A Bayesian approach to measuring evidence in L2 research: An empirical investigation

What this research was about and why it is important

Researchers often use null hypothesis significance testing (NHST) to distinguish the random findings (noise) from replicable ones (signal) in their studies. While previous research criticizes the NHST approach, no clear alternative is offered to replace it. In this study, we proposed an alternative approach using the Bayesian methodology, called the Bayes Factor approach. Then to evaluate the potential of the proposed approach, we subjected a large sample of previously published results from four major L2 journals to a reanalysis. The findings showed that considerable disagreements arise between the two methods over the rejection of the null hypothesis. Based on these findings, we proposed a new threshold for declaring a statistically significant result in second language research through which the two methods of inference, NHST and Bayesian, could be reconciled.

What the researchers did

• They consulted previous surveys of L2 methodological practices to select the following four journals: Language Learning, Modern Language Journal, Studies in Second Language Acquisition, and System.
• They coded 119 studies that used t-tests in their designs focusing on tests that (a) were not used in pair-wise comparisons following (post-hoc) a larger analysis (often ANOVAs), (b) were not used for planned comparisons, (c) were not from relational analyses that use t-tests (e.g., regression) and (d) if from independent-samples, were calculated under the assumption of equality of group variances.
• They identified 418 such tests. This sample included 172 independent-samples t-tests, and 243 paired-samples t-tests, and 3 one-sample t-tests. The complete raw dataset for the study is publicly available at: https://raw.githubusercontent.com/izeh/l/master/l.csv.
• They developed a computer program available at: https://github.com/izeh/l/blob/master/d.r to extract the relevant data from each published study and reanalyze the results using their proposed Bayesian approach.
• They compared the original results of the published studies with those from their proposed method.

What the researchers found

• They found that in 64.06% of cases where p-values fell between .01 and .05 (i.e., evidence to reject the null), the Bayesian analysis found the evidence in the primary studies to be only at an ‘anecdotal’ level (i.e., insufficient evidence to reject the null).
• They discovered that that adoption of more stringent (i.e., lower) thresholds for researchers to declare a statistically significant finding could reconcile the two methods of inference.
• They showed that the use of a lower threshold for statistical significance can lead to more replicable findings, that is repeating one’s study under the original study’s stated conditions is more likely to lead to the same set of findings.

Things to consider

• Testing a null hypothesis is not enough to reject it.
• Specifying alternative hypotheses is needed to evaluate the findings under alternative positions as well.
• While a finding can be considered unlikely under the null hypothesis it may be even less likely under the alternative hypothesis.
• Statistically significant findings from small-sized research are often likely to be exaggeration or mis-representations of the relationships/effects of interest.