



Lexical Ambiguity in Statistics: What do students know about the words association, average, confidence, random and spread?

Jennifer J. Kaplan
Michigan State University

Diane G. Fisher
University of Louisiana - Lafayette

Neal T. Rogness
Grand Valley State University

Journal of Statistics Education Volume 17, Number 3 (2009),
www.amstat.org/publications/jse/v17n3/kaplan.html

Copyright © 2009 by Jennifer J. Kaplan, Diane G. Fisher, and Neal T. Rogness all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

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 a pilot study that is the first in a sequence of studies designed to understand the effects of and develop techniques for exploiting lexical ambiguities in the statistic classroom. In particular, this paper describes the meanings most commonly used by students entering an undergraduate statistics course of five statistical terms.

1. Introduction

Written and spoken language play a crucial role in the classroom. They are the 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 & Rubenstein, 2000](#)). As students begin to take specialized subjects in middle or high school, they become exposed to each subject's specialized

vocabulary ([Lemke, 1990](#)). Upon their entry into a new subject, students do not yet speak the language of the domain. According to [Lemke \(1990\)](#), the use of a specialized vocabulary with a novice in a domain creates a "mystique" about the subject. The subject may seem to the student "dogmatic, authoritarian, impersonal and even inhuman" (pg. xi). Furthermore, the use of specialized language that is unfamiliar to the student may portray the subject as more difficult than it is, a subject that can only be mastered by geniuses. While Lemke's work has focused mainly on the language used in science, his claims are equally relevant to statistics. In fact, [Makar and Confrey \(2005\)](#), in their study of pre-service teachers use of non-standard language to discuss variation, also found that neglecting students' use of non-standard language makes the subject seem unreachable and more complex or difficult than other subjects.

[Lemke \(1990\)](#) further observed that people connect what they hear to what they have heard and experienced in the past. [Konold \(1995\)](#) has done extensive research on student understanding of various probabilistic and statistical concepts. Three major findings he has noted are that "(1) students come into our courses with some strongly-held yet basically incorrect intuitions, (2) these intuitions prove extremely difficult to alter, and (3) altering them is complicated by the fact that a student can hold multiple and often contradictory beliefs about a situation" (Konold, pg. 2). Furthermore, he has concluded "we have a variety of data suggesting that these intuitions are persistent and, to this point, survive our best teaching efforts" (ibid, pg. 6). Although Konold's work does not deal exclusively with the meanings of specific words, it seems reasonable that they can be applied to the learning of language.

The authors hypothesize, as an extension of the findings of Konold to the learning of statistical vocabulary, that 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. The words or phrases that are the same or similar but can be used to express two or more different meanings are said to have lexical ambiguity ([Barwell, 2005](#)). To date there has not been a large scale formal study of language use in statistics classrooms, but statistics instructors have anecdotal evidence of students' misunderstandings and interpretation of words such as correlation, spread, and outlier, just to name a few. As such, this research seeks to investigate and understand lexical ambiguity as it relates to statistical terms and the learning of statistics by undergraduate students. Specifically, the stage of research being discussed in this article is an assessment of prior knowledge and meanings attached to select words used in introductory statistics that also have meanings in everyday usage of language.

Research done with elementary school children as subjects provides "evidence that awareness of linguistic ambiguity is a late developing capacity which progresses through the school years" ([Durkin & Shire, 1991b](#), pg. 48). [Schultz and Pilon \(1973\)](#) conducted a study on the development of the ability to detect linguistic ambiguity. They looked at four different grade levels – first, fourth, seventh, and tenth – and four different types of ambiguity. Detection of lexical ambiguities exhibited a steady, almost linear improvement across grades ([Schultz & Pilon, 1973](#)). The authors, therefore, conclude that college students, once made aware of the ambiguities, should be able to correctly process the statistics meaning of the ambiguous words. This suggests that there is need for study of the language used in statistics, and in particular, those words that have lexical ambiguity, and the effects on student learning. Once lexical ambiguities in statistics have been identified and their effects studied, appropriate measures can be taken by instructors to address the lexical ambiguities in classrooms. The primary research question of this paper is: What are the meanings of select target words commonly used by students entering an undergraduate statistics course?

2. Literature Review

The intent of this section of the paper is to inform the reader about the construct of lexical ambiguity in general. Under the assumption that the reader is unfamiliar with lexical ambiguity, the review of the lexical ambiguity literature is comprehensive. The literature cited below provided the authors with a blueprint for a long-term research program to study lexical ambiguity in statistics. The ultimate goal of the research program is to provide statistics instructors with ways to address ambiguous words so that students will develop better understandings of the words and related statistical concepts. This paper, the first of several planned to explore how lexical ambiguity connects to introductory statistics, focuses on the meanings for five target words that students bring into such a course. Subsequent papers will focus on the meanings students attach to these words at the end of the course as well as how the findings connect to the existing lexical ambiguity literature as well as how the construct relates to content found in introductory statistics. While the findings presented in this article are from undergraduate, introductory statistics students, the review of the literature primarily draws upon lexical ambiguity research done with K – 12 students as well as undergraduate students from disciplines such as mathematics education and chemistry education.

2.1 Language and Language Acquisition

Language acquisition is not a trivial process; Leung states "that learning a word is not a simple and straightforward matter of getting and learning its definitive meaning/s" ([2005](#), pg. 131). 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" has connotations that are not necessarily included in the core meaning of cat. In the case when a word does have a clear core meaning, vocabulary studies next consider a set of necessary and sufficient conditions that must be met by an object or concept in order to be represented by a word ([Leung, 2005](#)).

While the idea that a definition of a word is a set of necessary and sufficient conditions might appeal to mathematicians, [Leung \(2005\)](#) illustrates the impracticality of this idea in practice by providing four different published definitions of the word "square." Each of the definitions lists a different set of necessary and sufficient conditions for an object to be classified as a square. Therefore, even after establishing that knowing the meaning of a word means knowing its core meaning (when such a thing exists) as well as its non-core meanings and the relationships between the various possible representations of meaning, there is still the issue of how such meanings are acquired. Leung's final argument for her claim that learning a word is not straightforward is to cite Schmitt's observation that learning the meaning of a word is an incremental process. The first time a person hears a word, he will remember only one particular meaning sense of the word, the one in which it was used on that occasion. It is only after repeated exposures to the word that "basic formal and semantic features....are built up and consolidated" ([Leung, 2005](#), pg. 131) so learning the nuances of word meanings is a long process that happens via repeated exposure.

2.2 Lexical Ambiguities in the Classroom

Durkin and Shire (1991a) identified four types of lexical ambiguities in mathematics education:

- Homonymy: words that share the same form but have different meanings
 - Leaves: part of a tree/ seven minus four leaves three
- Polysemy: words that have two or more different but related meanings
 - Product: something that has been made/ the results of a multiplication

- Homophony: words with different spelling (and meaning) but same pronunciation
 - Sum/ some; Pi/ pie; Two/too/to
- Shifts of application: words that can mean different things when considered from different perspectives
 - Number: nominal, ordinal, cardinal or visual

Within the mathematics and science education literature, several authors suggest practical strategies for helping students to deal with lexical ambiguities in mathematics classrooms. These suggestions are based on both research results ([Durkin & Shire, 1991a](#)) and classroom experience ([Adams, et al., 2005](#)) and classroom observations ([Lemke, 1990](#)). Two of the major suggestions made by the authors are to acknowledge and exploit the lexical ambiguities and to help students to "build their voices." [Adams, et al. \(2005\)](#) suggest that students list the ambiguous word pairs and write sentences for each meaning. In addition, they propose that teachers ask students what they think words mean before giving a technical definition so that the new knowledge can be attached to prior knowledge. [Lemke \(1990\)](#) suggests that students be asked to translate between technical and colloquial statements of questions. Finally, [Durkin and Shire \(1991a\)](#) propose that teachers use words in contexts where colloquial meanings coincide with technical meanings to build a solid foundation for students.

The major suggestion to help students develop their voices in a domain is to give students the opportunity to speak and write in the domain. [Thompson & Rubenstein \(2000\)](#) suggest silent teacher, where the teacher does not speak and instead lets the students be the sole oral communicators, or choral response, in which the class gives singsong oral responses in unison as activities to help student to develop oral communication skills. They also suggest that the teacher should listen carefully to student talk and make corrections or help with rephrasing when necessary and having students discuss writing samples of varying quality and use journaling to develop written skills. [Lemke \(1990\)](#) suggests that students be explicitly taught to combine technical terms in complex sentences as well as being exposed to examples of major and minor genres of science writing.

In a case study of a fifth grade teacher, [Khisty and Chval \(2002\)](#) suggest that Ms. Martinez helped her students learn mathematical communication by creating an environment filled with "rich words that students appropriate as their own" (pg. 154). The authors acknowledge that the process of learning vocabulary is not a matter of "giving" words to students. They write,

The words represent meanings that are waiting to be developed and eventually internalized. Therefore, which words are presented to the students and how they are developed are vitally important. Just as important is that students have opportunities to use these words in their talk and as they work (pg. 155).

Ms. Martinez began populating her classroom with sophisticated vocabulary, such as the word "inverse" during the first week of school. She modeled "appropriate academic and mathematical discourse" (pg. 160) for her students, drew specific attention to new words and connected the new words to words or ideas that students might have known previously. As the year went on, Ms. Martinez moved from rephrasing and repeating student responses to a more silent teacher giving her students the space to practice their developing mathematical voices.

In a study of the use of language by pre-service teachers, [Makar and Confrey \(2005\)](#) found that the use of non-standard or informal language, such as clumps and chunks, allowed the pre-service teachers to demonstrate and articulate their understanding of variation in a way that the technical language, such as interquartile range or standard deviation, did not. Furthermore, they claim, when using non-standard language the prospective teachers were able to integrate ideas of distribution and variation. The statements made using technical language tended to be less relational and separated, rather than

integrated, the ideas of distribution and variation. While Makar and Confrey were not addressing lexical ambiguity directly, their work does provide insight into helping students "build their voice" and develop understanding using non-standard terminology. They suggest that students benefit from the use of non-standard terminology because they are using words that hold meaning for them and convey their own conceptions, every day meanings are more accessible to a large variety of students and allows for multiple entry points into the material and the students are better oriented to develop correct conceptions of the concepts they are learning.

Applying this research to undergraduate students, as words having lexical ambiguity are encountered, instructors might want to bring specific attention to these words and discuss how the usage in statistics compares and/or differs from their everyday usage. Further, instructors could employ journaling and develop journal entries designed to help students build their statistics voice and confront lexical ambiguity in statistics. An example of such an entry is: Discuss the meaning of *range* in statistics, in studying functions, and in everyday English. Are there meanings that are shared by all or some of the uses? What is special about each of the mathematical meanings?

2.3 Previous Studies of Language and Lexical Ambiguities

In order to create instructional materials that aid teachers in confronting lexical ambiguities in the statistics classroom, more must be known about the nature of lexical ambiguities in statistics and their effects on student understanding. There have been no formal studies of lexical ambiguities in statistics education, so the literature from mathematics, chemistry, and language education serves as a basis for the current research. [Durkin and Shire \(1991a\)](#) used a multiple-choice task to study the effects of lexical ambiguities in mathematics with 10-year-old students. After collecting a list of mathematics words with lexical ambiguities, they created two sentences for each word, one using the mathematical meaning and the second using the every day meaning. They also created meaning choices for each word: 1. a synonym for the mathematical meaning, 2. a synonym for the correct every day meaning, 3. a word with thematic relation to the mathematical meaning and 4. a word with thematic relation to the every day meaning. Students were given booklets containing one of the two sentences for each word and were asked to choose the best meaning of the target word as it was used in the sentence. [Durkin and Shire \(1991a\)](#) found that when children misidentified the meaning of an ambiguous word in a mathematical sentence, the sense they chose was often the every day sense. This happened significantly more often than a child interpreting an everyday use of an ambiguous word as though it were the mathematical usage.

[Tomlinson, Dyson and Garratt \(2001\)](#) studied undergraduate chemistry students' understanding of the vocabulary of error. The researchers claim "even the word 'error' is a source of confusion to many students since students commonly regard 'errors' as personal mistakes rather than recognizing that "*every physical measurement is subject to a host of uncertainties that lead to a scatter of result*" (pg. 1). The researchers, together with a number of colleagues, developed a list of terms about error with which first year students should be familiar. They then designed open-ended tasks to test students' procedural knowledge, defined by the researchers as the ability to use the words, rather than declarative knowledge, defined as the ability to define the words. The researchers classified responses as showing "good or some" understanding or showing "little or no" understanding. The researchers conclude that first year chemistry students would benefit from a better understanding of the vocabulary of error. Furthermore, they suggest that students will not learn the vocabulary through the textbooks in which they find a lack of clarity and consistency in the use of the vocabulary of error ([Tomlinson, et al., 2001](#)).

3. The study

3.1 Research Question

The study reported here is a pilot study of five words identified by the research team as possibly having lexical ambiguity: association, average, confidence, random and spread. The choice of these words will be discussed in subsequent detail. In order to "exploit" the lexical ambiguity of words and help students form strong mental connections between their existing word meanings and the statistical meanings, research must be done to ascertain the meanings of the words that are most commonly used by students. The research question for the study presented here was: What are the meanings of the five target words most commonly used by students entering an undergraduate statistics course? A secondary research question was whether it is possible to generate reliable and valid data about what students think a word means, or the definition that a student holds for a word, in particular, the five target words, through a pencil and paper task. If so, it would allow the research team to collect and analyze data from a larger and more diverse sample.

The research team began this project by brainstorming a list of words commonly used in college-level, algebra-based, introductory statistics courses that are believed to exhibit lexical ambiguity. The list of words is included in Table 1. The remainder of this section provides the basis on which the five study words were chosen. The nomination of the words to be used in this pilot study was done based on prior experiences the researchers had had in the classroom rather than on a review of literature. Since our study is the first that attempts to study language acquisition in statistics on a large scale, there was not a literature base from which to draw. Instead, the considerable experience in the classroom and research interests of the authors were used to choose the first set of words to study.

Table 1: List of words suspected to have lexical ambiguity in statistics (words examined in this study *italicized*)

Association	Control	Independence	Nominal	Range	Skew
<i>Average</i>	Correlation	Margin	Normal	Response	<i>Spread</i>
Bias	Distribution	Mean	Null	Sample	Standard
Blocking	Error	Median	Parameter	Scatter	Statistic
Center	Event	Minimum	Population	Significance	Statistics
<i>Confidence</i>	Experiment	Mode	<i>Random</i>	Simple	Variance

Association in statistics is a relationship between two variables. One of the researchers routinely asked students on an exam to describe the association present in two contexts: a two-way table and a scatterplot. There was evidence from the free responses given that students had difficulty interpreting the word *association*. In the case of the two-way tables, for example, students were supposed to describe the relationship between gender and transportation choice and instead students would write a response like, "the association is that both boys and girls have the same choices of school transportation." The practice of finding a commonality between two groups and calling that an association may stem from a common everyday use of the word *association*. An *association* tends to be a group of people who have joined together for a common goal or purpose or an affiliation between people. While these uses imply a relationship, they do not imply a relationship in statistical sense, in which certain values of one variable are more likely for certain values of the other variable.

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 \(2007\)](#), [Utts & Heckard, \(2004\)](#), or [Agresti & Franklin \(2007\)](#)). The experience of one member of the research team, however, both as a teacher and a statistical consultant,

shows that the word *average* is used in everyday language to have a variety of meanings that include what is "typical" and what is "normal". In addition, when used as a measure of center, many use *average* interchangeably with the ideas of "median" or even "mode". [Triola \(2006\)](#) specifically addresses this concern and says:

Unfortunately, the term *average* is sometimes used for any measure of center and is sometimes used for the mean. Because of this ambiguity, the term *average* should not be used when referring to a particular measure of center.

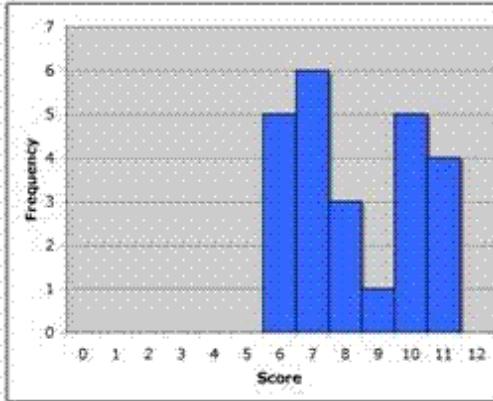
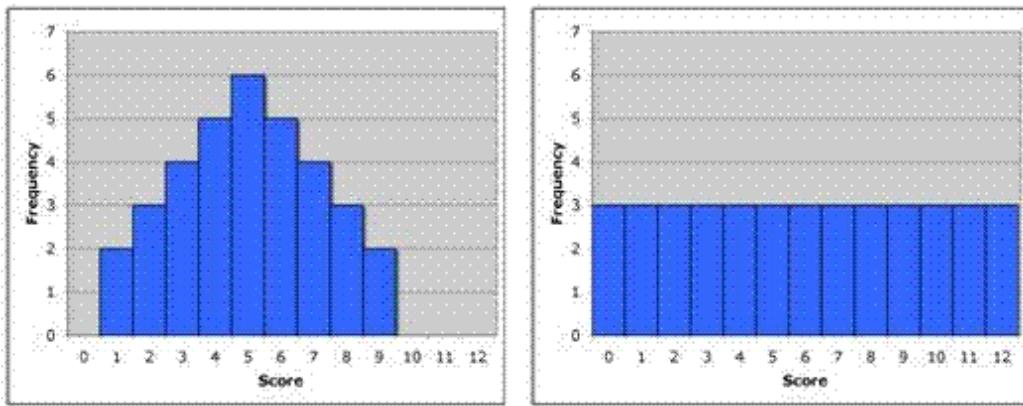
Because of the varying meaning that individuals can attach to the word *average* and thereby bring into an introductory statistics class, it is a word ripe for having lexical ambiguity.

Confidence in common usage is a trust, assurance, boldness or faith. In most usages it is assumed to be a word associated with strength of conviction. In only one definition in the Oxford English dictionary is a level of confidence discussed. In that sense, *confidence* is an assurance based on insufficient grounds or having an excess of assurance. In statistics, by contrast, *confidence* is associated with a probability. From the frequentist perspective used in most traditional textbooks, when a confidence interval is created or an interval estimate is given, there is the underlying assumption that confident does not imply being certain. Instead, *confident* is used in a probabilistic sense. It seems reasonable that this subtlety could be easily lost on beginning statistics students.

Random was nominated as a study word by one of the researchers because over the years it seemed to her to be the one word that is the most difficult to change what the students believe the word means.

Students have used the word *random* all their lives to mean haphazard, or to describe an event that was unlikely. In contrast, the statistical use of the word *random* implies considerable structure and a distribution of likelihood. To address the students' difficulties in understanding the statistical definition, the researcher not only talks in class about the importance of *random* in sampling and experimental design, but also gives several examples of court cases where the statistical definition of *random* is also the legal definition. In addition, this instructor does a "lottery" in class every semester where she chooses numbers randomly and gives out prizes to the lucky 3 winners. In class she takes note that sometimes all of the winners sit in the front or are all male. In other words, she points out that, when a random selection occurs, the outcome is not necessarily balanced or what seems fair.

The word *spread* in statistics is used as a synonym for variability or measures of dispersion. The measures of spread typically taught in an introduction to statistics course are: range, interquartile range, and standard deviation. While one common use of the word *spread* is to disperse or scatter, the word *spread* is also associated with an even covering. For example, in the acts of spreading butter on toast or a blanket over a bed one wants an even coating. Contrast that with the notion of estimating the relative measures of spread on a histogram. One of the researchers has asked students to compare two histograms and say which data set has higher variability or spread. Many students have indicated that the histogram with small range but irregular bar heights, such as that shown in Figure 1, has more spread than a histogram showing a normal or uniform distribution with large range, such as those shown in Figure 2. Given the common usage of the word *spread*, students may think that spread measures the evenness of the heights of the bars in the vertical direction when, statistically, spread is a measurement in the horizontal direction.

Figure 1: Histogram with small range and irregular bar heights**Figure 2:** Histograms with large range and regular bar heights

3.2 Research Design

A 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. 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. One class met on Mondays and Wednesdays from 1:00 to 2:15 pm and the other met on Tuesdays and Thursdays from 11:00 am to 12:15 pm.

There were approximately forty students enrolled in each section. Sixty-seven students completed the pretest of the pilot study, 45 women and 22 men. All of the students enrolled in the two sections were U.S. based students, and there is no evidence that any are English language learners. Twenty of the students (30%) were nursing majors; there were 25 other majors reported, such as psychology, advertising, health information management and biology, but no other major had more than 4 subjects. The distribution of the self reported GPAs of the subjects was unimodal and roughly symmetric with a mean GPA of 2.98 and standard deviation, 0.47. The distribution of self reported ages of the subjects was unimodal with right skew; the median age was 20 years and the middle 50% of the ages between 19 and 21. No students under 18 years of age were surveyed due to IRB constraints.

During the first week of class, before any of the five words were discussed, the students were given a questionnaire asking five sets of questions. For instance they were asked to

- a. Define or give a synonym for the word "association."
- b. Use the word "association" in a sentence.

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

3.3 Analysis

The research team made a list of the common definitions of each word in the study using the Oxford English Dictionary Online as a reference. The first researcher to code the data had those definitions in mind, but used the data to modify the definitions so the responses given by students could be reasonably mapped to the definitions. Once the first researcher had finished creating coding categories for the definitions and had coded all the responses, draft versions of coding categories and the instruments were then sent to the other two researchers. Those researchers independently coded the responses and suggested modifications and edits to the coding categories.

Coding data were compiled for all three raters. [Table 2](#) gives the agreement percents at each stage of the coding as well as the percent of the subject definitions that were ultimately coded. The first row of [Table 2](#) shows the percent of subjects for which all three coders agreed on the first pass of coding. The second row of [Table 2](#) shows the percent of subjects for which two of the three coders agreed after the first coding. For the cases in which there were exactly two coders who agreed, the coder who was in disagreement revisited that subject to see if the coding was an error or if he or she agreed without discussion to change the coding. Row three of the table indicates the percent agreement that was achieved after this stage of the coding.

Table 2: Agreement between coders

Word	Association	Average	Confidence	Random	Spread
Initial Agreement – 3 coders	87%	78%	92%	58%	55%
Initial Agreement – 2 coders	95%	91%	98%	84%	83%
Agreement without discussion – 3 coders	88%	82%	94%	68%	58%
Definitions classified in final coding	97%	96%	100%	91%	95%

Subsequently, the researchers discussed each case in which there was disagreement. At this point changes were made to the coding rubrics. In particular, the rubrics for *spread* and *random*, which had the lowest initial inter-rater reliability, underwent the largest changes. Originally the rubric for *spread* had three separate categories for scattering or dispersing over a large area depending on whether the verb was active, passive or reflexive. These categories accounted for 75% of the disagreements and, thus, were judged to be too fine grained an analysis for reliability. The research team decided that there was no qualitative difference between sentences like, "she spread the money out to count it," (active verb use) and "the disease spread across campus" (passive verb use) so the categories were collapsed into one. Within the rubric for *random*, the researchers struggled to decide whether it was necessary to separate definitions that indicated a random event from those that indicated random selection. In addition,

descriptions of categories were expanding to be more clear and inclusive. The words *confidence* and *association* had the highest initial inter-rater reliability and were the easiest to code. They are also the words with the fewest coding categories. The final coding rubric for each of the five words is discussed in the next section.

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 the secondary research question is whether such a study could be done a large scale. While it is possible to interview subjects to gain more insight into the meanings they hold for certain words, that is not the intent of this study. Examples of responses that could not be classified are included for each of the target words.

3.4 Results

This section contains the results for each word in alphabetical order. Each subsection begins with an overview of the results for that word and contains a table describing the definitions used by the students along with the number of students who used each definition. The tables are followed by examples of the definitions as well as examples of students whose definitions were either unable to be classified or were clearly an incorrect definition of the word.

3.4.1 Association

The two most popular uses of the word association were: a group of people who have come together for a common purpose, such as the American Medical Association or the National Basketball Association, and to be grouped in a loose affiliation. Many students who referred to loose affiliations used the term "guilt by association," but another example of a loose affiliation is, "she does not like to be associated with the girly girls." In contrast, a more formal relationship is characterized by statements such as, "The association between Brian and Andrew is that they are brothers." An example of a mental connection between two objects is given by the statement, "When the baby seen (sic) white objects he screamed and cried because of his association between white objects and loud abrupt noises."

Table 3: Definitions of Association (65 students able to be coded)

Definition	Number of Subjects
A body of persons who have combined together for a common purpose or to advance a common cause	36
To be affiliated, grouped with or in a loose relationship; union in companionship; fellowship, for example "guilt by association"	20
Statistical definition: an association exists between two variables if a particular value for one variable is more likely to occur with certain values of other variables Selecting without bias	4
A formal relationship between two people, objects and/or ideas	2
A mental connection between objects and/or ideas	2

Of the four students who were classified as giving the statistical definition of the word association, three had sentences that incorporated the correct statistical use of the word, but without matching definitions. One had a correct statistical definition without an accompanying sentence. This seems to indicate that even when students enter a statistics class with a notion of association as a relationship between variables, this idea is not well developed enough to fully articulate.

Correct Sentence without matching definition:

Sentence: The amount of ice cream eaten in the summer vs. the winter is associated with the temperature of the outside air.

Definition: statistically speaking, a grouping of information

Correct Definition without matching sentence:

Sentence: Math is an association with numbers.

Definition: Association exists when a value for a variable is more likely to occur with certain values of another variable

Three student responses for *association* could not be coded. In the first case, the student gave a definition "in relation to" and sentence, "Sam's club is close in association with Wal-mart." The research team could not infer whether the student was implying that there was a relationship between Sam's Club and Walmart, either formal or informal, or whether the student was indicating that the two stores tend to be in close physical proximity to each other. The second unclassifiable response was similarly vague. The definition was, "familiar" and the sentence was, "He is with the popular association." The research team could not discern whether the student was using the definition: a group of people with a common purpose, or whether the student was implying some type of informal relationship. Finally, the last unclassifiable response indicated a lack of understanding of the meaning of the word *association*. This student gave as a definition "when something is associated with something else" and sentence "Give the association of the two numbers."

3.4.2 Average:

The most common definition, given by 25 students, for the word *average* was typical or mediocre, something that is neither outstanding nor poor. Each traditional measure of center was represented by a category. A total of twenty-eight students gave one of these definitions for average: 15 gave mean, 11 gave median and 2 gave mode. Another four students discussed average as a value that represents most of the data, without specifying a specific measure.

Table 4: Definitions of Average (64 students able to be coded)

Definition	Number of Subjects
Ordinary, Normal, typical, mediocre, not extraordinary, common, neither outstanding nor poor, standard	25
Mean	15
Median or in the middle	11
Sum	6
Overall summary on something, general value that represents most of the data, overall outcome	4
Mode, most common number	2
Majority	2
A value we can use to compare one person's performance to the group	1
Combination of	1
The division of two statistics to get a whole answer	1

There were ten students who gave definitions that were not indicative of the center and another three students who had responses that were not obviously connected to the idea of average as a measure of center or what is typical. Of the six students who defined average as the sum of a group of numbers, three

wrote reasonable sentences about batting average, grade point average and class average. The other three wrote sentences that were difficult to interpret:

- She has an average height of 5'8".
- I found the average for the number of students who like the color pink.
- The average of high school students who made straight A's during the first semester is 35%.

Similarly, the other four students in this category either used a common phrase in their sentence, such as "average American girl" or used the word ambiguously. The counts in the table sum to 68 because four students gave responses that indicated more than one category. In each of the four cases, the students had two definitions from the three most common categories: typical, mean, or median.

The students whose definitions could not be categorized wrote sentences that made sense, but the sentences and definitions were common or vague enough that it was not possible to interpret what the word meant to the student. For example, it was unclear whether the following student was viewing average as "mean" or as "not extraordinary, normal."

Definition: normal or mean

Sentence: He was a C average student.

The following student is an example of an incomprehensible definition with a correct sentence.

Definition: a group of something that have around about the same height or weight. Just something others relate to

Sentence: What is the average weight you should maintain?

3.4.3 Confidence

Sixty-two of the 66 students who responded gave the same definition for confidence: belief in someone or something, self esteem, assurance or determination. Furthermore, all of the responses to this item were able to be coded. This, perhaps, indicates that student prior knowledge of the word *confidence* has little variability and that issues with this word might be easily addressed. On the other hand, the word *confidence* as used in the statistical term *confidence interval* implies a level of certainty as well as uncertainty. While the definitions were not formally coded in this way, some responses included explicitly the idea that confidence has levels while others implied that confidence is a synonym for only a strong belief.

Table 5: Definitions for Confidence (66 students able to be coded)

Definition	Number of Subjects
Self-esteem, belief in someone or something, assurance, boldness, determination	62
The confiding of secret matters to another	3
Statistical Definition	1

Example with explicit mention of levels of confidence

Definition: level of assurance

Sentence: Because I studied hard, I will have a high level of confidence while taking the test.

Example in which confidence is strong belief

Definition: believing in yourself greatly

Sentence: She gave her speech confidently in front of millions

Example in which understanding of level of confidence is ambiguous

Definition: belief in someone or thing

Sentence: I have confidence in myself that I will pass statistics.

3.4.4 Random

The most common use of the word *random* by students is an occurrence that is unplanned, unexpected or haphazard. Some examples are: "The apartment is a random arrangement of modern and traditional" and "During class some students ask very random questions that do not pertain to the lesson." Twenty-nine students gave this definition. A further 27 students used the word *random* to describe a method of choosing. There were four categories of types of choosing that students labeled as random: 1. choosing without criteria, plan or prior knowledge, 2. choosing without order or pattern, 3. choosing without bias and 4. choosing using a pattern without realizing it. Finally, there were five students who used the phrase "by chance" to define random and six students whose responses could not be categorized.

It is interesting to note how far from the statistical definition of a random event the students' notion of a random event is. In statistics, a random event is one for which no one outcome can be predicted, but there is knowledge of the long-term distribution of the outcomes. For students, in contrast, a random event is one that has very small chance of occurring or that happens with no warning (or no known underlying distribution). Furthermore, it is unclear that those students who give definitions that are close to a statistical notion of random, using terms such as "bias" and "chance", have a clear understanding of what the words they use actually mean. For example, a student gave the definition, "to pick without bias" and the sentence, "I randomly selected the color of my truck." It is unclear from the statements what meaning the word "bias" has for this student. Similarly, a student who gave the definition, "by chance" wrote the sentence, "They do random drug tests at my job." Does this imply that the subjects or dates for the drug tests are chosen "by chance" or that the drug tests are haphazard in their occurrence?

Table 6: Definitions of Random (61 students able to be coded)

Definition	Number of Subjects
An occurrence that has no definite aim or purpose, unplanned, haphazard, spontaneous, different	29
Selecting without prior knowledge or without criteria or agenda, not being specific in choosing, no definite way, non-repeating, blindly chosen	13
Selecting without order or pattern	10
By chance	5
Selecting without bias	3
Selecting unknowingly in some type of pattern	1

The six unclassified responses all had understandable sentences. In three cases, the definition was vague and the sentence use so common that the students thinking about the word random could not be inferred. For example, the definition "nothing in particular" is paired with the sentence "Tom took part in a random drawing for a car." Two other unclassified definitions were "awkward" and "non-repeating." The last unclassified response had the definition, "Anything of a certain topic or set" and sentence, "Pick me a random number from the box." The authors hypothesize that the student first wrote the sentence and then used the sentence to try to give a definition. The sentence is one that a student might have heard, but not understood. He may have thought that in choosing a number the word random is an adjective meaning

"one of the set that is in the box."

3.4.5 Spread

The most common definition for the word *spread*, given by 25 students, was: to distribute, disperse, separate or scatter to extend over or cover a large space. A further seventeen students used the word spread to mean cover evenly, as spreading butter or jam on bread. Eleven students used the word spread as a synonym for range or difference between numbers, often invoking the phrase "point spread." The correct meanings given by students were: to extend as in "spreading his wings" and butter, jam or dip. Three students' definitions could not be classified.

Table 7: Definitions of Spread (63 students able to be coded)

Definition	Number of Subjects
To scatter distribute, disperse; to go apart or separate; to extend over a larger space	25
To distribute in a thin layer, smear or cover evenly	17
Range or difference between two numbers as in point spread	11
Extend, open out as in spread his wings	3
Butter, jam, dip	3
Spreadsheet	2
A large group	1
Graph	1

The most interesting non-routine use of the word spread was in the word "spreadsheet." While spreadsheet is a legitimate English word, its use as a synonym for *spread* does not indicate that the student has an understanding of the meaning of the word *spread*. In fact, two of the students with responses that could not be classified used the word "spreadsheet" in their sentence. The following are the responses from the students who used "spread sheet" in their sentences.

Sentence: The doctor used a spread sheet to determine how many people had the disease in the United States.

Definition: large area covered

Sentence: I need to use a spread sheet for the test.

Definition: A word used to say that something was everywhere.

Sentence: I used a spread to display charts and graphs.

Definition: layout

Sentence: I looked at the information on the class spread sheet.

Definition: information gathered and documented

The research team hypothesized that these students may have written their sentences first and then created definitions based on the sentences. It may be that these students do not know what the word spread means; they know of one use of the word and tried to define the word based on that usage.

The other two uses of the word spread were not common, each given by only one student, but all were recorded in the event that further research uncovered more students with similar misunderstanding. One student used "graph" as a synonym for spread and then wrote the sentence, "The spread displays the

number of each food sold in the month of March." Another student wrote the sentence "We took a survey over a spread of random college students" and gave a definition "over several numbers of variables." This meaning was interpreted as "a large number or group." Finally, the last response that could not be classified was the definition "set of information" coupled with the sentence "the spread was very interesting." It was not possible to infer what the student knows about the word spread from this response.

4. Discussion

4.1 Summary of Findings

With regard to the primary research question, the meanings of the five target words most commonly used by students entering an undergraduate statistics course, the preliminary findings discussed above show each of these words to be problematic for different reasons. Average and spread both have a variety of meanings for the students entering an introduction to statistics class. In the case of average, many of these meanings that students hold at the beginning of the course use the mean to describe something in the middle. Unfortunately, with that diversity of prior understanding of the word *average*, it is unclear what students will assimilate about the statistical use of the word *average* by the end of the course. Spread is even more problematic. Not only are there many common uses for the word, but these uses are not consonant with the statistical use of the word. The students who associate "spread" with buttering toast are thinking about making an even coating. When translated to the idea of estimating standard deviation from a histogram, these students may view data with a uniform distribution as having a "good" type of spread. The diversity in the meanings of *average* and *spread* that students hold at the beginning of a course may make these words more difficult to address in the classroom.

Random and confidence were fairly convergent in terms of common usage from the student perspective. Because students have similar understandings of these words they are good candidates for a study that addresses student learning when lexical ambiguity is directly addressed in the classroom. The everyday meanings should be easy to address during instruction. Association was also a word with a fairly convergent meaning. Most students view an association as a grouping, either formal or loose. A few students think of association as a relationship, either idealized or concrete. These relationships, however, are focused on the similarities between the members of the groups. Linking the everyday and statistical meanings of *association*, therefore, should be done with care and will probably necessitate learning more about student's understanding of variables and relationships. The similarity of pre-existing definitions, however, suggests that the ambiguity surrounding the word *association* might be reasonably routine to address.

With regard to the secondary research question, findings are believed by the research team to show that this method of studying language provides a valid and reliable vehicle for learning what students know about the meaning of a word. Modifications of the design described here have been used to begin other research studies, which will be discussed in the section on future research.

4.2 Implications for Teaching.

The findings of the research discussed here have already changed the approach that the research team takes with language in our own classrooms. The first author has routinely begun to introduce the idea of random by saying, "in common usage random has come to mean haphazard, or something that happens with small probability; something that is really unlikely." I hear students say 'That comment she made in class was so random.' or 'I can't believe I just bumped into her; that was so random.' When something is statistically random, however, it is not haphazard at all. In fact it is quite structured. It is possible to know the likelihood of each outcome. For example, when a die is rolled, each number occurs with probability

1/6. None of the outcomes is particularly unlikely. There is a known expected distribution in the long run of rolling a die. If the die is rolled 600 times, roughly 100 instances of each number are expected to occur. What cannot be predicted in statistical randomness is the exact outcome on the next roll. One particular roll of the die cannot be predicted, but a lot is known about what should happen in the long run. In fact, the first author is mindful, every time she discusses *random* in class, whether talking about random events, random assignment or the taking of a random sample, to remind the students explicitly about the difference between the use of the word *random* in every day usage and in statistics.

In addition the research team has become more thoughtful in the use of *spread* as a statistical synonym for *variability* and the word *average*. While students tend to use *spread*, perhaps because it appears in the text or because the word does have meaning for them as a synonym for *variability*, the researchers tend to let students use the word in the classroom before using it themselves. In addition, more care is taken in the use of the word *average*, being specific about the measure of center under discussion and using *mean* or *median* as appropriate. *Confidence* and *association* provide a bigger challenge for instruction.

Statistical confidence from the frequentist perspective includes a level of surety based on probability, whereas many students enter the course using *confidence* to mean a high degree of assurance. Statistical association is focused on relationships with differences, whereas in common usage, association is focused on relationships with commonalities. It is unclear at this time how to best help students to understand these differences and clearly more work is needed in this area, which will be discussed in the section of future directions.

4.3 Limitations

This study was a small pilot study conducted with the students of one instructor at one institution. The findings, therefore, may not generalize to other institutions or students in different regions. It is possible that the coding rubrics developed in this study would not be valid or reliable for a more diverse population. With that in mind, the research team conducted a larger scale study with a similar design at three institutions, using at least two instructors at each site. Preliminary analysis of the data indicates that the rubrics discussed above do generalize to a more diverse population.

A second limitation occurs in the interpretation of the data. As instructors, grading student exams, an attempt is generally made to find any bit of meaning on the part of the student to award partial credit. This research, however, attempted not to do that, which led to the number of responses that were unable to be coded. While it is recognized that using interviews might allow us to understand more clearly the meanings all of our subjects have about the target words, basing a study on interviews would conflict with the goal of conducting a large-scale study. Furthermore, the goal of this study was to provide background information on the general use of the target words by college students, which the research team believes it does.

4.4 Future Directions for Research

This study reports the findings from the first stage of a multiple-stage study. The subjects described in this study were administered a post-test at the end of the semester asking them to provide both every day and statistical meanings for the target words. Those responses were used to create coding rubrics for the statistical meanings of the target words. These rubrics will be presented in a subsequent publication ([Kaplan, Fisher & Rogness](#)). In addition, the every day meanings were coded using the rubrics described in this paper. Furthermore, data on the target words has been collected from the three institutions of the researchers and a subset of those data have been used to validate the every day and statistical coding rubrics created based on the pilot data. The authors have chosen five additional words, *bias*, *error*, *independent*, *normal*, and *significant* and have collected pre- and post-test data from students at their institutions. Coding rubrics will be created and validated for the second set of words in a similar fashion.

In the future, the research team will be working with the linguistic software SPSS Text Analysis for Surveys to aid in the coding of the data. In addition, a review of commonly-used introductory statistics textbooks should be done to determine if and how these words are used by the authors. The findings from 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, there is also a plan to measure the impact the ambiguity has on student performance in statistics. The research team is 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 overburdened curriculum.

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 both designing the research and refining the paper. 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

Adams, T.L., Thangata, F., & King, C. (2005). "Weigh" to go: Exploring mathematical language. *Mathematics Teaching in the Middle School*, 10, 444-448.

Agresti, A. & Franklin, C. (2007). *Statistics: The Art and Science of Learning from Data*. Upper Saddle River, NJ: Pearson Education, Inc.

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education* 19(2), 118 – 126.

Durkin, K. & Shire, B. (1991a). Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds.) *Language in Mathematical Education: Research and Practice*. Philadelphia, PA: Open University Press, 71 – 84.

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

Kaplan, J.J., Fisher, D. & Rogness, N. (2009). Lexical Ambiguity in Statistics: What do students learn about the words association, average, confidence, random and spread? Manuscript submitted for publication.

Khisty, L. & Chval, K. (2002). Pedagogic discourse and equity in mathematics: When teachers' talk matters. *Mathematics Education Research Journal*, 14(3), 154 – 168.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1), 1 – 11, <http://www.amstat.org/publications/jse/v3n1/konold.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), 127 – 135.

Makar, K. & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27 – 54.

Monaghan, F. (1999). Judging a word by the company it keeps: The use of concordancing software to explore aspects of the mathematics register. *Language and Education*, 13(1), 59 – 70.

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

Schultz, T. & Pilon, R. (1973). Development of the ability to detect linguistic ambiguity. *Child Development*, 44, 728 – 733.

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

Tomlinson, J., Dyson, P. & Garratt, J. (2001). Student misconceptions of the language of error. *University Chemistry Education*, 5, 1 – 8.

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

Utts, J. & Heckard, R. (2004). *Mind on Statistics, 2nd edition*. Belmont, CA: Brooks/Cole-Thompson Learning.

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

[Volume 17 \(2009\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Resources](#) | [Editorial Board](#) | [Guidelines for Authors](#) |
[Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)