The results about the word *confidence* are encouraging because of the numbers of students who recognize that in statistics *confidence* is a level of surety (39% pilot; 50% validation) rather than being very sure of something (19% pilot; 11% validation). On the other hand, the number of students who were not able to write coherently about the word *confidence* (26% in both samples) does suggest that there is potential to improve understanding of the word in general.

It is particularly discouraging that, in the validation sample, only 5% of the subjects were able to correctly define the word *random* as used in a statistical sense. While the reader might be encouraged by students' connection between *random* and *unbiased* or *representative*, other data, which will be discussed in future publications, indicates that students believe that random sampling means that the researchers started by stratifying the population in order to force a representative sample rather than understanding that the probability structure underlying random sampling provides the theoretical lack of bias in a random sample.

Finally, the results about the word *spread* seem to indicate that a similar number of students use the term as a synonym for *shape*, how the data look on a graph or where the data are (14% pilot; 35% validation), as those who use it as a synonym for *variability* (15% pilot; 42% validation).

3.2 Implications for Teaching

The findings discussed here provide initial ideas for teaching that might help students develop better understandings of the vocabulary associated with statistics. While these suggestions are preliminary and rather vague, the research team intends to use the preliminary findings to design a more comprehensive intervention for some of the words that have been investigated. For two of the words studied, *average* and *spread*, the initial findings indicate that exploiting the lexical ambiguities may not be an appropriate or fruitful avenue to pursue. Instead, the statistics teaching community might consider removing these words from general technical use. Previously published results (Kaplan et al., 2009) showed the large number of existing colloquial definitions for the word *spread*, such as a vast area, a coverlet, and a buffet of different food items. Given the number of existing definitions for *spread*, the choice of this word as a colloquialism for the statistical term of *variability* or *scale parameter* is perhaps a poor one. We suggest that instructors use the technical word, *variability*, in classes and dispense with the word *spread*. Similarly, *average* does not need to be used in statistics classrooms as a technical word. Instructors could refer to *measures of center* or use the name of the particular measure of center, the mean, median, or mode, depending on the context. This attention to detail would not take time away from instruction of other material and might dispel some of the confusion that students have about the descriptions of distributions of quantitative variables.

The other three target words in this research are part of the statistical lexicon in an essential way and cannot be dealt with through replacement. They provide a greater challenge for teaching. Clearly, more research is needed about these three words, which will be discussed in the following section. The research team does have some initial thoughts about addressing student difficulties with *association, confidence* and *random.* With regard to *association*, we suggest that instructors stress, every time an association is discussed, that *association* is a synonym for *relationship*. Moreover, instructors should be clear that the relationship is between two variables. The thread of this discussion should include both categorical and quantitative variables. When discussing *association* between categorical variables, instructors should emphasize that *association* does not mean similarity. For *association* between quantitative variables, instructors should point out that correlation is a measure of the linear *association* but not a synonym for *association.*

Previous results about the word *random* suggest that instructors contrast the statistical and colloquial meanings of the word every time it is used in class (Kaplan et al., 2009). A similar suggestion is made by Lavy and Mashiach-Eizenberg (2009).The data presented here suggest further that instructors need to be very specific that taking a random sample or enacting random assignment means that there is an equal probability associated with the process. Furthermore, instructors are encouraged to contrast simple random sampling/assignment with stratified random sampling/assignment to help students understand that stratification is not an element of *random*.

Confidence, like *random*, is generally used in introduction to statistics courses, and by the students, as an adjective. With *random* the students discussed either random samples or random assignment and these can be contrasted with stratified sampling or voluntary assignment. *Confidence* tends to be used as a modifier for either *level* or *interval*; but it is unclear, at this time, with what *confidence interval* can be contrasted. More research on the word *confidence* is necessary before the research team can make suggestions about its use in class.

3.3 Future Directions

This study reports the findings from the second stage of a multiple-stage research program. In particular, the goals of this program are to use data to highlight specific words and document barriers to students' comprehension that are associated with misunderstandings of those words, design and implement an intervention to investigate whether the explicit examination of the lexical ambiguity of certain words during instruction promotes deeper understanding of statistics and assess the success of the intervention on student learning outcomes. To this end, the authors have already chosen five additional words, *bias, error, independent, normal,* and *significant* and have collected beginning of course and end of course data from students. Coding rubrics will be created and validated for the second set of words in a similar fashion to those described here.

In the future, the research team will be working with linguistic software to aid in the coding of the data. A review of commonly-used introductory statistics textbooks will be done to determine if and how these words are used by the authors. Furthermore, the research team intends to videotape statistics classes to find out how instructors introduce and use the target words in classrooms and interview students to resolve issues that arise within the large scale data sets. The findings from the text and video analyses, together with the large-scale studies of lexical ambiguity in statistics, will be used as the basis for the creation of instructor resources that will provide suggestions for instruction in which the statistical and everyday meanings of words can be explicitly linked for students so that they will develop strong statistical meanings of technical vocabulary words that are similar to common English words. At this stage of the research, we also plan to measure the impact the ambiguity has on student performance in statistics. We are confident that addressing lexical ambiguity of statistics terms is one path to helping students develop better understanding of statistics without adding topics to an already over burdened curriculum.

APPENDIX A – Sample Post-test Instrument

Name (please print) ____

Data Collection Form B-1

1. **Average**
   a. Write a sentence with the word "average" using its primary meaning to you in everyday language.

   b. Provide a definition for the word "average" that maintains the same meaning as you used in the prior sentence.

   c. Write a sentence with the word "average" using its primary meaning to you in statistics.

   d. Provide a definition for the word "average" using its primary meaning to you in statistics (i.e. maintaining the same meaning as you used in the prior sentence).

2. **Confidence**
   a. Write a sentence with the word "confidence" using its primary meaning to you in everyday language.

   b. Provide a definition for the word "confidence" that maintains the same meaning as you used in the prior sentence.

   c. Write a sentence with the word "confidence" using its primary meaning to you in statistics.

   d. Provide a definition for the word "confidence" using its primary meaning to you in statistics (i.e. maintaining the same meaning as you used in the prior sentence).

3. **Random**
   a. Write a sentence with the word "random" using its primary meaning to you in everyday language.

   b. Provide a definition for the word "random" that maintains the same meaning as you used in the prior sentence.

   c. Write a sentence with the word "random" using its primary meaning to you in statistics.

   d. Provide a definition for the word "random" using its primary meaning to you in statistics (i.e. maintaining the same meaning as you used in the prior sentence).

Acknowledgements

This work was completed under the CAUSEmos research cluster grant funded by NSF DUE-0618790. The authors would like to thank their research mentors, Sterling Hilton, John Holcomb, and Marsha Lovett for their invaluable help with designing the research. Furthermore, the advice given by the two referees and associate editor were invaluable toward the presentation of the topic and results discussed in the paper.

References

Barwell, R. (2005). Ambiguity in the Mathematics Classroom. *Language and Education 19(2)*, pp. 118 – 126.

DeVeaux, R.D., Velleman, P.F. and Bock, D.E. (2009) *Intro Stats, 3rd Edition*, Boston, MA: Pearson Education, Inc.

Durkin, K. and Shire, B. (1991). Primary school children's interpretations of lexical ambiguity in mathematical descriptions. *Journal in Research in Reading, 14(1)*, pp. 46 – 55.

Kaplan, J.J., Fisher, D., and Rogness, N. (2009). Lexical Ambiguity in Statistics: What Do students know about the words: association, average, confidence, random and spread? *Journal of Statistics Education.* http://www.amstat.org/publications/jse/v17n3/kaplan.html

Konold, C. (1995). Issues in Assessing Conceptual Understanding in Probability and Statistics. *Journal of Statistics Education, 3(1)*, pp. 1 – 11.

Lavy, I. and Mashiach-Eizenberg, M. (2009). The interplay between spoken language and informal definitions of statistical concepts. *Journal of Statistics Education, 17(1)*, Available online: http://www.amstat.org/publications/jse/v17n1/lavy.html

Lemke, J. (1990). *Talking Science*: *Language, Learning and Values*. Norwood, NJ: Ablex Publishing Corporation.

Leung, C. (2005). Mathematical vocabulary: Fixers of knowledge or points of exploration. *Language and Education, 19(2)*, pp. 127 – 135.

Moore, D. (2007). *The Basic Practice of Statistics, 4th Edition*. New York, NY: W. H. Freeman and Company.

Shultz, T. and Pilon, R. (1973). Development of the Ability to Detect Linguistic Ambiguity. *Child Development, 44*, pp. 728 – 733.

Thompson, D. and Rubenstein, R. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *Mathematics Teacher*, *93(7)*. pp. 568 – 574.

Triola, M. F. (2006). *Elementary Statistics, 10th Edition*. Boston, MA: Pearson Education, Inc.

Jennifer J. Kaplan
Michigan State University
Department of Statistics and Probability Michigan State University
A413 Wells Hall
East Lansing, MI 48824
Email: kaplan@stt.msu.edu

Diane G. Fisher
Mathematics Department
University of Louisiana at Lafayette
403 Maxim Doucet Hall
P.O. Box 41010
Lafayette, LA 70504-1010
Email: dgf9042@louisiana.edu

Neal T. Rogness
Department of Statistics
Grand Valley State University
1 Campus Drive
Allendale, MI 49401
Email: rognessn@gvsu.edu
