COLLEGE INSTRUCTORS' ATTITUDES TOWARD CONFIDENCE INTERVALS

Hyung Won Kim[1], Aaron Thomas Wilson[1] and Ho Kyoung Ko[2]
[1]School of Mathematical and Statistical Sciences, University of Texas Rio Grande Valley, USA
[2]Graduate School of Education, Ajou University, Suwon, South Korea
hyung.kim@utrgv.edu

This study explores college statistics instructors' attitudes toward teaching confidence intervals (CIs) in elementary statistics courses and using them for inferential statistics. An instrument was developed using a taxonomical construct that classifies attitudes into the three pedagogical components: affective, cognitive and behavioral. Data were collected from 270 college instructors, who were grouped by gender, academic background and statistics teaching experience. The study analyzes the data across these groupings to explore attitudes toward CIs themselves, CIs in terms of the three pedagogical components, and CIs compared between general and teaching contexts. The results indicate that the three groupings are moderate to strong predictors of attitudes toward CIs, and that group differences were more evident in behavioral component than the other two.

BACKGROUND

In the research community of the social sciences, the process of null hypothesis significance testing (NHST) has been the overwhelming choice for quantitative data-analysis procedures for many years (Fidler & Loftus, 2009). Thus, the *p*-value has been widely used in quantitative social science research. However, an increasing number of scholars in statistics education depreciate the *p*-value as an inadequate measure of evidence for hypothesis testing in inferential statistics (Hubbard & Lindsay, 2008; Marden, 2000). Recently, many statistics reformers have studied the adequacy of confidence intervals (CIs) in inferential statistics (e.g., Beyth-Marom, Fidler, & Cumming, 2008). In the pedagogical context, Schmidt and Hunter (1997), for example, claimed that students might have greater success learning how to use CIs than learning how to conduct NHST, and, further, that CIs enable researchers to make better inferential decisions than the *p*-values used in NHST. In general, reformers of statistics practices recommend the use of CIs in place of the typically used *p*-values for the following advantages: (1) Information in the form of point and interval estimates using measurement units is easily accessible and comprehensible; (2) while remaining within a familiar Frequentist framework, CIs offer more information than *p*-values; (3) CIs help combine evidence from multiple experiments by promoting meta-analytic thinking; and (4) CIs help make sense of statistical non-significance (Cumming & Finch, 2001, 2005; Kalinowski, 2010). These advantages have led the reformers to call for wider usage of CIs in preference to *p*-values (Cumming & Finch, 2001). As a result, interval estimates are gaining more attention and becoming more popular in the quantitative research community (Cumming & Finch, 2005), leading to a need for more studies on the usage and teaching of CIs.

The literature on this topic, although scant, indicates that a number of statistics educators are actively exploring cognitive factors in CI knowledge attainment – in particular, misconceptions of CIs commonly developed by statistics students (e.g., Hoekstra, Morey, & Rouder 2014; Kalinowski, 2010). However, although the role of non-cognitive factors of teachers in improving the cognitive performance of students has been investigated and stressed in the statistics education community (e.g., Kim, Wang, Lee, & Castillo, 2017; Martins, Nascimento & Estrada, 2012), no attention has been paid thus far to the non-cognitive factors of elementary statistics instructors (ESIs) in teaching CIs in their courses and in using CIs to do inferential statistics. Aligning with the previous research that has demonstrated the significance of non-cognitive factors of teachers in students' learning, we aim to explore ESIs' attitudes toward CIs in the context of teaching elementary statistics as well as in the context of using CIs for inferential statistics. To this end, we draw on a taxonomical construct to classify attitude into three pedagogical components – affective, cognitive and behavioral – which is widely accepted in studies of non-cognitive factors in psychology and education (e.g., Estrada, 2002; Martins et al., 2012). The study in this paper uses Martins et al.'s (2012) description of the components: "affective" addresses feelings about the objects in question, "cognitive" addresses self-perception or beliefs about the objects in question and "behavioral" addresses inclinations to act in a particular way about the objects in question. Our

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

main research question is: What attitudes do ESIs have toward teaching CIs in elementary statistics courses and using them for inferential statistics? Specifically, we ask: (1) How do ESIs' attitudes compare across the three components of attitude in the domain of CIs? (That is, how do ESIs feel about (affective), perceive (cognitive) and behave with respect to using CIs for inferential statistics and teaching them in elementary statistics?), and (2) How do ESIs' attitudes toward CIs compare between the general context of using CIs and the context of teaching CIs in elementary statistics?

METHOD

To measure ESIs' attitudes toward CIs, the research team developed an instrument, which is divided into three parts. Part I collected biodata of the participants. Part II (12 items) addresses ESIs' attitudes toward using and teaching CIs with no explicit connections with the aforementioned advantages of CIs. Part III (18 items) addresses ESIs' attitudes toward CIs in relation to such advantages. This paper uses participant responses to the items in Parts I and II only. The items in Part II of the questionnaire are grouped by the three pedagogical components, and address two contexts: general and teaching (Table 1); for example, "CIs provide less evidence for making statistical inferences than the *p*-value methods" (general) and "In introductory statistics classes, I emphasize CIs as a tool for inferential statistics as much as *p*-value methods" (teaching). An online survey was sent to 10,096 faculty in math or statistics departments at two or four year colleges across the US during the summer of 2017. While 292 responded, the analysis was conducted using the responses of 270 participants who answered more than 80% of the 12 items in Part II of the survey. Based on these responses the instrument demonstrated acceptable internal consistency, with a Cronbach's alpha of 0.751.

The 12 items in Part II had a 7-point Likert scale format. To avoid apparent acquiescence, seven items were phrased positively and five were phrased negatively (as in the first example item above); the response scores of the latter were reversed for the analysis. Two sets of analyses were conducted. The first set compared the survey responses to examine the ESIs' attitudes according to the three pedagogical components (affective, cognitive and behavioral). The second set compared their attitudes in the two contexts (general and teaching). In each set, the survey responses were compared across the participants by grouping them in three ways based on their biodata: (1) by gender (male or female); (2) by academic background (Math: having a graduate degree with the highest degree in math without holding a graduate degree in statistics; Stat: having a graduate degree in statistics; and MaEd: having a graduate degree with the highest degree in math education without holding a graduate degree in statistics); and (3) by statistics teaching experience (novice: 3 years or less; low-Exp: 4–9 years; mid-Exp: 10–15 years; and high-Exp: 16 years or more). In the three sets of analyses, the gender group comparison used a one-sided hypothesis, "the score of one group is higher than the other's." Therefore, these comparisons were done using one-sided *t*-tests. Both the academic background and the statistics teaching experience group differences were analyzed via Welch ANOVAs, followed by post hoc multiple comparisons to identify the specific groups that were the sources of differences. The homogeneity of variance was not assumed in any of the tests.

Table 1. *Survey Items Framed by the Taxonomical Constructs*

	Affective (A)		Cognitive (C)		Behavioral (B)	
General	9	10	5	6	12	7
Teaching	1	3	2	4	11	8

RESULTS

Significance tests were conducted at α = .05, .01 and .001. Due to the limitations involved in making dichotomous conclusions based on *p*-values, we further provide 95% 2-sided confidence intervals for the *t*-tests. Throughout the analysis, we use the acronyms GN, AB and STE for gender, academic background and statistics teaching experience, respectively. Each individual's score could range from 1 to 7, as with the means. The results are summarized in Table 2. Column 1 of Table 2 shows the means and standard deviations of the attitude scores. While the difference associated with GN is significant (p = .039) at α = .05, the *p*-values of AB (.001) and STE (.000) indicate the attitude differences associated with AB and STE are significant at α = .01 and α = .001,

Table 2. *Mean (SD) of responses across the three components and contexts*

	1. All	2. Affective	3. Cognitive	4. Behavioral	5. General	6. Teaching
95% CI	(5.081, 5.266)	(4.849, 5.051)	(5.426, 5.649)	(4.909, 5.161)	(4.942, 5.150)	(5.203, 5.401)
Mean (SD)	5.173 (.770)	4.950 (.843)	5.538 (.929)	5.034 (1.049)	5.046 (.868)	5.302 (.827)
GN, *n*=270	*p* =.039	*p* =0.040	*p* =0.703	*p* = 0.015	*p* =.040	*p* =.086
Male, *n*=172	5.247(.058)	5.032 (.800)	5.553 (1.010)	5.154 (1.001)	5.128 (.874)	5.367 (.822)
Female, *n*=98	5.045(.077)	4.806 (.901)	5.511 (.771)	4.826 (1.091)	4.904 (.842)	5.187 (.827)
AB, *n*=268	*p* =.001	*p* =.106	*p* =.067	*p* =.000	*p* =.001	*p* =.009
Math, *n*=179	5.108^{**}(.778)	4.947 (.839)	5.488 (.939)	4.877^{***} (1.023)	4.964^{**} (.859)	5.244^{**} (.833)
Stat, *n*=55	5.518^{ref} (.712)	5.132 (.725)	5.805 (.905)	5.636^{ref} (.972)	5.443^{ref} (.873)	5.609^{ref} (.715)
MaEd, *n*=34	5.040^{**}(.770)	4.750 (.961)	5.466 (.720)	4.919^{**} (.982)	4.911^{**} (.710)	5.172^{*} (.853)
STE, *n*=270	*p* =.000	*p* =.002	*p* =.069	*p* =.000	*p* =.001	*p* =.001
Novice, *n*=66	4.801^{ref}(.821)	4.607^{ref} (.873)	5.274 (1.028)	4.530^{ref} (1.029)	4.665^{ref} (.920)	4.936^{ref} (.817)
Low, *n*=74	5.292^{**}(.609)	5.026^{*} (.747)	5.640 (.819)	5.210^{***} (.779)	5.138^{**} (.703)	5.449^{**} (.742)
Mid, *n*=61	5.298^{**}(.714)	5.108^{**} (.748)	5.619 (.827)	5.175^{**} (1.103)	5.209^{**} (.814)	5.390^{*} (.784)
High, *n*=69	5.292^{**}(.824)	5.058^{*} (.915)	5.609 (.996)	5.207^{**} (1.134)	5.169^{*} (.931)	5.416^{**} (.872)

* $p < .05$; ** $p < .01$; *** $p < .001$. Superscript “ref” marks the reference group in each column.

respectively. Regarding AB, those with a graduate level statistics background had more positively developed attitudes toward CIs than those without a statistics background at the graduate level. Regarding STE, while the attitude difference between the novice group and the three experienced groups of ESIs is clear, no clear differences were detected among the three experienced groups.

Columns 2 to 4 of Table 2 compare the three pedagogical components of attitude grouped by GN, AB and STE. The GN results are significant, at α = .05, for the affective and behavioral components. The AB results are significant for the behavioral component at α = .001, but do not show significance for the other two components at α = .05. Two *p*-values for STE indicate significant differences (affective at α = .01 and behavioral at α = .001). Regarding STE, the scores of the novice group are significantly lower than those of the three experienced groups for both affective and behavioral components. It is important to note that the behavioral component shows significant differences (at different levels of α) for the three groupings based on GN, AB and STE.

Regarding the general and teaching contexts, Columns 5 and 6 compare the ESIs’ attitudes toward CIs in the two different contexts (general and teaching). Male instructors had significantly more positive attitudes toward CIs both in the teaching context (*p* = .086; significant at α = .10) and in the general context (*p* = .040; significant at α = .05). The group differences in AB are significant in the general context (*p* = .001; significant at α = .01) and in the teaching context (*p* = .009: significant at α = .01). Regarding STE, ESIs with teaching experience of four or more years have significantly more positive attitudes toward CIs than their novice counterparts, both in the general context (*p* = .001; significant at α = .01) and in the teaching context (*p* = .001; significant at α = .01). Thus, while there is no clear difference among the STE groups’ attitudes between the two contexts, differences in attitudes are more significant in the general context than in the teaching context for both GN (*p* = .040 < *p* = .086) and AB (*p* = .001 < *p* = .009) groups.

CONCLUSIONS AND DISCUSSION

The findings suggest that while GN is a moderate predictor of ESIs’ attitudes toward CIs (*p* = .039), both STE and AB are strong predictors of ESIs’ attitudes toward CIs (*p* = .001 and .000, respectively). In particular, the post hoc analysis for AB found that the ESIs with a graduate level statistics background hold significantly more positive attitudes toward CIs than those without such a background, and that for STE, the ESIs with four or more years of experience have more positive attitudes toward CIs than their novice counterparts. The group differences showed different patterns based on the taxonomical construct for which the comparison was conducted. For instance, among the three pedagogical components, the group differences were clear for the behavioral component (GN: *p* = .015; AB: *p* = .000; STE: *p* = .000) but less clear for the affective and cognitive components. The findings imply that ESIs with less statistics background or statistics teaching experience may have difficulty emphasizing CIs in their teaching and may not have developed an understanding of how to use CIs for inferential statistics. For the comparison of the two contexts (general and teaching), the difference was more drastic in the general context (p = .040, .001 and .001) than in the teaching context (p = .086, .009 and .001). The lower scores in the general context further indicate that ESIs' attitudes toward teaching CIs in their elementary statistics courses are more positive than their attitudes toward using it for inferential statistics.

The findings have implications regarding the culture of statistics learning at the college level. Many scholars in social sciences emphasize the importance of confidence intervals, in terms of teaching them in elementary statistics as an alternative statistical method for inferential statistics and conducting actual quantitative research. If such recommendations to emphasize CIs are to be successfully implemented in both the teaching community and the research community, it is critical to understand how ESIs perceive them. This study's exploration of ESIs' attitudes toward CIs is a first step in this direction. It is hoped that the study lays a foundation for future studies to explore the origins and factors of ESIs' attitudes toward CIs, the dynamics of how ESIs' attitudes affect their teaching and pedagogical decision-making, and how the challenges in teaching CIs relate to elementary statistics students' attitudes.

This study has limitations. Although the sample size (n = 270) allowed for quantitative data analysis for statistical inferences, a larger scale study and factor analysis of latent variable structure are necessary to arrive at firmer conclusions. Further, the data collection relied on voluntary responses from instructors of mathematics and statistics, and the response rate was low (2.67%), constraining the extent to which the findings can be generalized.

REFERENCES

Beyth-Marom, R., Fidler, F., & Cumming, G. (2008). Statistical cognition: Towards evidence-based practice in statistics and statistics education. *Statistics Education Research Journal, 7*(2), 20–39.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals based on central and noncentral distributions. *Educational and Psychological Measurement, 61*, 530–572.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist, 60*, 170–180.

Estrada, A. (2002). *Análisis de las actitudes y conocimientos estadísticos elementales en la formación del profesorado* [Analysis of attitudes and elementary statistical knowledge in training teachers] (Unpublished doctoral dissertation). Universidad Autónoma de Barcelona.

Fidler, F., & Loftus, G. R. (2009). Why figures with error bars should replace *p* values: Some conceptual arguments and empirical demonstrations. *Zeitschrift für Psychologie/Journal of Psychology, 217*(1), 27–37.

Gal, I. & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, *2*(2).

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, *21*(5), 1157-1164.

Hubbard, R., & Lindsay, R. M. (2008). Why *p* values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology, 18*, 69–88.

Kalinowski, P. (2010). Identifying misconceptions about confidence intervals. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics*. [CDROM]. IASE, Lijbljana, Slovenia, Refereed paper.

Kim, H., Wang, X., Lee, B., & Castillo, A. (2017). College instructors' attitudes toward statistics. In Anna Chronaki (Ed.), *Proceedings of the Ninth International Mathematics Education and Society Conference* (Vol. 2, pp. 611–621). Volos, Greece: University of Thessaly.

Marden, J. I. (2000). Hypothesis testing: From *p* values to Bayes factors. *Journal of the American Statistical Association*, *95*, 1316–1320.

Martins, J. A., Nascimento, M. M. S., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Education Research Journal, 11*(2), 26–44.

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests*? (pp. 37–64). Hillsdale, NJ: Erlbaum.