

The importance of supporting inferences with evidence: Learning lessons from Huffman (2014) in the hope of providing stronger evidence for extensive reading

Stuart McLean
Osaka Jogakuin University
Kansai University
Japan

A research paper does not have validity. A study's methodology—for example, the accuracy of its measurements—does not have validity. It is the inferences that researchers make that have or lack validity. Thus, researchers are responsible for presenting evidence that the inferences made are appropriate considering their research design and any limitations.

The strength of the inferences and conclusions that Huffman (2014) makes regarding the effectiveness of extensive reading (ER) relative to intensive reading (IR) at developing reading rate is excessive, and not valid considering the limitations of the methodology of his study. Huffman provides evidence of the effectiveness of ER in developing learners' reading speed. When measuring reading speed, Huffman reports the participants' self-reported time on task, provides evidence of participants' comprehension of timed reading passages, and utilizes the standard word unit (Carver, 1982): six letter spaces including punctuation and spacing. Huffman should be commended for these achievements in his publication. However, the validity of the inferences made within the paper is limited. Thus, the strength of Huffman's conclusions is excessive considering the limitations of the methodology of Huffman (2014).

Huffman (2014) refers to statistical analysis: the ER group statistically ($t(64) = 5.97, p = .000$) improved their reading rate relative to the intensive reading group, and the eta squared index indicated that 36% of the variance of the reading rate gain variable was accounted for by a student's membership in the ER group or in the IR group. Based on this, Huffman states that statistical analysis "unequivocally support[s]" (p. 27) the *a priori* hypothesis that "Reading rate gains will be significantly greater for students in a one-semester college extensive reading course than those in an intensive reading course" (p. 22), and concludes that "This study provides solid empirical data supporting the effectiveness of extensive reading over intensive reading for reading fluency development" (p. 28). Huffman concentrates on and bases inferences on the single independent variable of treatment type. However, limitations in the research design result in the presence of not one, but three independent variables. The weekly practice of timed reading by the ER group, but not by the IR group introduces a second independent variable. The third independent variable is time on task, which Huffman mentioned as a limitation of the paper. However, the investigation into the two treatment groups' different amounts of treatment time was limited, and this author feels that the explanation for the difference in time on task was incorrect.

A second independent variable: Only ER group participants conducted timed reading practice throughout the semester

Timed reading practice is not listed in the methodology as an activity conducted by the IR group by Huffman (2014). In contrast, Huffman states that the ER group participants “also did a series of six timed readings in addition to the pre- and post-course reading rate measures” (p. 25). Further, Huffman when explaining the large gain in reading speed relative to a previous study states that “students in this study engaged in timed reading activities during class time throughout the semester and were regularly encouraged to work on increasing their reading speed” (p. 28). However, the practice of timed reading by the ER group, but not by the IR group is not stated by Huffman as a limitation of the paper. This is problematic as previous studies (Bismoko & Nation, 1974; Chang & Millet, 2013; Chung & Nation, 2006; Cramer, 1975; Macalister, 2008, 2010; Yen, 2012) have provided evidence that timed reading practice alone increases a learner’s reading speed.

The practice of timed reading by the ER group, but not the IR group during the semester is further an issue because “A practice reading rate test consisting of one text was administered one week before the actual test in order to familiarize students with the procedure” (Huffman, 2014, p. 26), and “The posttest was administered in the same fashion, without the warm-up test” (p. 26). This is problematic because, with timed reading practice throughout the semester, and thus possibly only a week prior to the posttest, the ER group participants were to some degree prepared for the timed reading posttest. It can be argued that in contrast, the IR group, without timed reading practice since the beginning of the semester and without a posttest timed reading warm-up, was not “primed” for the posttest reading rate instrument to the same degree as the ER group. Beglar, Hunt, and Kite (2012) administered a practice reading rate test before posttests were administered to both experimental and control groups in order to not introduce a further independent variable. The presence of a second independent variable and the difference in the ER group’s versus the IR group’s preparation for the reading rate posttest does not prevent Huffman’s findings from supporting the premise that ER can improve reading rate relative to IR. However, the presence of this difference in measurement preparation and this second independent variable do require Huffman to state their presence within the limitations, and to hedge the strength of claims of the effectiveness of ER at improving reading rate relative to IR.

A third independent variable and its limited investigation: time on task

Huffman (2014) gives due attention to a limitation of Robb and Susser’s (1989) methodology: “Time-on-task was nearly double for the ER group, so the reading rate gains may be due simply to increased time spent reading rather than the pedagogical approach itself” (p. 21). However, despite Huffman stating the limitations of Robb and Susser’s (1989) methodology and highlighting the issue of time on task in Huffman, the extent to which the two treatment groups conducted respective treatments for different amounts of time was not fully investigated, and the explanation for it was limited.

Huffman (2014) commendably highlights the limitation of ER participants spending greater amounts of time reading relative to IR participants:

A final limitation is that the students in the ER group spent considerably more time reading during the semester than the intensive reading students, based on their self-reported data. It is possible and even likely, therefore, that the reading rate gains achieved by the ER group are due not only to the difference between the extensive and intensive reading approaches themselves, but also to the additional time the ER group students spent reading during the semester (p. 29).

This reduces the validity with which inferences regarding the relative effectiveness of ER and IR can be made. However, Huffman fails to hedge the strength of the inferences made, and states that “This study provides solid empirical data supporting the effectiveness of extensive reading over intensive reading for reading fluency development” (p. 28). Further, Huffman’s investigations into the degree ER and IR group participants conducted their respective treatments are limited.

Huffman (2014) states that “The mean number of hours per week spent reading was 3.59 (SD = 1.79) for the ER group and 2.44 (SD = 1.38) for the IR group” (p. 27). The statistical significance between groups was not reported in Huffman. This author calculated the statistical difference and effect size using the reported mean time spent reading, standard deviation, and sample size of each group. The two groups are significantly different in time spent reading ($t(64) = 2.91$ $p = .005$; $g = .71$). The effect size of .71, according to Cohen’s (1998) effect size criterion is between medium (.50) and large (.80). There are now three potential independent variables at work here: the reading treatment (ER versus IR), the weekly practice of timed reading (by the ER group but not by the IR group), and the statistically significant difference in time on task. As a result, this author argues that we should entertain the possibility that time on task and the weekly practice of timed reading only by ER participants had a greater influence on reading rate than the ER treatment itself, or even that the ER treatment had no significant influence on reading rate. As a result, Huffman’s inference that the statistical analysis “unequivocally support[s]” (p. 27) the *a priori* hypothesis that “Reading rate gains will be significantly greater for students in a one-semester college extensive reading course than those in an intensive reading course” (p. 22), and his conclusion that “This study provides solid empirical data supporting the effectiveness of extensive reading over intensive reading for reading fluency development” (p. 28) are excessive.

Limited evidence for the explanation of the difference in time on task

Regarding ER and IR group participants reporting different amounts of time spent conducting their respective treatments, Huffman (2014) stated the following:

From an experimental standpoint it would be ideal to control time on task, but from a pedagogical standpoint it can be argued that this difference in time spent reading is in itself an argument in favor of the effectiveness of extensive reading. It is also difficult; to justify placing artificial limits on the time students spend reading for the purpose of an experiment (p. 29).

Huffman's (2014) statement regarding the limitation of controlling for time on task suggests that the greater amounts of time spent reading by the ER group resulted from some characteristic of ER itself. However, no description, explanation nor evidence is provided that an unnamed characteristic of ER resulted in the ER group participants conducting ER for longer periods than the IR group participants conducted IR.

The only information readers of Huffman (2014) have regarding motivations for conducting ER is the amount read by the ER group, and the amount read relative to reading goals. Students were "told that they would be evaluated primarily on the number of pages they read, and that they needed to submit a book report to show that they had read each book. The amount read was evaluated on a sliding scale from 400 pages (passing) up to 800 or more pages (highest possible grade)" (p. 25). Despite this "the students read an average of 545.85 pages" (p. 24), substantially less than that required to receive the highest possible grade. This would suggest that students read for the period they did (at least partly) because they were being evaluated primarily on the number of pages they read. The high word goal of 800 or more pages may well have been the reason for the amount read, and not, as Huffman suggests, an unnamed characteristics or quality of ER.

Thus, it might be argued that the ER group participants' significantly greater gains in reading rate, which "unequivocally support the hypothesis" (Huffman, 2014, p. 27) that "Reading rate gains will be significantly greater for students in a one-semester college extensive reading course than those in an intensive reading course" (p. 22), are conceivably *not* the result of "the effectiveness of extensive reading over intensive reading for fluency development" (p. 28). It could be argued instead that the ER group participants' significantly greater gains in reading rate resulted in part from their weekly timed reading practice, regular encouragement to work on increasing their reading speed, and the significantly greater amount of time spent conducting reading by the ER participants than the IR group participants. It could be further argued that it is *not* true that "the difference in time spent reading is in itself an argument in favor of the effectiveness of extensive reading" (p. 29), but that this difference is an argument for the effectiveness of setting reading targets, evaluating students "primarily on the number of pages they read" (p. 25), and informing students of this evaluation criterion. As a result, Huffman's conclusion that "This study provides solid empirical data supporting the effectiveness of extensive reading over intensive reading for reading fluency development" (p. 28) is not valid.

Conclusion

Huffman's (2014) research methodology includes a number of novel characteristics which future ER research would benefit from utilizing. However, a significant difference in time on task between groups, combined with weekly timed-reading practice by only the ER group participants, results in the presence of not one independent variable (ER versus IR) but three (ER versus IR, weekly timed reading practice versus a lack thereof, and time on task). As a result, this author argues that Huffman cannot validly infer so strongly that ER more effectively develops learners' reading rate than IR. It is hoped that this discussion of Huffman (2014) will encourage

researchers to make appropriate inferences, and so provide stronger evidence for the inclusion of ER in language programs.

References

- Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning*, *62*, 665–703. doi:10.1111/j.1467-9922.2011.00651.
- Bismoko, J., & Nation, I. S. P. (1974). English reading speed and the mother-tongue or national language. *RELC Journal*, *5*, 86–89. doi: 10.1177/003368827400500109
- Carver, R. P. (1982). Optimal rate of reading prose. *Reading Research Quarterly*, *18*, 56–88. doi: 10.2307/747538
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Chang, A. C. S., & Millett, S. (2013). The effect of extensive listening on developing L2 listening fluency: some hard evidence. *ELT Journal*, *68*, 31–40. doi:10.1093/elt/cct052
- Chung, M., & Nation, I. S. P. (2006). The effect of a speed reading course. *English Teaching*, *61*, 181–204.
- Cramer, S. (1975). Increasing reading speed in English or in the national language. *RELC Journal*, *6*, 19–23. doi: 10.1177/003368827500600203
- Huffman, J. (2014). Reading rate gains during a one-semester extensive reading course. *Reading in a Foreign Language*, *26*, 17–33.
- Macalister, J. (2008). The effect of a speed reading course in an English as a Second Language environment. *The TESOLANZ Journal*, *16*, 23–33.
- Macalister, J. (2010). Speed reading courses and their effect on reading authentic texts: A preliminary investigation. *Reading in a Foreign Language*, *22*, 104–116.
- Robb, T. N., & Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, *5*, 239–251.
- Yen, T. T-N. (2012). The effects of a speed reading course and speed transfer to other types of texts. *RELC Journal*, *43*, 23–37. doi: 10.1177/0033688212439996

About the Author

Stuart McLean holds an M.S.Ed. in TESOL from Temple University, Japan, and is currently working on his Ph.D in Applied Linguistics at Kansai University. His research interests include second language vocabulary instruction, extensive reading, language assessment, and research methodology. He has published in *Reading in a Foreign Language*, *TESOL Quarterly*, *Vocabulary Learning and Instruction*, and *Language Teaching Research*. E-mail: stuart93@me.com