Design and Methodology in the Science of Reading

Christopher Schatschneider

Dana Santangelo

Christine M. White

Cristian E. Vazquez

Emma D. Friedmann

Florida State University

Florida Center for Reading Research

This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200020 to the Florida Center for Reading Research at Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Published in The Reading League Journal, Jan/Feb 2023

Design and Methodology in the Science of Reading

by Christopher Schatschneider, Dana Santangelo, Christine M. White, Cristian E. Vasquez and Emma D. Friedmann

What is the science of reading? The Reading League, in their document "Science of Reading: Defining Guide" (TRL, 2022), proposes that "The science of reading is a vast, interdisciplinary body of scientifically-based* research about reading and issues related to reading and writing." (p. 11.) The asterisk is carrying a lot of weight here. What does it mean for something to be scientifically based? It means that the evidence for a claim is weighed by the design and methods used to collect and analyze it, and the claim takes into account the limitations of that evidence. In this paper we will discuss some of the designs and methods used to investigate how children learn to read for understanding and provide a brief history into the use of these designs in reading and educational research.

The methodologies used in the science of reading are numerous. They range from eye-tracking, fMRI, cognitive psychology experiments, connectionist models, individual differences studies, group-based designs, and single-case studies, to ethnographic studies, and observational studies, among other approaches. While all contribute to the science of reading, design and methodology have implications for the evidentiary value one should place upon a study's results, or more plainly, how the results of a study should be used. We would like to start with a story about a mismatch between a claim of knowledge and the information used to back that claim.

In 1967, Professor Ken Goodman published a very influential paper titled "Reading: A Psycholinguistic Guessing Game". In this paper, he argued against the common perception of reading as a "precise process [that] involves exact, detailed, sequential perception and identification of letters, words, spelling patterns and larger language units" and instead proposed the following:

> Reading is a selective process. It involves partial use of available minimal language cues selected from perceptual input on the basis of the reader's expectation. As this partial information is processed, tentative decisions are made to be confirmed, rejected or refined as reading progresses. More simply stated, reading is a psycholinguistic guessing game. (pp 126-127)

These were strong claims. To support them, Goodman provided anecdotes of mistakes made by children reading aloud, a method he referred to as miscue analysis. He argued that the mistakes made by children while they were reading indicated that they were actively guessing at what the next word or words should be, and that they were not reading every word in the passage. This view of reading as a psycholinguistic guessing game would form the cornerstone of the Whole Language approach to reading instruction that still exists in different forms today.

Goodman took his observations and declared this view of reading as fact. However, he placed too much evidentiary value upon his own personal observations. Arguably, observations are only the beginning of scientific knowledge. There are times when our observations can lead us to the correct conclusions, but there are other times they can lead us astray. Observations give us ideas about what could possibly be occurring, but from these observations, hypotheses should be formed, and these hypotheses should be tested. Goodman's claims went beyond the limitations of his methodology and did not hold up to rigorous testing. About a decade later, eye tracking methodologies were developed, and we were able to observe that our eyes land on almost every word we read (Rayner & Juhasz, 2004)—and if we do skip a word, it is usually a short high frequency word or noun marker such as *a* or *the*. Further, Stanovich, Cunningham, and Feeman (Stanovich et al., 1984) discovered that guessing a word in text is a hallmark feature of poor readers, not proficient readers who will spend more time sounding a word out.

Goodman's mistake was not in making observations. Observing and generating hypotheses concerning observations is foundational to science. However, Goodman placed too much evidentiary weight on anecdotal evidence. His observations should have been a starting point, not an endpoint.

This story implies that there is a hierarchy to the different types of scientific evidence in regard to how much weight we should place upon them, and this in turn is dependent on what types of questions are being asked. Given that the science of reading is built upon a vast number of studies across many disciplines and study designs, it's important to be able to identify a study based upon its design and methods and come to a conclusion about what it provides in terms of what kind of information this type of study can generate.

One such set of studies are observational studies. These include case studies, ethnographic studies, demographic studies, survey studies, focus groups, structured interviews, or any study that has as its goal to describe phenomena. In these studies, the researcher does not try to influence the phenomena in question. In fact, they often take great care in trying to be as neutral or invisible as possible. Observational studies are important for describing a phenomena and its co-occurrences. They provide the information needed to generate ideas about why a particular phenomenon may be occurring. In other words, they provide information as to *what is happening* and can help generate hypotheses as to *how it's happening*. So, the evidentiary value of observational studies depends on the question being asked. In the example above, Goodman (1967) stated that our eyes only sample enough words from a text to make meaning. This could be reframed as, "Do our eyes look at every word we read?" That is, we want to know what is happening with our eyes when we read. This was answered by the numerous eye-tracking studies referenced above that found that our eyes land on almost every word we read. These eye tracking studies were also observational studies. They just happened to use a more rigorous methodology to observe what our eyes are doing when we read.

Another related set of studies are individual differences studies. In a sense, these are observational studies with a measurement component with the goal of numerically quantifying a phenomenon such as reading ability so that we assess "how much" skill each child has, which allows us to look for other phenomena that may be related to reading. These studies are looking for *relations* among phenomena, but ultimately they are observational studies that inform us as to what is happening and allow us to generate hypotheses about how something is happening.

There have been numerous individual differences studies on reading development. One such study was conducted by Wagner et al. (1993) in which he and his team measured what they believed to be different aspects of phonological processing and investigated whether these skills were related to reading in a group of kindergarten and second grade children. The measures developed by Wagner and colleagues were grounded in a theory of phonological processing proposed earlier by Wagner and Torgesen (1987). In this study, they found five distinct but correlated skills that tapped phonological processing and found that they correlated strongly to reading ability, demonstrating a strong relation between reading and phonological processing skills. This study, and many other individual differences studies since then, have demonstrated this relation. But establishing that a relation exists does not provide evidence of what is *causing* this relation. That is, evidence from individual differences studies cannot be used to claim that, for example, acquiring or increasing phonological processing skills will increase reading skill. Why not? It is because learning to read is a complex process that may have many causes, and it's possible that there are alternative explanations as to why phonological processing skills are related to reading. For instance, it's possible that there are other phenomena such as vocabulary knowledge that may be influencing both phonological processing and reading. The evidential value of individual difference studies does not support causal claims. They can suggest potential causal mechanisms, but in order to support a claim of causality, we need to run an experiment.

As opposed to observational or correlational studies, experiments are studies where the investigator introduces something into the environment to measure its effect. Observational and individual differences studies do not introduce anything new into the environment. Experiments look to answer the question, "If I do something different, what will the effect be?" Shadish et al. (2002) suggest that experiments are well suited to studying cause and effect relationships because they (a) ensure that a presumed cause is deliberately manipulated and, thereby, precedes the observed effect, (b) incorporate procedures that help determine whether the cause is related to the effect, and (c) incorporate procedures to minimize and/or assess the influence of extraneous factors that could produce the effect presumed to be attributed to the cause.

The best way to determine if one thing is causally related to another would require the use of a time machine. For example, if we wanted to know if a particular classroom practice was an effective way to teach reading, we could administer it to a classroom over a year and examine how children grew in their reading skills. Then we would send them all back in time to the beginning of the year, but this time, we would not deliver that particular classroom practice. We would follow them over the course of the year and again measure how much skill in reading was gained. Why would this be an ideal way to determine causality? Because the time machine would produce a condition that would be an excellent representation of what would have happened had we not delivered that particular classroom practice.

Why do we need experiments in the science of reading? First, we need to acknowledge that reading is a complex behavior with many factors that help or hinder its development. Students show up in the classroom with a myriad of backgrounds, experiences, and skills. Their home environments, classroom environments, cognitive skills, motivations, and larger societal contexts all influence their reading development. It is these multiple sources of influence that make it challenging to determine if a particular program or classroom practice will increase reading skills. So how can we control all of these other sources of influence when we want to know if a particular practice is effective? Well, the short answer is that we can't. However, we can try to minimize their influence, and our best chance to do this is to use random assignment. Specifically, since we have no way to control these other factors, we need a way to create another condition that has the best chance of representing what would have happened had we not delivered the treatment. Random assignment accomplishes this by distributing non-experimental factors, such as motivation to read or prior experiences with reading, randomly across the treatment and control conditions so the two groups are equivalent on all of the other potential influences on reading development. This helps minimize any other plausible reasons why one group may grow differently in reading skills, and removing other plausible alternatives is one of the criteria for establishing causality.

There are many examples of the use of random assignment in the investigation of reading development. One such example was a study conducted by Blachman and colleagues (Blachman et al., 2004). Blachman identified 69 students in second and third grade who had poor word reading skills and randomly assigned them to either receive intensive one-on- one tutoring for eight months using an intervention that emphasized the phonologic and orthographic connections in words (the treatment group) or to receive the standard remediation services provided by their school–often referred to in reading research as a business-as-usual control. Multiple measures of reading and related skills were administered at the beginning of the year, the end of the year, and one year after the intervention was concluded, allowing Blachman and colleagues to estimate change in reading ability and to compare the change in reading skills between the treatment group and the control group. This study was able to address the three conditions for causality: (a) the cause (the intervention) preceded the effect, (b) by using measurements they were able to establish that the cause was related to the effect, and (c) they were able to remove other plausible reasons why the two groups would grow differentially over time by using random assignment. By directly comparing the growth rates between those who received treatment to those who received the business-as-usual intervention, they were able to estimate the *effect* of the intervention. This study showed a large effect of the intervention condition compared to the regular treatment on reading and reading related skills, and interestingly, the effect on word reading skills persisted 10 years later (Blachman et al., 2014). It is because of the randomized design of this study that we would give its effects higher evidentiary weight than studies that did not employ a randomized design when investigating which practices produce positive gains in reading skill.

However, there may be times when randomizing people to conditions is either not possible or may not test the question being asked. In this case, researchers may employ a quasi-experimental design. A quasi-experimental design has all the features of an experimental design except for one crucial difference: the groups are not formed via random assignment. This has a large impact on the ability to make causal claims. Because the groups are not formed via random assignment, it becomes challenging to rule out other plausible reasons why there may be a difference between the group getting the treatment and the control group. This possible pre-existing difference between the groups is known as *selection bias*. The potential for selection bias means that much more effort must be put into comparing the groups on any pre-existing differences that could be related to growth in reading in a quasi-experiment compared to a randomized design. But even if you could establish that the groups were equivalent at the beginning of the study on all measured variables, there still exists the possibility that they may be different on some unmeasured variable that may influence reading outcomes. Thus, many will give quasi-experimental studies less evidential weight when looking for causal relations.

One relatively famous reading study that employed a quasi-experimental design was conducted by Foorman et al (1998). Foorman and colleagues wished to compare the Whole Language approach to reading instruction to two different code-focused approaches–one where there was explicit instruction in code-based skills and another where the code-based instruction was implicit. In this study, the treatment was delivered by the classroom teachers, as opposed to the Blachman study described previously that delivered one-on-one tutoring by the researchers. Since the Whole Language approach is part practice and part philosophical belief, it would not be possible to randomly assign teachers to be Whole Language teachers. The prevalent instruction in the area was Whole Language, so the research team went from school to school and visited each principal to see if there were teachers who would be amenable to delivering a code-focused curriculum. This introduces the potential for selection bias in that the teachers who were open to delivering a code-focused curriculum may be different from those teachers who were unwilling to do so, in a way that may impact their students' performance. It was also possible that the schools that had principals willing to try a new curriculum were different, both in terms of the students who attended that school, and in supports available to teachers. These possibilities exemplify how selection bias may influence the estimation of the treatment effect.

Since random assignment was not used, the researchers needed to investigate any potential differences between the classroom environments that were using either a Whole Language approach or a code-focused approach. The team collected a large number of school, teacher, and student variables that had the potential to influence the results. Their investigation revealed no significant differences among the treatment groups on a number of teacher and student level variables. This helps strengthen the conclusions that can be drawn from a quasi-experimental study in that the influence of those variables on selection bias could potentially be ruled out. This study found that the students who received explicit instruction in letter-sound correspondences over the course of the year performed significantly better on end-of-year assessments of phonological awareness and word reading skills than the students who received implicit code-based instruction or the Whole Language approach.

All of the studies presented here contribute to the science of reading. From observational studies to randomized control trials, all of them add to our knowledge of reading. But different types of studies are designed to address different types of questions, and the claims we can make based on the results of a study depend on the design and methodology used to answer those questions. Observational studies are designed to answer descriptive questions and to help identify *what is happening*. Individual differences studies answer similar questions to observational studies while adding a strong quantitative measurement component and quantifying the degree to which phenomena such as phonological processing and reading skills are related to each other. Finally, experiments and quasi-experiments aim to address what causes what, with randomized control experiments producing higher-quality evidence than quasi-experiments because they minimize the influence of extraneous factors that may bias the results.

References

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of Intensive Reading Remediation for Second and Third Graders and a 1-Year Follow-Up. *Journal of Educational Psychology*, *96*, 444–461. https://doi.org/10.1037/0022-0663.96.3.444

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, *106*(1), 46–57. https://doi.org/10.1037/a0033663

Foorman, B.R., Francis, D.J., Fletcher, J.M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 1-19.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, *6*(4), 126–135. https://doi.org/10.1080/19388076709556976

Rayner, K., & Juhasz, B. (2004). Eye movements in reading: Old questions and new directions. *European Journal of Cognitive Psychology*, *16*(1–2), 340–352. https://doi.org/10.1080/09541440340000385

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Relation between early reading acquisition and word decoding with and without context: A longitudinal study of first-grade children. *Journal of Educational Psychology*, *76*(4), 668–677. https://doi.org/10.1037/0022-0663.76.4.668

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, *101*, 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*, 83–103. https://doi.org/10.1037/0022-0663.85.1.83