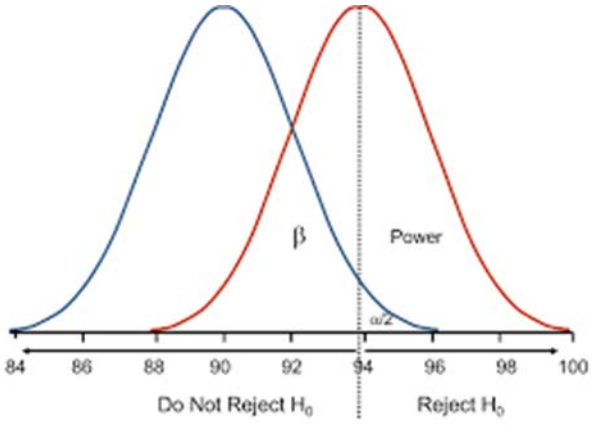
Null Results in Bilingualism Research

Ellen Bialystok
Department of Psychology
York University
Toronto, Canada

Like democracy, empirical studies are the worst form of scientific methodology, except for all the others. Experiments are inherently imperfect, the designs are inevitably incomplete, and the results are buffeted by countless uncontrolled variables, both known and unknown. And yet they are our most effective means of deriving new knowledge, testing ideas, creating and evaluating effective interventions, and understanding our world. In psychology our goal is to develop principled descriptions of human behavior that reveal some underlying pattern that is generally more true of humans than it is false. And because empirical studies rarely yield simple clear answers, how we talk about those results, particularly when there is ambiguity or inconsistency, is essential to the scientific project.

Psychology has recently been drawn into the growing movement called the "replicability crisis". The outcome of an experimental "null effect", once used as a signal to re-examine the design, reconsider the task, or re-evaluate the statistical procedures, is now used to challenge entire areas of study. The resulting discussion has in some cases reduced complex problems to simple binary choices: an effect exists or it does not. Research on bilingualism has been one such target of this trend.

Studies over the last 20 years have demonstrated systematic changes to cognitive and brain systems across the lifespan that can be traced to bilingual experience. Many inconsistencies exist, many questions remain, and research is ongoing. However, recent studies reporting null results between language groups has begun to overshadow the substantial body of research showing positive effects of bilingualism, leading some to conclude that bilingual effects simply do not exist. But what is the correct interpretation of these null results?



Consider first the epistemological status of null results. Because most statistical approaches are based on probability testing, null results in themselves have no meaning, but replicating that outcome with various experimental modifications allows one to rule out the experimental manipulation as a factor. In this way, null results provide an important constraint on interpreting the boundaries of an effect. In bilingualism research, null results reported for more than a decade have shown that there are no

behavioral differences between young adult monolinguals and bilinguals performing simple executive function tasks, increasing confidence in that conclusion.

But these results are only part of the evidence needed to determine whether bilingualism affects mind and brain; because the tasks used in psychological research provide imperfect assessments of cognitive function, evidence from brain involvement is also required. Here, the evidence is clearer: even when behavioral performance is equivalent, bilinguals use different functional networks than monolinguals and structural brain analyses reveal consistent difference language groups. Something has changed with bilingual experience.

A second issue is the need to provide a theoretically-grounded rationale for why the targeted process is *expected* to distinguish between the groups. If a null result is used to refute a previous positive result, then it must *prima facie* be a test of the same ability. Although most research in bilingualism relies on a common set of executive function tasks, increasingly research reporting null results is based on tasks and abilities that have not previously demonstrated positive outcomes and have no obvious theoretical reason for expecting they would. Yet these null results are interpreted as counter-evidence to claims about bilingual effects on cognition more broadly. Null results are empirical tools, not theoretical ones, and are specific to the evidence from which they emerged.

Finally, bilingualism has special status among the independent variables in psychology. There is no standard definition for "bilingual" and no objective demarcation point beyond which experience with another language crosses some boundary from monolingualism to bilingualism. There is emerging evidence that bilingualism is best conceived as a continuum, but there is not yet an established metric for describing that continuum. Therefore, studies of bilingualism demand scrupulous attention to the description of the population.

For this reason, no two bilingual samples can be identical, and research conducted in different communities will reflect the details of that bilingual experience – status relation between the languages, proficiency with the languages, usage patterns including switching, among others. Thus, comparing results across linguistic, national, cultural, or other grouping factors is precarious. Bilingualism research needs to document in detail the type of bilingual experience represented so that results can be properly interpreted and responsibly compared to others that may have led to different outcomes.

The best tool we have at present for advancing a scientific agenda is the model provided by empirical research. The recent interest in replication is an essential part of science – it is imperative to establish the reliability of research results by providing checks and balances. The problem is in how to interpret the null effect when the replication fails. There are many reasons that experiments end in null results, including absence of the effect being investigated, but they could equally signal experimental conditions or theoretical principles.

Some argue that null results provide as much information as significant ones and that all null results deserve to be published, with failure to do so constituting publication bias. However, such a policy would be just as egregious a violation of the scientific method as would publishing all experiments that achieve $p < .05$. The only way to move forward on this project of trying, however imperfectly, to understand the human mind is to leave bias behind and interact responsibly with all the data to evaluate its meaning, both in terms of what is statistically significant and what is not.

Further Reading: Bialystok, E. (2016). The signal and the noise: Finding the pattern in human behavior. *Linguistic Approaches to Bilingualism, 6,* 517-534.