Routledge
Taylor & Francis Group

Confidence Trick: The Interpretation of Confidence Intervals

Colin Foster

School of Education, University of Nottingham, Nottingham, United Kingdom

Abstract: The frequent misinterpretation of the nature of confidence intervals by students has been well documented. This article examines the problem as an aspect of the learning of mathematical definitions and considers the tension between parroting mathematically rigorous, but essentially uninternalized, statements on the one hand and expressing imperfect but developing understandings on the other. A small-scale study among schoolteachers sought comments on four definitions expressing differing understandings of confidence intervals, and these are examined and discussed. The article concludes that some student wordings could be regarded as less inaccurate than they might seem at first sight and presents a case for accepting a wider range of more intuitive understandings as a work in progress.

Résumé: La fréquente mésinterprétation de la nature des intervalles de confiance de la part des étudiants est bien documentée. Cet article analyse la question en tant qu'aspect de l'apprentissage des définitions mathématiques, et considère la différence entre d'une part la répétition d'énoncés parfaitement rigoureux sur le plan mathématique, mais qui n'ont pas été intégrés, et d'autre part l'expression de concepts encore imparfaitement maîtrisés, mais qui dénotent une certaine compréhension. Une étude réalisée auprès d'un petit groupe d'enseignants a sollicité leurs commentaires au sujet de quatre définitions exprimant différents degrés de compréhension du concept d'intervalle de confiance, commentaires qui ont ensuite fait l'objet d'une analyse et d'une discussion. L'article conclut que certains énoncés des étudiants sont moins inexacts qu'ils ne pourraient sembler à priori, ce qui suggère qu'on peut accepter une plus vaste gamme d'énoncés intuitifs comme l'expression d'une 'compréhension en devenir'.

Regardless of the text, there is almost invariably a peculiar pair of caveats presented as from on high: Never accept the alternative hypothesis, and never say the probability is 0.95 that the mean lies in a 95% confidence interval for the mean.

Dillon (2011, p. 34)

INTRODUCTION

It is one thing to assess students' procedural fluency at a particular mathematical technique, but it is quite another to attempt to uncover their underlying understanding. Skemp (1976) famously

contrasted an *instrumental understanding* of how to get the answer with a more *relational understanding* of the bigger picture, including the interconnections between the different elements in the process. In a related formulation, Sfard (1992) distinguished between *operational* and *structural* approaches to mathematical concepts. A student engages operationally when he or she is focused on the dynamic process of achieving a particular transformation. The complementary structural aspect entails awareness of a more static, abstract, reified entity, and this can be difficult to achieve. Asking students to explain their answer, or the ideas behind it, is one way of attempting to develop and assess the depth of their understanding, but it may be that explanations that satisfy the teacher can also be learned off pat, just as easily as recipes for performing some technique. Being able to prove a geometric theorem, for instance, may indicate a strong grasp of geometry but, on the other hand, it could simply mean that the student has memorized a proof. Students may take the pragmatic position that they have to learn a method and then, in addition, learn a sociomathematically acceptable explanation for the method. The quote attributed to Nietzsche, "It is hard enough to remember my opinions, without also remembering my reasons for them," comes to mind. Mathematical understanding is not supported by students electing to memorize an explanation to go with a particular mathematical procedure.

The statistical topic of confidence intervals would seem to present an interesting example of an area where students may adopt a double-speak strategy of thinking in one way privately but conforming in the classroom to a form of words that is acceptable to the teacher—and that they hope will ultimately satisfy an examiner. It is common for teachers to characterize students' informal articulations regarding confidence intervals as evidence of deep-seated misconceptions and to seek to confront or challenge these by various means so that students can overcome them (Kalinowski, 2010). However, Smith, diSessa, and Roschelle (1993) charted students' development from views that might be termed *misconceptions* through to more expert articulations, taking issue with what they regard as dismissive attitudes towards students' preexisting ideas. Adopting a constructivist perspective, they promoted the view that students' "prior knowledge is the primary resource for acquiring new knowledge" (p. 151), claiming that:

> Persistent misconceptions, if studied in an evenhanded way, can be seen as novices' efforts to extend their existing useful conceptions to instructional contexts in which they turn out to be inadequate. Productive or unproductive is a more appropriate criterion than right or wrong. (p. 147)

In this article, I explore common student explanations of confidence intervals and question the appropriateness of regarding these as evidence of misconceptions.

CONFIDENCE INTERVALS AND THEIR MISINTERPRETATIONS

Many users of statistics have argued that confidence intervals may in certain circumstances be preferable to hypothesis testing (Gardner & Altman, 1986; Nakagawa & Cuthill, 2007), although others have disagreed (Poole, 1987). A confidence interval is a way of describing the reliability of an estimate. A statistician may be interested in the mean (or some other measure) of a population but, except in rare cases, data will not be available for the entire population and a sample must be taken instead. When the mean is calculated from a sample, it is unlikely to match precisely the mean of the entire population, but it can be used as an estimate of the (unknown) population mean. How good an estimate it is will depend on the size of the sample: other things being equal,

a bigger sample will give a better estimate for the population mean. Specifying a confidence interval is more informative than just giving the sample mean, because it provides a band of values either side of the sample mean. A 95% confidence interval would have a wider band than a 90% confidence interval for the same data, because the more certain you wish to be about capturing the population mean, the less tightly you can set the limits. If you repeatedly use 95% confidence intervals, then on 95% of occasions the population mean will be captured in your interval—and one time in 20 it will lie outside it.

The precise form of words used to describe this capturing process is the issue at stake in this article. Anecdotal accounts support the quotation at the beginning that, for many students, besides the business of learning how to calculate confidence intervals, there is the additional requirement—frequently presented as little more than examination technique—to express their interpretation using sociomathematically acceptable language. The reasons for the prohibitions on certain formulations may be poorly understood by the students, making success a matter more of social conformity than of mathematical sophistication.

It is widely reported that students (Kalinowski, 2010)—and not just students (Belia, Fidler, Williams, & Cumming, 2005)—frequently misinterpret confidence intervals. Fidler (2006) found that university statistics students exhibited a variety of misconceptions in relation to confidence intervals, and he divided these into *definitional* (relating to what a confidence interval is) and *relational* (to do with "how the various determinants of a [confidence interval] affect each other"; Fidler, 2006, p. 210). Hagtvedt, Jones, and Jones (2008) reported that "students typically believe that a given parameter is contained in a confidence interval with a known probability" (p. 53), commenting that none of the 30 students they tested before they embarked on their teaching was able to offer what they regarded as a correct interpretation. Robison-Cox (1999) perceptively characterized the essence of the error as placing confidence in the *interval* found rather than in the *process* by which it is found:

> Students mistakenly say "The probability that μ [the population mean] lies in [7.5, 9.2], the computed interval, is 0.90" instead of saying "The process by which the interval [7.5, 9.2] was computed yields intervals which include μ 90% of the time." (p. 81)

He then went on to describe a practical experiment (repeatedly throwing a tennis ball at a blackboard) designed to illustrate the meaning of a confidence interval, commenting that "we do not know if a particular interval has covered the target, we can only have confidence that the process will cover the hidden target a certain percent of the time in the long run" (Robison-Cox, 1999, p. 82). Others have constructed similar practical activities (Richardson & Haller, 2003) and used information technology to attempt to illuminate the nature of a confidence interval (Bertie & Farrington, 2003; Cumming, 2007).

MATHEMATICAL DEFINITIONS

Mathematical definitions, viewed as statements about what some mathematical object or idea is, are a vital aspect of learning mathematics (Morgan, 2005). It might seem that the teacher has little option but to *give* definitions to students and that students have little option but to accept them without argument. Hewitt (2001) distinguished between the *arbitrary* and the *necessary* in the learning of mathematics and advocates *telling* students things that are mathematically

arbitrary (e.g., what we mean by the word *mean*) but actively *avoiding* telling students anything that is mathematically necessary (e.g., what happens to the overall mean when two data sets of known mean are combined). Real mathematics concerns *establishing* the necessary, and educators committed nontrivially to constructivism will insist that this should be done *by* learners and not *for* them. On this basis, the definition of a confidence interval could be seen as an arbitrary fact, and students might simply be told a good definition and asked to remember and use it.

However, definitions are not always simply arbitrary statements that need to be remembered; frequently they need to be worked on in order to be understood. Tall and Vinner (1981) used the term *concept definition* to describe "a form of words used to specify [a] concept" (p. 152). The concept definition is a formal and mathematically accurate statement, given by the teacher or developed from student discussion, that accords with that accepted in the wider mathematical community. They contrasted this with what they call the *concept image*, by which they mean "the total cognitive structure" that a student associates with a concept, including "all the mental pictures and associated properties and processes" (p. 152). When absorbed in the details of a mathematical problem, the student may find that it is the concept image, rather than the concept definition, that controls their thoughts and actions. Tall and Vinner (1981) highlighted possible conflicts between the two, which may be completely unrecognized by the student. Vinner (1991) commented that:

> There is no harm if the students memorize the formal definition and repeat it in various occasions. ... But [the teacher or textbook writer] should have no illusions about the cognitive power that this definition has on the student's mathematical thinking. (p. 80)

Although students may have been given a sound definition by the teacher, what they have in fact understood by it—or take it to be in practice—may be quite different.

Edwards and Ward (2004) contrasted what I would term *descriptive* and *prescriptive* definitions. They compared Landau's (2001, p. 165) "extracted definitions," which are *descriptive* definitions reflecting the way words are actually used in practice, with Robinson's (1954, p. 59) "stipulative definitions," which are *prescriptive*—imposed with a clear intention that others should follow them. Edwards and Ward (2004) "think of mathematical definitions as stipulated, whereas most 'everyday language' definitions are extracted" (p. 412), thereby identifying stipulative definitions with Tall and Vinner's (1981) *concept definitions*. When students are supplied with mathematical definitions, it is frequently with purposes in mind that may not yet be easily communicated to them. A mathematics lecturer may say, "Suppose we define . . ." and then go on to make a definition that may seem very odd until its consequences are explored. Definitions may be set up to avoid problems of which the student thus far has no experience or to create a distinction between the object being studied and something else that they have not yet learned about. Alcock and Simpson (2002) described "working from definitions" (p. 32) as an aspect of studying university mathematics that contrasts sharply with many students' typical experiences in school.

THE STUDY

I carried out a small-scale exploratory research study in order to examine these issues further. In particular, the questions I sought to answer were the following:

1. To what extent do mathematics teachers see the description of confidence intervals as potentially problematic?
2. Do they regard particular formulations as revealing or contributing to misconceptions?
3. What particular forms of words do they regard as correct or incorrect?

The research participants were a convenience sample of 12 high school and college teachers of mathematics that I knew personally through mathematics teacher conferences. I sent each person a brief questionnaire by e-mail. Although this by no means constitutes a random sample of mathematics teachers, it did ensure that I obtained a 100% return rate, and the teachers included covered a range of different backgrounds and experiences. However, the small sample size means that any conclusions drawn must be tentative and can be indicative only.

The questionnaire consisted of four statements about a 95% confidence interval and asked the participants which of these they thought were correct or incorrect and whether they had any alternative formulations or comments about misconceptions or any difficulties associated with teaching confidence intervals. The four statements were as follows:

A About 95% of the time the true population mean lies inside the confidence interval.
B I'm 95% sure that the confidence interval contains the true population mean.
C The probability that the true population mean is within the confidence interval is 95%.
D There is a 95% chance that the true population mean is inside the confidence interval.

These statements were designed to draw out from the participants their views about the correctness or incorrectness of particular aspects of their formulation. Statement A takes an objective, long-term average, frequency perspective, whereas statement B presents the matter in a more subjective way, with a personal pronoun and the notion of being 95% sure. Statements B–D focus on one particular confidence interval, whereas statement A appears to suggest the existence of a (large enough) set of them. Statement C explicitly invokes the probability of the true population mean lying within a certain range—something that is generally seen as problematic within the literature. The purpose of this statement was to see the participants' reactions to a classic instance of what is widely regarded as a misconception. Statement D was intended to be a less familiar wording (a concealed version) of something very similar in content to statement C. Would more teachers reject C, perhaps due to its familiarity as a commonly presented misconception, than the less easily recognized D?

No participant saw any other participant's responses. Two participants followed up their initial e-mail response 2 days later (Respondent 1) and 3 days later (Respondent 4) with further thoughts, which were also included in the analysis. The extent to which respondents discussed the matter with colleagues outside this study, or consulted textbooks or other resources, is unknown. However, I would suggest that the likelihood is minimal for a busy teacher on a working school day responding within a few hours to an informal e-mail enquiry.

All respondents were aware of difficulties associated with how students expressed the meaning of confidence intervals and seemed to regard certain formulations, especially statement B, as representing a common misconception. They reported treading carefully and warning their students exactly how to describe confidence intervals; particularly what not to say, often with reference to examiners' expectations. For example, Respondent 1 commented that "The examiner does ask for the meaning of a confidence interval as a standard question in S3 [a 'Further Mathematics' unit

in the UK A-level Further Mathematics qualification] and he definitely doesn't like answers like B, C or D." Respondent 2 gave a typical answer, also concluding with what examiners require:

> I think that the issue is that there is nothing special about any particular confidence interval. It is based on one particular sample and another sample would produce a different confidence interval. You would expect that 95% of intervals constructed in this way will contain the mean. I think that is the sort of statement that [is] expected in exams.

In a high-stakes examination culture, what examiners want or expect is frequently taken as final.

Although all of the participants contacted were experienced teachers of mathematics, three hedged their responses to some degree. For example, Respondent 2 bracketed his comments with "It's been some time since I taught this topic but" and "but as I say I am rather out of touch!" Respondent 3 did something similar, with

> I have only in recent years studied and certainly only ever taught hypothesis testing in the S2 [the second statistics unit in the A-level course in the UK] sense. So it could be that below is referring to something that I have only possibly never looked into properly (I'd need to go to the exact definition etc.) . . . But I think you are asking something different which is beyond S2?

The other nine respondents seemed to offer more confident answers, although there was a general sense that my question was a difficult one in a complex area.

Two respondents did not see important differences between the statements. Respondent 1 commented, "I think that A is closest to what I teach. I usually say '95% of confidence intervals constructed in this way will contain the mean,'" not seeming to ascribe importance to any differences between their formulation and statement A. This was followed with a conventional explanation:

> Remember that the mean is fixed and each time we take a sample and work out an interval, we get a different interval. On 95% of those occasions we will have successfully trapped the mean inside our interval.

Two days later, the same participant sent a follow-up e-mail:

> Thinking about this again, perhaps B is acceptable too. What we need to remember is that the mean is fixed and has no probability attached to it. The probability is attached to the interval so the wording of B is acceptable but not C or D. (Respondent 1, 2 days later)

Respondent 4 also regarded some of the differences as unimportant (e.g., "C and D are just semantics"). Initially this respondent was uncertain about statement B but sent a second e-mail, 3 days later, with a different conclusion; almost exactly opposite to the point of view of Respondent 1:

> Doesn't it depend if you look at it from a frequentist or Bayesian [point] of view? Either way, clearly B is subjective, so not really provable (thus not true, I guess. It depends what info you knew at the start.) (Respondent 4)

> B is true—it is how you define a [confidence interval] but C and D are just semantics, so I think fine as well. I think C and D say the same thing as B really. A is too imprecise to be correct or incorrect. (Respondent 4, 3 days later)

These changing opinions seemed to indicate uncertainty over what was correct, even among these experienced, highly competent classroom practitioners.

Half of the respondents found statement A acceptable and statement B was rejected by three-quarters, yet Richardson and Haller (2003), for instance, seemed to accept phrases such as "90% confident" while adhering to the standard line, commenting that:

> if we claim that we are 90% confident that a proportion lies within the endpoints of a confidence interval, we are saying that the endpoints of the confidence interval were calculated by a *method that gives correct results in 90% of all possible samples*. We cannot say that the probability is 90% that the true proportion falls within the endpoints of the confidence interval. No randomness remains after we draw one particular sample and construct from it one particular interval. The true proportion either is or is not between the confidence interval endpoints. (p. 8)

One respondent accepted D but not C on the grounds that "You can't talk about the probability of the population mean being something, because it is fixed," and one rejected all four, substituting his own statement involving references to taking lots of samples in which 95% of them contained (or caught) the population mean. I obtained the impression that some of the teachers navigated this area by having one and only one way of expressing the matter, which they had thought about and were sure (or had been told) was correct, and were uncertain about the accuracy of *any* alternative. They therefore did not wish to accept any statement different from the one that they were most used to and comfortable with. For example:

> C is wrong. I might accept A but I don't like any of the others. Since Mu [the population mean] is fixed and M [the sample mean] is a random variable these statements should be about a random measurement M and not a fixed value Mu. I would accept: "The probability that the random interval (calculated from a sample value M) contains the value of Mu is 0.95" but not: "The probability that Mu lies inside the . . . interval is 0.95." (Respondent 5)

It would seem from this exploratory study that mathematics teachers do recognize the potentially problematic nature of the way in which confidence intervals are described and may be led by habit or by the publicized expectations of examiners to prefer particular forms of wordings—and consequently to pass them on to their students. However, at least among these research participants, there was little consensus regarding which of my statements were acceptable or unacceptable. A rough summary of their overall responses is given in Table 1.

TABLE 1
Respondents' Conclusions ($n = 12$)

Statement	Considered acceptable	Considered unacceptable	Ambivalent or no opinion clearly stated
A	6	1	5
B	0	9	3
C	2	4	6
D	2	3	7

DISCUSSION

The two statements:

- "The population mean lies within the confidence interval" and
- "The confidence interval contains the population mean"

would seem at face value to be linguistically equivalent, just as "The milk is in the fridge" and "The fridge contains the milk" are. To say that the first implies that the population mean is changing whereas the second implies that the confidence interval is changing seems unwarranted. Would we assume that the milk is mobile in the first statement but that the fridge is in the second? However, it must be admitted that "The fridge contains the milk" does sound unusual and might well cause us to question why the statement was being made that way round. Mathematical language needs to be both logically sound and conventionally acceptable.

Nonetheless, in the case where we take just one sample, and so obtain just one confidence interval, it would appear that we simply have two quantities—an *unknown* population mean and a *known* interval. Whether the population mean lies within the confidence interval or not depends on the *relative* location of the two things, not simply on one or the other. Treating the population mean as though it could take various values (an unknown variable) is not necessarily absurd; indeed, when finding maximum likelihood estimators, for instance, the population mean is tacitly treated as a variable, to the extent that differentiation is performed with respect to it. It might also be felt from an epistemological perspective that the notion of "fixed but unknown" is potentially problematic, because on what basis can we know that the population mean is fixed if it is completely unknown?

If saying that the population mean lies in the confidence interval is exactly equivalent to saying that the confidence interval contains the population mean, then we can label this event as X and recast the four statements A–D above as A'–D' below:

A' About 95% of the time X.

B' I'm 95% sure that X.

C' The probability that X is 95%.

D' There is a 95% chance that X.

When written in this way, they might indeed all appear to be conceptually equivalent, if "95% of the time," "95% sure," "probability is 95%," and "95% chance" all mean the same as "95% confident." Nevertheless, that does not mean that they necessarily have identical effects on the reader. Moreover, A' is a frequency statement, whereas B' is a more subjective one, which might not be regarded as the same by everyone, as I discuss below.

CONCLUSION

Many teachers would regard it as important for students to recognize that the population mean is fixed but unknown, whereas the limits of the confidence interval depend on the nature of the particular sample taken. Cumming (2007) defended cumbersome phrasing, commenting that

"The wording used . . . is meant to remind us that it's the intervals that vary, and any probability refers to what happens in the long run in the set of all intervals" (p. 91). Typically, statistics texts are careful with their language; for example, Garner (2010) explained as follows:

> We identify a confidence level for this interval, usually 95% or 99%, meaning that, in the large number of samples we might use to construct an interval, the true population value will fall into the interval 95% (or 99%) of the time. The procedure, if it were repeated over and over again, would 'catch' the population value in 95% (or 99%) of the attempts. (p. 130)

However, less precise (arguably more intuitive) explanations are also common, such as this from Woodbury (2002):

> The confidence interval tells us that we are 95% sure that the true mean age for all college statistics students is in between 21.67 and 23.73 years old. There is a 5% chance that the mean is not in this interval . . . (p. 313)

This is reminiscent of Fidler's (2006) students, who expressed being "95% confident that the population mean would lie in this interval" (p. 206). Similarly, Cohen, Manion, and Morrison (2011) wrote that:

> The confidence level, usually expressed as a percentage (usually 95 per cent or 99 per cent), is an index of how sure we can be . . . that the responses lie within a given variation range. The confidence interval is that degree of variation or variation range . . . that one wishes to ensure. (p. 145)

It is questionable whether these more user-friendly formulations do real damage to students' understandings or indeed say anything inaccurate about the situation. Suppose that a fair coin is thrown, caught, and covered before anyone sees. Whether it is heads or tails is fixed and cannot be changed, yet because we lack knowledge we say that the probability of heads is 0.5, because we know that 50% of the time it will be heads. It is unclear how this is any different from saying that we do not know whether the population mean lies within our confidence interval, but we know that in the long run it will 95% of the time, so we say that it does with probability 95%. Saying that the population mean must be either inside or outside the interval (it is all or nothing) is no different from saying that one throw of a coin must be either completely heads or completely tails. This idea of a fixed but unknown situation has some resonances with the paradox of Schrödinger's cat (Gribbin, 2012).

Precision may be a good course for the teacher, but expecting students to parrot their formulations is likely to be counterproductive. According to Sfard (1992), "For an abstract object to be born, a long period of incubation may sometimes be necessary" (p. 83). She shows that trying to force a structural perspective on students can lead to the creation of superficial *pseudo-objects*, which are not well supported by understanding. If students have had the opportunity to build up the imagery and network of ideas expressed by well-thought-out phrasing, through engaging with appropriate activities, then it does make sense for them to work on expressing such ideas accurately and in ways that capture essential aspects of them. However, forcing students into formal articulations of confidence intervals that they do not own might fail to influence any deeper understanding. It might be good pedagogic practice for the teacher to emphasize active verbs, such as in saying that "the confidence interval *captures* the population mean," to try to indicate what is going on in a dynamic manner, but it would be inadvisable to create a shibboleth for students over particular forms of words.

What students say and what students think are not simply related. Edwards and Ward (2004) commented that:

It is not uncommon for students (or any person for that matter) to repeat something they have heard without full understanding. For instance, students may say, perhaps to please instructors, that "mathematics is necessary in all walks of life," without being able to give one meaningful example beyond the day-to-day interactions involved in commerce. (p. 414)

If students give explanations mechanistically, this is unlikely to help them in the process of developing their nonverbal concept images. If a student is simply told "*never* say the probability is 0.95 that the mean lies in a 95% confidence interval for the mean" (Dillon, 2011, p. 34), they may assent without modifying their concept image and thus continue to think of the confidence interval in exactly the same way. The consequence of a fascist "word police" agenda is that confidence intervals become a minefield through which students (and sometimes teachers) step with *lack of confidence*. Smith et al. (1993) stressed the need to "conceptualize learning in terms of refinement rather than replacement" (p. 150) of students' preconceptions, commenting that "it is the nature of mathematical and scientific knowledge that the most elegant and valued expression consists of very general, compact, and abstract propositions" (p. 150), and we must not assume that this can be achieved quickly by brute force. It is likely to be more effective to work willingly with students' current conceptions as they are, rather than to attempt to impose something that is quite alien to them.

For some educators, the only solution is to teach statistics from a Bayesian perspective, even at an elementary level (Berry, 1997). Jackman (2009), for example, located the problem within the frequentist paradigm:

One often heard interpretation of the 95% confidence interval ... is "there is a .95 probability that μ lies between \$40 000 and \$50 000." ... From the frequentist perspective, the statement ... is valid, since for frequentists "probability" is at least tacitly understood to mean "relative frequency in repeated sampling." But this is not how most practitioners use confidence intervals. Rather, subjective statements of the sort "I am 95% sure that $\mu \in$ [\$40 000, \$50 000]" are quite typical. ... Alas, the correct frequentist interpretation is the less helpful statement about the performance of the 95% confidence interval in repeated sampling. This leads to considerable confusion, and frankly, makes teaching and learning statistics harder than it should be. (p. xxxiii)

Bayarri and Berger (2004) took the view that "one can teach much of elementary statistics from [the Bayesian] perspective, without changing the procedures that are taught" (p. 63). Howell (2013) confessed that he has "worried and fussed" (p. 193) over how to present the meaning of a 95% confidence interval, saying that "in some ways the argument is very petty and looks like we are needlessly splitting hairs" (p. 194). In the end, however, he too sided "for teaching purposes" with the Bayesians, since "Going with the traditionalists requires such convoluted sentences that you look like you are trying to confuse rather than clarify" (p. 194). Some authors, such as Masson and Loftus (2003) seem to be attempting to combine to the two perspectives by saying, for example, that:

there is a 95% probability that the interval is one of the 95% of all possible confidence intervals that includes the population mean. Put more simply, in the absence of any other information, there is a 95% probability that the obtained confidence interval includes the population mean. (p. 204)

It is clear that various approaches are possible, and teachers of statistics might like to reflect on their own concept images for the notion of a confidence interval and to consider ways in which they might choose to word the definition for students. Students could be provided with opportunities to consider different mathematical definitions and to explore how different wordings might highlight different, perhaps unintended, properties or features. In the end, whatever view teachers take, definitional ambiguity can be seen as an opportunity for clarifying thinking through discussion (Foster, 2011) rather than for imposing an inflexible mantra and declaring all else heretical.

ACKNOWLEDGMENT

REFERENCES

Alcock, L., & Simpson, A. P. (2002). Definitions: Dealing with categories mathematically. *For the Learning of Mathematics*, *22*(2), 28–34.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, *19*(1), 58–80.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, *10*(4), 389–396.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *American Statistician*, *51*(3), 241–246.

Bertie, A., & Farrington, P. (2003). Teaching confidence intervals with java applets. *Teaching Statistics*, *25*(3), 70–74.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. Oxford, England: Routledge.

Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, *29*(3), 89–93.

Dillon, M. (2011). Aftermath: Statistics à la mode. *Math Horizons*, *19*(2), 34.

Edwards, B. S., & Ward, M. B. (2004). Surprises from mathematics education research: Student (mis)use of mathematical definitions. *The American Mathematical Monthly*, *111*(5), 411–424.

Fidler, F. (2006). From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology (PhD Thesis, University of Melbourne). Retrieved from http://www.botany.unimelb.edu.au/envisci/docs/fidler/fidlerphd_aug06.pdf

Foster, C. (2011). Productive ambiguity in the learning of mathematics. *For the Learning of Mathematics*, *31*(2), 3–7.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than *p* values: Estimation rather than hypothesis testing. *British Medical Journal (Clinical Research Ed.)*, *292*(6522), 746–750.

Garner, R. (2010). *The joy of stats: A short guide to introductory statistics in the social sciences*. Toronto, ON, Canada: University of Toronto Press Incorporated.

Gribbin, J. (2012). *In search of Schrödinger's cat*. London, England: Black Swan.

Hagtvedt, R., Jones, G. T., & Jones, K. (2008). Teaching confidence intervals using simulation. *Teaching Statistics*, *30*(2), 53–56.

Hewitt, D. (2001). Arbitrary and necessary: A way of viewing the mathematics curriculum. In L. Haggarty (Ed.), *Teaching mathematics in secondary schools: A reader* (pp. 47–63). London, England: RoutledgeFalmer.

Howell, D. C. (2013). *Statistical methods for psychology* (8th ed.). Scarborough, Canada: Wadsworth Cengage Learning.

Jackman, S. (2009). *Bayesian analysis for the social sciences*. Chichester, England: John Wiley & Sons.

Kalinowski, P. (2010). Identifying misconceptions about confidence intervals. In C. Reading (Ed.), *ICOTS-8 Proceedings: Towards an evidence based society*. Voorburg, The Netherlands: International Association for Statistical

Education, International Statistics Institute. Retrieved from http://icots.net/8/cd/pdfs/contributed/ICOTS8_C104_KALINOWSKI.pdf

Landau, S. I. (2001). *Dictionaries: The art and craft of lexicography* (2nd ed.). Cambridge, England: Cambridge University Press.

Masson, M. E., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, *57*(3), 203–220.

Morgan, C. (2005). Words, definitions and concepts in discourses of mathematics, teaching and learning. *Language and Education*, *19*(2), 102–111.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, *82*(4), 591–605.

Poole, C. (1987). Beyond the confidence interval. *American Journal of Public Health*, *77*(2), 195–199.

Richardson, M., & Haller, S. (2003). Confident in a kiss? *Teaching Statistics*, *25*(1), 6–11.

Robinson, R. (1954). *Definition*. London, England: Oxford University Press.

Robison-Cox, J. F. (1999). Having a ball with confidence intervals. *Teaching Statistics*, *21*(3), 81–83.

Sfard, A. (1992). Operational origins of mathematical objects and the quandary of reification—the case of function. In G. Harel & E. Dubinsky (Eds.), *The concept of function: Aspects of epistemology and pedagogy* (Vol. 25, pp. 59–84). Washington, DC: Mathematical Association of America.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, *77*, 20–26.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115–163.

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.

Vinner, S. (1991). The role of definitions in the teaching and learning of mathematics. In D. Tall (Ed.), *Advanced mathematical thinking: Mathematics education library* (pp. 65–81). Dordrecht, the Netherlands: Springer.

Woodbury, G. (2002). *An introduction to statistics*. Pacific Grove, Canada: Wadsworth Group.