

CONFIDENCE MEANS WHAT?!?: LEXICAL AMBIGUITY IN THE INTERPRETATION OF CONFIDENCE INTERVALS

Jennifer J. Kaplan and Kristen E. Roland
 Middle Tennessee State University, Murfreesboro, TN
 Appalachian State University, Boone, NC
Jennifer.kaplan@mtsu.edu

Associations of Statistics and Psychology recommend reporting interval estimates, such as confidence intervals, when communicating data-based results. Previous research in statistics education, however, has identified the word confidence as having lexical ambiguity for students learning statistics. This study investigates the nature of this ambiguity with the interpretation of confidence intervals. Analysis of transcripts from task-based interviews revealed four themes, (a) sureness/belief, (b) confident in outcome, (c) confusion of chance/probability, and (d) equivalence of chance/probability, none aligned with the statistical concept definition when the word is applied to confidence intervals. The results of this study identify confusion between the words confident, sureness, probability, and chance as areas for further research as well as classroom attention. Researchers should continue to investigate these issues while instructors exploit these ambiguities.

INTRODUCTION

Making sense of the modern level of news reporting requires global citizens to be knowledgeable consumers of data and results of statistical inference. Introductory statistics courses at the tertiary level, whether taught using traditional frequentist methods or randomization techniques, typically include the study of two inferential techniques: hypothesis testing and confidence intervals (CIs). As early as 2001, the American Psychological Association (APA) recommended that published research include CIs and requested the removal of reports of hypothesis testing (APA, 2001). Similarly, the American Statistical Association (ASA) issued a statement cautioning against the over-reliance of *p*-values in scientific research (Wasserstein & Lazar, 2016) and strongly suggested reporting point and interval estimates and/or effect sizes accompanied by measures of uncertainty, such as standard errors or intervals (Wasserstein et al., 2019). This should lead to more emphasis on instruction of CIs in tertiary introductory statistics courses.

Published research about how individuals understand CIs was sparked, in part, by the internal debate among researchers of the APA concerning the use of hypothesis testing over CIs in published works (Wilkinson, 1999). Studies of published researchers, particularly in the fields of psychology, neuroscience, and medicine indicated that most researchers did not understand CIs fully. Within the statistics education field, research focused on student misconceptions of the interpretation of CIs (e.g., Crooks, 2014; Fidler, 2006) and on aids for instruction (e.g., Bertie & Farrington, 2003; Gordon & Gordon, 2020; Hagtvædt et al., 2008). While these studies are valuable, they do little to illuminate how students or researchers come to these misconceptions nor how to help them build normative conceptions of a confidence interval.

Complicating the learning of CIs is the lexical ambiguity associated with the word *confidence*. Lexically ambiguous words are those with multiple meanings (Barwell, 2005). If a commonly used English word, such as *confidence*, is co-opted by a technical domain such as statistics, the first time students hear such a word in their statistics course they may incorporate the technical usage as a new facet of the features of the word they learned or knew previously. This may encourage students to make incorrect associations between the common English meaning of words and the technical use of the words, in this case *confidence* in the domain of CIs.

BACKGROUND

According to Kaplan et al. (2009) in common usage *confidence* means trust, assurance, boldness, or faith and is generally

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. (p. 7)

In fact, Kaplan et al. (2009) reported that an overwhelming majority of undergraduate students at the beginning of a statistics course define confidence as assurance, boldness, self-esteem, or a belief in someone or something. Kaplan et al. (2010) reported undergraduate students' statistical definitions of the word *confidence* at the end of a one semester introductory statistics course. The most frequent category of student responses was a level of surety or certainty (50%), which was broken into three sub-themes: (a) about the location of a value (9%), (b) of something vague or unspecified (35%), and (c) that something is correct (6%). A smaller proportion defined confident implying a high level of certainty (11%). Kaplan et al. (2010) were troubled by the large proportion of students responses that lacked thematic coherence or connection to a normative meaning of *confidence*.

The findings from Kaplan et al. (2010) are particularly concerning given the interpretation of a CI discussed in many United States introductory statistics textbooks (e.g., Agresti et al., 2017; De Veaux et al., 2018; Larson & Farber, 2019; McClave & Sincich, 2017; Triola & Iossi, 2018) and courses that use these textbooks, which for a 95% confidence interval is: We are 95% confident that the parameter is contained within the calculated interval. This interpretation, however, is not without controversy. Morey and colleagues (2016) also expressed concern about using the word confident in the interpretation of a Neyman (frequentist) confidence interval. Morey et al. (2016) defined three fallacies associated with the interpretation of a CI. Of particular relevance to this study and aligned with the findings of Kaplan et al. (2010) is the *Fundamental Confidence Fallacy*, the idea that people tend to interpret the word confident as a probability. Thus, Morey et al. (2016) argue the textbook interpretation of a CI is incorrect and suggest an increased reliance on Bayesian inferential methods. Although the issues raised by Morey et al. are valid, their recommendations require a major overhaul of the introductory statistics curriculum. In fact, Hoekstra et al. (2018) explored Dutch statistics textbooks to determine how many books contained the Fundamental Confidence Fallacy. Of the 23 books analyzed, 14 (61%) of them contained the Fundamental Confidence Fallacy, leading Hoekstra and colleagues to conclude the textbooks are contributing to the confusion in the interpretation of a CI and the word confident.

Callaert (2007) also commented on the difficulty surrounding the use of the word *confident* in interpreting the confidence interval. From his experiences at Advanced Placement (AP) statistics readings, he concluded that students had learned that confident was the correct word to use when interpreting a confidence interval, but like Kaplan et al. (2010), he questioned whether students knew what the word meant in the technical sense. In discussions Callaert had with other AP readers, other instructors expressed concern about the blanket teaching of *confident*, indicating that students and teachers alike may read "95% confident" and "95% probability" as equivalent (Callaert, 2007) and providing a potential explanation to the findings of Kaplan et al. (2010). From these studies, the interpretation of a CI and the meaning of the word confident are not as straightforward as it would appear.

METHODOLOGY

This paper reports a subset of results of a larger study designed to illuminate students' demonstrated knowledge of the interpretations of a CI and confidence level. The theoretical framework of the larger research project was based in the development of a formal concept image (Tall & Vinner, 1981) for the concept of CIs. Roland and Kaplan (2022) present the formal concept image for the concept of CIs, which was created from the theoretical definition of a CI. The theoretical framework and education experience of the participants aided in the decision to use the following constructs:

1. The interpretation of a CI is: We are confidence-level-% confident that the parameter is contained within the interval
2. The word *confident* in this sentence is defined as: confident in the random process that was used to derive the CI estimator with a given coverage probability, and the process is expected to produce confidence-level-% CIs that contain the parameter under repeated sampling.

In the definitions provided above, *confident* implies a confidence in the random process associated with deriving the CI estimator with corresponding coverage probability rather than the probability that any one CI estimate (outcome) contains the value of the unknown parameter. Therefore, the interpretation of the frequentist CI estimate should not include the words probability or chance and instead should use

the word *confident* (for a discussion about the difference between use of *confident* and probability within the interpretation of a CI estimator and a CI estimate, see: Roland & Kaplan, 2022). Furthermore, in the context of the interpretation of CIs, confidence is not equivalent to or a synonym for probability or chance, and probability and chance were treated as equivalent words or synonyms within this study.

Participants in the study were recruited from introductory, intermediate, and senior level statistics classes and from the statistics graduate program at a large research-focused university in the southwestern part of the United States. The participants were volunteers but received a small financial incentive to participate in the study. Participants engaged in three clinical interviews over the course of the project, only two of which were analyzed for this study. The interviews were designed to elicit the conceptualizations participants had about the interpretation of CI and interpretation of the confidence levels in a conversational way around a series of probing task questions. Interviews of 11 of the 16 subjects are included in the analysis presented in this paper. Three participants were removed from the study due to data collection issues and/or demonstrated knowledge about CIs below what would have been expected from a post-instruction introductory statistics student. One participant self-removed from the study after the third interview. The final participant was removed by the researcher because of a demonstrated alternative working definition of probability, rather than the frequentist probability that underpinned the work in this study. Of the 11 participants, one (Tiana) was an introductory student, three (Kiara, Aiden, and Logan) were intermediate students who were business majors, four (Diana, Brody, Emma, and Gabe) were senior statistics students, and three (Liam, Jace, and Joel) were statistics PhD graduate students. Jace and Joel were the only two non-native English speakers in the analyzed data presented in this analysis.

Over the course of the first two interviews, participants were asked to interpret CIs and confidence levels, and discuss statements with normative and non-normative conceptualizations of interpretations. The data for this paper came from three tasks. Task 1 appeared on the first interview and asked participants to interpret a previously calculated 95% CI and discuss what the 95% represented. Two friends, Jamie [a recurring character in the interviews who is a newspaper reporter at a fictional university, Hill Top State University (HTSU)] and Alex, created a group playlist for a long car ride. After 33 of 50 songs played using random shuffle that had been loaded by Jamie, the pair estimated the proportion of songs on the playlist that belonged to Jamie.

Among other questions, the following two task sub-questions were posed:

- Task 1.1. How should Jamie and Alex interpret the 95% CI you just calculated?
- Task 1.2. Jamie and Alex cannot remember what the 95% represents in the calculation and the interpretation. How would you remind them what the 95% represents?

Tasks 2 and 3 were on the second interview. Task 2 contained several statements asking the participants to reflect on their understanding and belief about correctness of each statement:

- Task 2.1. Jamie is 93% confident that the actual mean monthly rent for all students at HTSU is between \$705 and \$793.
- Task 2.2. Jamie is 93% sure that the actual mean monthly rent for all students at HTSU is between \$705 and \$793.
- Task 2.3. Jamie has a 93% probability that the actual mean monthly rent for all students at HTSU is between \$705 and \$793.
- Task 3. *Jamie's Colleague*. Participants were asked to comment on Jamie's colleague's statement: Jamie talked to a fellow reporter about constructing 93% confidence intervals. Jamie's colleague said that prior to collecting his sample, there is a 93% probability that the confidence interval will capture the actual mean monthly rent of all HTSU students. The colleague continued the explanation by saying that once Jamie collected a sample, the probability of the [previously constructed 93% confidence] interval (\$705, \$793) actually containing the mean monthly rent of all HTSU students is now either 0 or 1.

The corpus of data from the larger study from which these results were taken was over 19 hours of video, all of which were transcribed. (For a detailed description of the treatment of the data and its analysis, see Roland, 2020.) In the initial analysis phase, it became apparent that participants held very different definitions and conceptualizations of the words *confident* and *confidence level*. After these differences were identified, a more thorough thematic analysis (Braun & Clarke, 2012) was conducted

to uncover themes among these definitions and conceptualizations. This paper presents the themes found among definitions and conceptualizations of the word *confident*.

RESULTS

Through the thematic analysis process, four themes arose from the participants' personal and implied definitions of confidence: (a) sureness/belief, (b) confident in the outcome, (c) confusion of confidence/chance/probability, and (d) equivalence of confidence/chance/probability.

Sureness/Belief

The first theme that arose was from responses that defined the word *confident* as being equivalent to the words *sure* and/or *belief* or described the word *confident* as a measure of sureness and/or belief. These responses were quite specific in nature and were evidenced in the responses of five subjects: Liam, Aiden, Kiara, Gabe, and Tiana. Liam and Gabe implied belief in their definition: "Jamie has, like 93% belief that the actual mean falls within those two numbers" (Liam). Aiden referred to *confident* as "an arbitrary declaration," implying belief or likelihood. Kiara stated that confidence implied "more likely than not ... you're almost super certain ... way more likely than not, like a 51%." For Aiden, Kiara, and Tiana, the equivalence appeared to be a slippery slope towards defining the word *confident* in terms of probability. These three stated explicitly that *confident* does not imply probability, but their discussions in the two interviews suggested they had come to understand there is probability associated with the CI estimate.

Confident in the Outcome

Responses that support the second theme, *confident in the outcome*, typically occurred when subjects were asked specifically to define the word *confident* as used in the interpretation of the confidence interval. Four senior and graduate participants (Brody, Liam, Diana, and Gabe) and an intermediate participant (Logan) appeared not to have thought about what the word *confident* meant in the textbook interpretation of the confidence interval. For these participants, *confident* meant the interval was "doing what it's supposed to," "representing whatever you're trying to represent," or a way to "quantify" the certainty about the interval. As Brody said, *confident* is a "quantifiable measure of how good you feel about your estimate." Notice these definitions do not explicitly relate the word *confident* to sureness or belief. Instead, these definitions demonstrate a general acceptance of the word *confident* in the interpretation without explicit connection to the random process. As a sub-theme of this theme, Brody, Liam, Gabe, and Logan expressed that the use of the word *confident* was a way to connect the interpretation of the CI to the name of the statistical method being used. Logan stated: "*confident* implies that we are estimating, and we are in the percentage level of confidence in that estimate, which is why I like the word." Although four of the five participants with conceptualizations in this theme were able to state correct interpretations of a CI and confidence level, it is unclear whether either of these meanings are helpful definitions for *confident* in the statistical context because of the lack of connection to the random process.

Confusion: Confidence/Chance/Probability

The next theme contains responses that demonstrated *confusion over the lack of equivalence of the words probability, chance, and confidence*. There were two types of confusion with respect to the equivalence of the three words: (a) confusion about the word *confident* that has been accepted as 'fact,' and (b) contradictory statements about the equivalence of the words across the interviews. Joel and Liam stated they had previously struggled with the use of the word *confident* in the interpretation of a CI. Liam stated his struggle with understanding the usage of the word *confident* in the interpretation started in AP statistics. He was able to convince himself using an explanation that *confident* and probability cannot be equivalent because of a capture/not capture scenario: once a CI estimate has been calculated, the interval either captures $P(\text{Capture} = 1)$ or does not capture the parameter $P(\text{Capture} = 0)$. Joel, on the other hand, simply learned that the word *confident* (in both English and Chinese) was to be used: "to be honest, I don't know [referring to the difference between probability and confidence]. I have the same question before, but I did not find the answer. ... Since ... the official name [is] confidence interval. I just follow that." Aiden, Kiara, and Tiana, in contrast, originally interpreted the definition of *confident* as equivalent to chance. Over the course of the two interviews, however, they eventually stated the two

words were not equivalent. Aiden's contradiction was immediate: stating that he felt chance was never the right word to use as an explanation of the definition of confident during the first interview. Kiara and Tiana, however, both stated that confident and chance were equivalent on the first interview but that the two words were not equivalent during the second interview.

Equivalence: Confidence/Chance/Probability

The final theme contained responses that stated that confident was *equivalent to chance and probability*. During both interviews, Logan, Emma, and Jace stated explicitly that the words confidence, sure, chance, and probability were equivalent. In Jace's case, there may have been a translation issue. The subtle difference between these words in the English language is difficult to understand for native English speakers. The difference in meaning, statistically, is quite large but may be difficult for non-native English speakers to understand.

DISCUSSION

Four definitional categories for the word *confident* as used when subjects interpreted CIs emerged from this study: (a) Sureness/Belief, (b) Confident in the Outcome, (c) Confusion about Confident, Chance, and Probability, and (d) Equivalence of Confident, Chance, and Probability. The *Sureness/Belief* conceptualization, which overlapped with categories found by Kaplan et al. (2009), defined the word confident as equivalent to sure, certainty, and belief, all of which can be used to express ideas of formal or informal probabilities. The Confident in the Outcome category of conceptualizations of the word confident contained statements that implied the word confident meant that the CI was "doing what it was supposed to" or was referring to the statistical procedure being conducted and were elicited when subjects were specifically asked to define the word confidence in the context of the interpretations of a confidence interval. The *Confusion about Confidence, Chance, and Probability* conceptualization contained responses that either expressed initial confusion or responses that were contradictory over the course of the two interviews about the equivalence of the words confidence, chance, and probability. The responses in the *Equivalence of Confidence, Chance, and Probability* category stated explicitly that the three words were equivalent.

Given the number of subjects who expressed either confusion about the relationship between confidence and chance or probability or who believed confidence is equivalent to chance or probability, aligning with the Fundamental Confidence Fallacy (Morey et al., 2006), further research is needed to explore how to help students understand the word confidence in the statistical context of CIs. Kaplan et al. (2014) suggest exploiting the lexical ambiguity associated with technical terms and providing students opportunities to build their voices around lexically ambiguous terms as instructional strategies to overcome adverse learning effects associated with lexical ambiguity. In order to design instructional interventions, more must be known about how students develop the notion that confidence is equivalent to probability, the Fundamental Confidence Fallacy, in statistics so instruction can disrupt that pathway. More investigation of the confusion that students have about the meaning of the word confidence in statistics relative to probability may illuminate effective instructional strategies that can be used when CIs are introduced to exploit the ambiguity of the word confidence and reduce the initial student confusion. Finally, more research is needed to investigate whether using sure or belief as equivalent to confident is productive in helping students understand the way the word confidence is used in statistics as well as the interpretation of the CI and/or the interpretation of the confidence level.

This is the first study to investigate the interaction between students' definitions of the word confident and their understanding of the concept of CIs. Given previous findings around the ambiguities associated with the use of the word confident, these results have the potential to impact instruction in a way that will increase statistical literacy around confidence intervals.

REFERENCES

- Agresti, A., Franklin, C., & Klingenberg, B. (2017). *Statistics: The art and science of learning from data* (4th ed.). Pearson.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). American Psychological Association.
- Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education*, 19(2), 117–125. <https://doi.org/10.1080/09500780508668667>

- Bertie, A., & Farrington, P. (2003). Teaching confidence intervals with Java applets. *Teaching Statistics*, 25(3), 70–74. <https://doi.org/10.1111/1467-9639.00134>
- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology* (Vol. 2, pp. 57–71). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Callaert, H. (2007). Understanding confidence intervals. In D. Pitta-Pantazi & G. Philippou (Eds.), *Proceedings of the Fifth Congress of the European Society for Research in Mathematics Education* (pp. 692–701). ERME. <http://erme.site/cerme-conferences/cerme-5/>
- Crooks, N. (2014). *Does comparison promote gains in conceptual knowledge? The case of learning about confidence intervals* [Doctoral dissertation, University of Wisconsin-Madison]. ProQuest.
- De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2018). *Intro stats* (5th ed). Pearson.
- Fidler, F. (2006). Should psychology abandon p values and teach CIs instead? Evidence-based reforms in statistics education. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS7)*. ISI/IASE. https://iaese-web.org/documents/papers/icots7/5E4_FIDL.pdf?1402524965
- Gordon, S. P., & Gordon, F. S. (2020). Visualizing and understanding hypothesis testing using dynamic software. *PRIMUS*, 30(2), 172–190. <https://doi.org/10.1080/10511970.2018.1534295>
- Hagtvedt, R., Jones, G. T., & Jones, K. (2008). Teaching confidence intervals using simulation. *Teaching Statistics*, 30(2), 53–56. <https://doi.org/10.1111/j.1467-9639.2008.00308.x>
- Hoekstra, R., Morey, R. D., & Wagenmakers, E.-J. (2018). Improving the interpretation of confidence and credible intervals. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan*. ISI/IASE. https://iaese-web.org/icots/10/proceedings/pdfs/ICOTS10_8A2.pdf?1531364291
- Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2009). Lexical ambiguity in statistics: What do students know about the words association, average, confidence, random and spread? *Journal of Statistics Education*, 17(3). <https://doi.org/10.1080/10691898.2009.11889535>
- Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2010). Lexical ambiguity in statistics: How students use and define the words: Association, average, confidence, random and spread. *Journal of Statistics Education*, 18(2). <https://doi.org/10.1080/10691898.2010.11889491>
- Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Exploiting lexical ambiguity to help students understand the meaning of random. *Statistics Education Research Journal*, 13(1), 9–24. <https://doi.org/10.52041/serj.v13i1.296>
- Larson, R., & Farber, E. (2019). *Elementary statistics: Picturing the world* (7th ed.). Pearson.
- McClave, J. T., & Sincich, T. (2017). *A first course in statistics* (12th ed.). Pearson.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. -J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103–123. <https://doi.org/10.3758/s13423-015-0947-8>
- Roland, K. E. (2020). *Student conceptualization of the interpretation of the confidence interval and the confidence level: Identifying similarities and differences in student concept images of confidence intervals* [Unpublished doctoral dissertation]. The University of Georgia.
- Roland, K. E., & Kaplan, J. J. (2022). Using concept images as a framework for the concept of confidence intervals. In S. S. Karunakaran & A. Higgins (Eds.), *Proceedings from the 24th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 490–497). RUME. <http://sigmaa.maa.org/rume/RUME24.pdf>
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2), 151–169. <https://doi.org/10.1007/BF00305619>
- Triola, M. F., & Iossi, L. (2018). *Elementary statistics* (13th edition). Pearson.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA’s statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$.” *The American Statistician*, 73(S1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>