TEACHING STATISTICS IN CONTEXT: EFFECTS OF STATISTICS HISTORY ON STUDENT LEARNING

Catalina Suarez-Rivera and Dean Langan (joint first authorship)
UCL Great Ormond Street Institute of Child Health, University College London, United Kingdom
catalina.rivera@ucl.ac.uk

Students often learn statistical techniques and theory without historical context. Would integrating history support student experiences and lead to improved learning? An online randomized cross-over experiment with 135 participants examined (a) comprehension of statistical concepts (Fisher's exact test and Pearson's correlation) after watching short videos with and without historical context, and (b) overall attitudes towards including the history of statistics in live/online lectures. Self-reported viewing habits and comprehension were comparable in participants who watched the videos with and without history. Despite this, participants generally had a positive attitude towards its inclusion. Likewise, at least three times as many participants agreed that videos with historical context were preferred, important, and worth the time.

INTRODUCTION

Nearly all techniques and theories in statistics were created by real people to solve real-world questions and problems. The Fisher's exact test was developed to assess a lady's tea tasting abilities; the *t*-distribution for small samples was born in the Guinness brewery; and 'regression to the mean' arose to explain that tall fathers tend to have short sons (Salsburg, 2001). Indeed, historical and social contexts in which statistical methods were developed are full of rich and colorful details (Bellhouse, 2005). In some cases, the darker side of this history reveals connections to imperialism, oppression, eugenics, and misogyny. For example, the development of non-parametric significance tests was directly linked to eugenics (Clayton, 2020).

Historical contexts serve as productive and real-world examples for the classroom, which in turn support student learning (Chick & Pierce, 2011). Moreover, history helps to increase motivation for learning, makes statistics more accessible, and explains its role in society (Fauvel, 1991). However, statistics students often learn methods without the historical references or context (Bellhouse, 2005). Research is needed to understand if the inclusion of historical and social contexts of statistics benefits students' experiences and attitudes. A meta-analysis found a positive but negligible effect of history of mathematics on students' attitudes (Bütüner, 2015). An experiment on mathematical history in primary children found a positive effect on learning (Karaduman, 2010). Does history applied to post-secondary statistics education increase student learning and generate positive attitudes towards statistics?

METHODS

We recruited 135 participants for an online study via university mailing lists to reach previous undergraduate and postgraduate health sciences students and delegates in short, applied statistics courses. Participants were randomized to watch one of two online videos about a statistical technique (with its history explained, or without it) and then assigned to watch the other video (as in a cross-over trial). The video with history began with a short (~2 minutes) explanation of the context in which the test was developed followed by the theory (~5 minutes). The Fisher's exact test video with history began with the story of 'the lady drinking tea,' whereas the Pearson's correlation coefficient video with history began with general information on Karl Pearson. After watching the video, participants answered questions about how much of the video they watched and their comprehension of the theory component. On the next page, participants were given the opportunity to watch the other version of the video they were not randomized to at first and answered multiple-choice and free text questions about viewing practices and attitudes towards historical context. The questionnaire was developed by the authors to measure learning and constructs of interest. Both the videos and the questionnaires can be accessed as supplementary material (tinyurl.com/ICOTS2022). Validity and reliability of the questionnaire were not assessed in this study. Seventy-three (54%) participants were allocated to watch videos (with and without history) on Fisher's exact test in the first phase of the experiment. The last 62 (46%) participants were allocated to watch a video on Pearson's correlation coefficient during the second phase.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

Data analysis involved multiple linear regression with/without history as the key predictor and a continuous comprehension score (0–100% range) as the outcome, which tested understanding of the video's statistical topic and weighted questions based on the amount of work involved. Moreover, Mann-Whitney *U* tests compared how much of the video the participants claimed to have watched (self-reported), and their attitudes towards having historical context included in lecture videos (measured on Likert scales). Finally, thematic analysis was conducted on free text responses (Maguire & Delahunt, 2017). Each response was coded regarding its overall sentiment (i.e., positive, neutral, or negative) and theme (e.g., reasons why statistical context helps or not).

RESULTS

Table 1 shows sample demographics of the 135 participants who completed the questionnaire, split by whether they watched the video with or without history. Demographic variables include self-reported competency (binary) of statistics in general (*) and of the topic of the video they were allocated to watch (†).

Table 1. Demographics of the sample (*n* (%)), split by topic of the video and condition

	Fisher's exact test		Correlation analysis	
	With history (n = 38)	Without history (n = 35)	With history (n = 24)	Without history (n = 38)
Undergraduate	2 (5.3)	0 (0.0)	3 (12.5)	6 (15.8)
Postgraduate	23 (34.2)	26 (74.3)	18 (75.0)	23 (60.5)
Staff	6 (15.8)	2 (5.7)	2 (8.3)	2 (5.3)
External (other)	7 (18.4)	7 (20.0)	1 (4.2)	7 (18.4)
Statistics competency*	16 (42.1)	17 (48.6)	14 (58.3)	12 (31.6)
Topic competency†	9 (23.7)	14 (40.0)	13 (54.2)	11 (28.9)

Self-Reported Viewing

Participants used a Likert scale (none/minimal/some/most/all) to report how much of the video they watched. Overall, 57/73 (78.1%) and 44/62 (71.0%) reported watching most/all the Fisher's exact test and correlation videos, respectively. More participants watched most/all the correlation video with history (81.6%) than they did without it (73.3%), but this was non-significant according to the Mann-Whitney *U* test (W = 456, p = 0.475). A similar observation was made for the Fisher's exact test video (75.0% vs. 68.4%, W = 577.5, p = 0.706).

Comprehension Scores

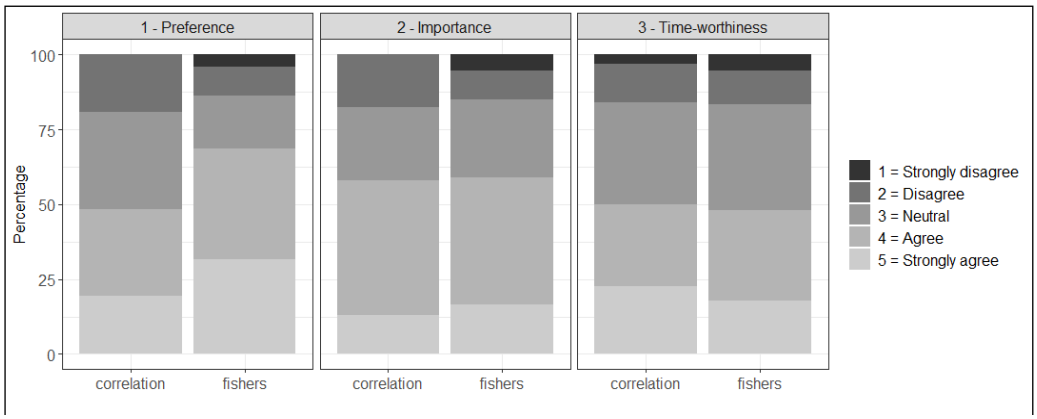
Participants were quizzed about their knowledge of the statistical method after watching the video through a short test, resulting in a continuous percentage score. Multiple linear regression models predicted test scores from predictor variables including whether participants watched the video with history, and self-reported competency with regards to statistics in general and the topic of the video. Results are presented in Table 2. Those watching either video with history did not score significantly different in the test. Those reporting they are statistically competent scored on average 9.96% higher on the test after watching the Fisher's exact test video (p = 0.020).

Table 2. Results from two multiple linear regression models of test scores, split by video topic

	Fisher's exact test (model 1)			Correlation analysis (model 2)		
Predictor	Est (B)	95% CI	*p*-value	Est (B)	95% CI	*p*-value
Intercept	68.44	-	-	48.39	-	-
With history	-2.96	[-9.53, 3.63]	0.374	-7.80	[-19.61, 4.01]	0.191
Statistics competency	9.96	[1.60, 18.33]	0.020	1.97	[-15.90, 11.97]	0.779
Topic competency	-3.35	[-12.42, 5.73]	0.465	-3.21	[-17.29, 10.88]	0.650

Attitudes Towards Historical Context

Attitudes towards historical contexts in terms of (a) preference, (b) importance, and (c) time-worthiness were measured on a Likert scale and are presented by video topic in Figure 1. Overall, more participants agreed than disagreed that historical context is preferred (59.2% vs. 16.3%, $p < 0.001$), important (58.5% vs. 16.3%, $p < 0.001$), and worth the extra study time (48.9% vs. 16.3%, $p < 0.001$). Notably, more participants watching Fisher's (68.5%) preferred history relative to participants watching the correlation video (48.4%) ($W = 2723$, $p = 0.0352$). No differences existed between video topics on importance and time-worthiness.



Note. The three statements (in full) are: (1) If this video was included in a statistics course, I would prefer the video with historical context. (2) In general, it is important to include the history of statistical concepts in an introductory statistics course. (3) Including historical context in statistics lectures is worth the extra study time that would be required.

Figure 1. Attitudes towards historical context by video topic

Free Text Comments

We received 63 (46.7%) text comments from participants offering some insight into the attitudes towards this topic that they felt were not captured elsewhere in the questionnaire. All percentages are calculated out of the 63 responses (unless specified otherwise). The sentiment towards providing historical context was largely positive; 40 (63.5%) had a positive sentiment, 19 (30.2%) were neutral, and four (6.3%) were negative. Common adjectives describing their attitudes towards historical context were 'interesting,' 'enjoyable,' 'useful,' and 'fun.'

The main overarching theme identified was the participants' understanding of why providing historical context helps learning. Nine (14.3%) comments suggested that it simply helps with the understanding of statistical methods without any detail. More specific themes suggested that historical context helps to memorize content (12, 19.0%) or concentrate by making the content more fun (5, 7.9%). Participants suggested that it helps with memorization "when methods are named after their inventor" and because "the brain works by associations (like a spider web) and if there is a story attached to a particular event/topic, [it] often helps [to] remember things." Relatedly, seven (11.1%) positive sentiments stemmed from participants' appreciation for history.

Another overarching theme referred to the specific historical content and how it could be integrated with statistical content. Indeed, 21 (33.3%) comments (the highest of all detected themes) suggested that context serves as a story that learners can relate to. Interestingly, 18/21 (85.7%) of the comments within this theme came from participants that had watched the video on the Fisher's exact test, which involved the 'lady drinking tea' story. A related theme suggested that it is not historical context that helps learning, but providing any context that learners can relate to, such as an example from their field of research, or something from current events. Six (9.5%) participants identified with this theme. Similarly, five (7.9%) emphasized that it is most important to have such an example integrated with the content of the lecture, rather than as a separate motivating example.

There is often a darker side to the history of statistical methods that was not present in the videos we created. Nonetheless, seven (11.1%) participants identified this issue, with a strong and

mixed sentiment. Six (85.7%) of these comments were in favor of providing these details within a lecture, with comments such as "It is important to acknowledge the relationship between statistics and eugenics, historical and ongoing oppression—particularly at UCL. Galton, Pearson, Fisher were all eugenicists, and this should be explicitly acknowledged in teaching modules." One negative comment towards this issue from a participant who identified as part of a minority group stated, "this upsetting reminder distracts me from the class and makes me not want to continue."

Other critical themes identified include the following: eight (12.7%) participants mentioned time and learning constraints in modules that may already include substantial content, with a risk of overwhelming the learners; and three (4.8%) stressed reasonably that historical content should not be assessed but included as supplementary content.

CONCLUSION

Statistics history did not improve participants' self-reported viewing or comprehension. Specifically, videos with and without history had similar self-reported viewing and comprehension scores. However, participants' attitudes with respect to preference, importance, and time-worthiness of historical contexts were largely positive; at least three times more participants agreed than disagreed with these statements. Text analysis revealed that 63% of responses were positive and 30.2% were neutral, with the most common theme suggesting that historical context provides a story that students can relate to. Although comparisons between video topics were not the main study goal, slight differences in results between the Fisher's test and correlation videos suggest that the specific context presented may play a role for student attitudes. It is possible that participants preferred Fisher's video because its historical context included an interesting story related to the test rather than the correlation video and its general information on Karl Pearson. In conclusion, although history of statistics may interest students during a lecture, this study shows little evidence to suggest that history supports learning.

Further research is needed to continue testing the effects of history on learning and attitudes towards statistics. This study used a non-validated questionnaire that was administered in an online setting, which led to unmeasured engagement and a very brief exposure. Future work should control the extent to which participants are familiar with the topic and how much they engaged with the video. Ideally, a lengthier test for learning could examine whether students learn more from videos with history. Likewise, a more prolonged exposure to history could lead to different results. Finally, we did not include information on the darker side of statistical history in our videos, but free-text analysis highlighted strong and mixed opinions on its inclusion. Thus, a careful investigation of the types of stories and ways to tell them is needed.

REFERENCES

Bellhouse, D. R. (2005). Probability and statistics ideas in the classroom—Lessons from history. In *Proceedings of the 55th Session of the International Statistical Institute*. International Statistical Institute. https://www.stat.auckland.ac.nz/~iase/publications/13/Bellhouse.pdf

Bütüner, S. O. (2015). Impact of using history of mathematics on students' mathematics attitude: A meta-analysis study. *European Journal of Science and Mathematics Education, 3*(4), 337–349.

Chick, H. L., & Pierce, R. (2012). Teaching for statistical literacy: Utilising affordances in real-world data. *International Journal of Science and Mathematics Education*, *10*(2), 339–362. https://doi.org/10.1007/s10763-011-9303-2

Clayton, A. (2020). *How eugenics shaped statistics*. Nautilus. https://nautil.us/how-eugenics-shaped-statistics-9365/

Fauvel, J. (1991). Using history in mathematics education. *For the Learning of Mathematics*, *11*(2), 3–6. http://www.jstor.org/stable/40248010

Karaduman, G. B. (2010). A sample study for classroom teachers addressing the importance of utilizing history of math in math education. *Procedia-Social and Behavioral Sciences, 2*(2), 2689–2693. https://doi.org/10.1016/j.sbspro.2010.03.397

Maguire, M., & Delahunt, B. (2017). Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland Journal of Higher Education, 9*(3), 3351–33514.

Salsburg, D. (2001). *The lady tasting tea: How statistics revolutionized science in the twentieth century*. Macmillan.