Humility and the Scientific Method

For Nature, or more precisely experiment, is an inexorable and not very friendly judge of his work. It never says “Yes” to a theory. In the most favorable cases it says “Maybe,” and in the great majority of cases, simply “No.” If an experiment agrees with a theory it means for the latter “Maybe,” and if it does not agree it means “No.” Probably every theory will someday experience its “No”—most theories, soon after conception. (Einstein)

This quote underscores the essence of the scientific method and the accumulation of evidence. Scientific inquiry yields evidence that is available for continuous scrutiny and modification. Claims based on scientific evidence, no matter the methodology and research design used to acquire the evidence, are only as honest as their replications and openness to being overturned. In the best of all worlds, scientific inquiry and the evidence produced must be transparent and objective. The scientific method is democratic—it does not take sides. The scientific method is humble—it takes no for an answer. Scientific evidence illuminates a path that in the future may twist and turn and, in some cases, be erased altogether. In short, when we evaluate scientific evidence, we do not make claims about its perpetuity.

Scientific Evidence and Education

For much of education’s history, decisions have been based on untested theories and a search for the next magic bullet. Whether in the policy, regulatory, or instructional domains, they have turned out to be counterproductive, and the search for the next evidence-free solution has begun anew. Some untested theories and their ensuing instructional practices are touted as research- or evidence-based despite a lack of objective data reflecting effectiveness. When these practices invariably fail, students are not the only casualties. Confidence in the concept of scientific evidence also suffers. Consistent resistance to the use of evidence in education has deep roots in the institutions that prepare future educational leaders and teachers. Many educators in training are exposed to principles and practices based on invalid assumptions, fads, appeals to authority, and personal philosophies. In too many colleges of education, the use of scientific evidence and effectiveness research to inform policy and practice is rejected in favor of value judgments bearing on “what is best for children.” For example, when scientific evidence obtained through rigorous study determined the importance of direct instruction on developing reading skills, this finding was downplayed or ignored because it did not reflect a belief that learning to read is natural. In essence, personal values and opinions have dominated educator training over and above objective evidence of effectiveness.
Prioritizing large-scale objectivity of scientific evidence over personal opinions and beliefs is going to be difficult to achieve because evidence-based education has a long history of conflict with individual beliefs regarding how and what children need to learn. One step toward achieving this will be to make the range of evidence-based programs, curricula, instructional designs and strategies, professional development models, and policy development and implementation criteria that have shown evidence of effectiveness available to educators. A second step in prioritizing objectivity over personal opinions and beliefs will be to assist educators and the public in how to ask nontechnical questions when they assess the credibility of educational program effectiveness claims. A third step is to utilize both scientific evidence and evidence from High-Performing Educators (see page 6 for a description of High-Performing Educators) and the schools that consistently demonstrate significant student achievement, even in the most vulnerable communities. Debating different points of view on how best to improve student achievement can be productive only when the objectivity of scientific evidence is prioritized. Unfortunately, education is awash with evidence. Educational publishers label their materials as evidence-based, media reports champion the implementation of evidence-based practices, and consultants promise to provide evidence-based guidance. Some are based on evidence that has tenuous implications for practice, drawn from only tangentially related data and reflecting misrepresentations of original research—or predicated on untrustworthy research. Faulty or weak claims about evidence threaten to swamp efforts by the very people who want to employ evidence in making educational decisions.

Fortunately, there is a substantial body of relevant and trustworthy scientific evidence. Scientific data are especially relevant for determining what works for whom under what conditions. Unfortunately, educators’ efforts to identify and implement evidence-based practices are frequently impeded by the often dense and difficult-to-navigate nature of scientific publications. Researchers typically write reports on their research for an audience of other researchers—they are rarely accessible to educators, who might most benefit from the findings, and it would be an onerous task for educators to sort through the thousands of options or to translate the scientific jargon into classroom practice. In curating evidence about beneficial educational practice, the EAC will find and disseminate resources as well as organizations that are trustworthy and accessible. The EAC will also share recommendations for decisions about education policy, educator preparation, assessment, curriculum and instruction, and effective implementation that are in alignment with scientific findings and clear about actionable and practical educational efforts.

For Whom Is the Evidence Curated?

The EAC serves diverse consumers of educational research. Those consumers include classroom teachers, building and system leaders, policy experts, nonprofit advocates, journalists, and institutions of higher education. The diversity of the consumers for EAC’s work on evidence requires consideration of a wide range of research outcomes, methods, and reporting.

Different consumers of educational research have different interests. Some may be concerned about the provision of educational services to different geographic areas or different student groups, including students with disabilities, linguistically diverse students, and students who are economically disadvantaged. Some may focus more on economic input than financial factors. Others are interested in variations in instructional practices, arrangements of groups of students, selection of curricula and assessments, or school organizational structures. Still, others may be concerned with how higher education institutions should prepare preservice educators and how in-service providers should be
determined.

The evidence bearing on the concerns of these different constituents comes from different disciplines, employs different research methods, is available from different sources, and is communicated in different ways. The EAC aggregates evidence across disciplines, methods, sources, and communication modes so that it can meet the diverse needs of consumers. Regardless of their background, almost all the consumers of EAC efforts share concerns about the outcomes of educational decisions that promote student success. Thus, the EAC aggregates resources and organizations for decision-makers who are focused on improving student outcomes.

★★★★

The EAC Focuses on Multiple Outcomes

Evidence within education has examined diverse outcomes. A comprehensive catalog of educational outcomes would be tortuous to read, but it would surely include educational results in the areas of academic achievement (reading, writing, and arithmetic). Beyond the “three Rs,” the list would also include academic subject areas (e.g., mathematics, sciences, social studies, career education). Social and behavioral development represent another domain of concern. The EAC will curate and disseminate both evidence and connections to organizations specializing in these areas.

Disseminating evidence about outcomes in academic subjects and social-developmental behavior does not mean that the EAC will curate and disseminate evidence for only those areas. Although educators have not achieved consensus on definitions for scientific evidence, accountability, and what metrics are used to gauge growth in student achievement, the EAC will be able to identify and share the evidence that is available and to help consumers assess whether it is relevant to their needs.

Factors that influence academic and social outcomes are important topics for EAC efforts. However, some factors associated with higher and lower outcomes are unchangeable by educators. For example, although socioeconomic status is correlated with academic achievement, it is not something that educators can influence; however, they can influence the type of instruction provided to economically disadvantaged students, which can make a significant difference in their life trajectories.

★★★★

What Research Is Relevant for EAC Work?

As is true of education in general, there are disagreements about how studies should be conducted. Researchers often prefer one set of methods over another; however, the EAC sees value in diverse methods. The decision about what methods to use depends on the research question. For example, questions about what teachers like and dislike about a curriculum are descriptive questions that can be addressed by interviews and observations. Questions about which of the two curricula result in higher levels of achievement can be addressed using systematic causal-comparative methods that require experiments.

Because the EAC frequently focuses on questions about what works, it will prioritize studies that use causal-comparative methods. Physical and biological sciences rely on the criteria within the scientific method to determine cause and effect. For example, hypothesizing that curricula using an instructional principle (Method “K”) will lead to higher achievement than curricula using a competing theory of instruction (Method “M”) must be tested objectively. Researchers could ask teachers whether they prefer K or M or which one they predict will lead to higher achievement. But to truly determine which has
more positive effects on student outcomes in schools with teachers who have equal support requires the scientific method. In research of this sort, designs, measurement procedures, and curricular implementation must be implemented with fidelity, and that fidelity should be documented.

Reports of experimental studies must faithfully represent how the research was conducted. External unbiased peer review of reports of such studies is an essential feature in establishing the veracity and credibility of the claims made in such studies. Scientific evidence derived from experimental studies reflects empirical observations that are quantified to assess the extent of the relationship between treatment factors and outcomes that are free from nonexperimental bias (i.e., “confounding factors”). In general, randomized control trials (RCTs), quasi-experimental designs (QED), regression-discontinuity designs (RCD), and single-case research (SCR) are able to isolate the experimental manipulations that were related to the outcomes.

No single study is infallible. Replication of studies, procedures, and results must be conducted. When results are not replicated, researchers consider what differences might have caused the failure to obtain similar results. Were they conducted differently? Were there important, undocumented differences between the participants (students, teachers, schools)? Might there actually be no difference between the methods compared?

When two studies using similar methods result in consistent findings, the degree of trust that can be placed in those findings increases. When multiple studies repeatedly result in similar findings, trust in those findings increases even more. When there are dozens of studies that have compared a particular instructional program, practice, curriculum, or procedure to other programs, practices, curricula, or procedures, it is possible to make broader statements about effectiveness. It is even possible to examine interesting concerns: if it works better with older students; if it is as effective in groups of 15 or more students as it is in smaller groups; if students from certain demographic groups do as well with this method as they do with other methods, etc. Such questions can be addressed when researchers combine findings from multiple studies, sometimes using the method called “meta-analysis.”

As preferred as they are, RCTs and other experimental studies are expensive. They also require substantial research expertise and dedication of resources to the management of study implementation at research sites. Uncontrolled school factors make it difficult to implement RCTs with fidelity; teacher turnover, leadership changes, policy revolutions, and other factors can affect the implementation and fidelity of the study. Sometimes, it is possible to examine these uncontrolled school factors using quantitative methods as a part of integrating multiple studies, but doing so has its own set of challenges.

The costs and demands of conducting systematic empirical analyses of educational methods have restricted how many rigorous experiments have been reported. As a result, EAC teams may need to depend on other kinds of evidence on methods. One potential source of evidence is careful analyses of instruction when a teacher, school, or LEA consistently yields impressive results. Researchers might identify what we call High-Performing Educators (HPEs). HPEs are educators who have consistently produced higher-than-predicted levels of achievement among their students over a period of many years, with expectations based on students in other schools with comparable demographics. HPEs can be teachers, principals, districts, and, in some cases, even states such as Mississippi. HPEs allow explanation of the “how,” not just the “what.” In other words, the knowledge of HPEs is contextual in that it is clear about the specifics of what works for which students in which settings when implemented by whom.

The EAC also curates and disseminates research on the results of instructional methods that are
subjective in nature. Although consumers may be interested in how to raise test scores and improve behavior, they may also want to adopt methods that their students, teachers, and other educators consider rewarding. EAC analyses can provide evidence about the social validity of practices, procedures, and methods. What do consumers (students, parents, teachers, principals, etc.) say about the goals of specific methods, curricula, or procedures? To what extent do consumers agree with the methods employed in pursuing those goals? How happy are consumers with the actual outcomes of the methods? When evidence about the acceptability of goals, procedures, and outcomes exists, EAC teams will be able to share that evidence with consumers, too.

⭐⭐⭐

**Summary of Types of Evidence**

The following list summarizes the types of evidence considered by the EAC.

1. **Experimental and Quasi-Experimental Studies.** Experimental and quasi-experimental studies (aka scientifically-based research) are best for determining the causal relationships between instruction and student outcomes. These studies are described in the accompanying table that, although specific to reading, captures the features of such studies; it is taken from The Reading League *Defining Guide.¹*

<table>
<thead>
<tr>
<th>REQUIRED COMPONENTS</th>
<th>DEFINITIONS</th>
<th>WHY IMPORTANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design that is experimental or quasi-experimental.</td>
<td>Experimental design features one or more experimental groups and at least one comparison group. Participants are randomly assigned to groups.</td>
<td>Experimental and quasi-experimental designs allow researchers to determine if a particular variable being studied is the reason for improved outcomes.</td>
</tr>
<tr>
<td>These designs specifically answer questions about why individuals have difficulty learning to read and write, as well as which practices are effective.</td>
<td>Quasi-experimental design does not utilize random assignment. Participants are sometimes compared to groups with similar profiles.</td>
<td>Random assignment, recognized as the gold standard, provides a clearer link between cause and effect because it helps control the effects of variables other than the experimental treatment. This allows for greater confidence that the treatment is what led to improved outcomes.</td>
</tr>
</tbody>
</table>
   | Detailed description of study methods and population for replication, generalizability, or refinement of findings. | To have confidence in findings, a convergence of evidence is needed. Detailed descriptions regarding design, participants, settings, instructional practices, measurements, and outcomes must be provided to replicate the study (i.e., conduct another study in a similar manner). Generalizability is the extent to which the findings of a study would be expected in real-world contexts. | It is important to show that scientific findings are unbiased and to determine for whom and under what conditions positive outcomes are produced. Replication is what leads to a large body of studies with similar results so that we can:  
   a. Conclude findings are consistent (e.g., "on the right road")  
   b. Conclude findings are not consistent (e.g., more research needed)  
   c. Discover new questions to be studied. |
   | Publication in a peer-reviewed (referenced) journal. | Peer-reviewed journals provide a rigorous review by multiple independent scientists with relevant expertise. | Clear descriptions of the context in which the study was conducted, the resources involved, and the participants allow readers to evaluate whether similar findings might be expected in their situations. |

² Defining features of types of research methods. Courtesy of The Reading League.

2. **Qualitative Studies, Brain Imaging Studies, Correlational Studies, Observational Studies, and Meta-Analyses.** These are methods and approaches of research that are most useful when the research questions are not seeking to address causal claims. Qualitative research is useful for the framing of future experimental research.

3. **Qualitative Research** includes evidence from high-performing educators (HPEs) who, as noted previously, over several years have consistently produced higher-than-predicted levels of achievement for their students. These HPEs and their schools are essential for moving the scientific research into practice. Outcomes from HPEs provide “proof points” of how practices based on scientific evidence are implemented with or without effectiveness (e.g., student data, dosage, frequency and duration, grouping, delivery).

4. **Evidence of Program Effectiveness.** When a specific program’s effectiveness has been researched, the EAC will elevate its methodology and data.

¹ https://www.thereadingleague.org/what-is-the-science-of-reading/defining-guide-ebook/
The Role of High-Performing Educators (HPEs)

In an article published by the American Federation of Teachers, two highly regarded reading research experts pointed out an overlooked factor that must be considered if educators are to be able to bring about improvement in student literacy performance, especially in reading, that the Science of Reading or evidence-based reading has shown to be possible.

“Having been researchers studying mechanisms for improving literacy outcomes for more than 30 years, we offer a more sober—and sobering—review of what is known about how to help struggling readers.

To begin, we confess that there are some rather large holes in our collective knowledge. We know more about the science of reading than the science of reading instruction. In other words, we know a lot more about what components are associated with improved outcomes for each stage of reading development (e.g., phonemic awareness and the alphabetic principle are essential for beginning readers) than we do about how to teach all these components to a class of students with diverse learning needs. Similarly, we know more about interventions for students with mild to moderate reading difficulties than we do about students with severe reading difficulties. .... Finally, in policy development, we have not capitalized on theory and science for effectively implementing new practices in schools.” (Vaughn & Fletcher, 2020)

The simple statement that researchers know more about the components associated with improved reading performance than they do about how to teach all these components in a classroom leaves educators who want to improve students’ literacy performance wondering what actually must be done in the classroom and where to go to get guidance.

In this paper, we present and define the term High-Performing Educators (HPEs). These are educators who, over several consecutive years, have produced significantly higher than predicted levels of achievement in their students. These High-Performing Educators can be individual educators, but they can also be whole schools and districts that consistently produce strong achievement, especially with economically disadvantaged students and marginalized populations. HPEs move the research to practice and thus provide the evidence of what works. These are educators at schools who have done it, not just once in a relatively short study but consistently over a period of years.

High-Performing Educators are teachers in the classroom who bring evidence-based instruction to their students, informed by practices cognitive researchers and experts, such as Barak Rosenshine and John Hattie, who have found it to be the most effective for learning. Furthermore, the principal who supports the teachers can also be considered an HPE. These HPE principals support their teachers by ensuring there are adequate resources in time, coaching support, staffing, materials, and assessments. They ensure that high expectations are in place and implemented for staff and students. Other HPEs outside of the school unit are district administrators who develop policies that support effective instruction and provide consistency in leadership among the different district departments. Finally, the superintendent in a school system that has high-performing schools is responsible for all activities throughout the district.

At the state level, High-Performing Educators can also be found in state departments of education when

---

policies are made and implemented to support gains in student performance.

How Do We Identify High-Performing Educators

School Level

As a first step in identifying High-Performing Educators, we use data. We look for individual schools, multi-school projects, districts, and states that have produced higher levels of student performance compared to schools, projects, districts, and states with similar populations.

For example, to identify reading programs that have been effective in teaching beginning reading to low-income populations in a state or large district, we might list all the schools that have a specified percent of students qualifying for free or reduced lunch and compare them by the percent of students scoring at proficient or advanced on their third-grade state test. Schools that were in the top 5 or 10 percent would warrant further study to see how the good results are being achieved.

State Level

In the last several years, the state of Mississippi (see Butler, 2024) has made very significant improvements. Comparing its accomplishments to other states, we can determine which states also implemented systems that resulted in strong overall achievement.

Illustrative HPE Practices for State-Level Administrators

An analysis of commonalities among states that have made dramatic improvements in student achievement can be the source of HPE practices for state-level education administrators. One analysis that looked at Mississippi and Tennessee can be found in the article New Report Highlights States that Are at the Vanguard of the Reading Revolution. (Olson, 2023)

Here are the commonalities:

1. A multi-year strategy, rather than a piecemeal approach, to changing literacy instruction.
2. Investment in initial statewide training for educators, followed by ongoing support as teachers applied what they learned in their classrooms.
3. Commitment that teachers and students use only high-quality instructional tools grounded in the science of reading.
4. Observations of classroom instruction using rubrics to determine how well teachers are implementing the new reading strategies so district and state officials can identify areas for improvement or areas where teachers might need additional training or more resources.
5. Screening of early elementary students three times a year to determine students at risk and development of an improvement plan for the identified students.
6. Provision of interventions for struggling students aligned to the appropriate curriculum used during core instruction so that what they learn during tutoring, summer programs, or before- and after-school programs connects with what happens in their classroom.

Butler, K. (2024). The Long View of Literacy in Mississippi: Major Components that Contributed to Reading Achievement in Mississippi… and Beyond! The Reading League Journal May/June 2024.

The Alabama Literacy Act illustrates a requirement to seek input from HPEs, in this case, seeking input from the leaders of the Barksdale Reading Initiative, which was a driving force in the Mississippi miracle. The act states that “…all public education preparation programs offer a minimum of nine hours of coursework based on the science of reading. The state also stipulated that Educator Preparation Programs (EPPs) submit syllabi to the Alabama State Department of Education for evaluation. To ensure programs are in compliance with the Alabama Literacy Act, the state contracted with the Barksdale Reading Institute to provide external review of course offerings.”(The Hunt Institute, 2023)

District

Data from states comparing the performance of districts can also be used to determine High-Performing Educators at the district level. In a recent book by Karin Chenowith, Districts that Succeed- Breaking the Correlation Between Race, Poverty, And Achievement, Chenowith describes what five districts serving low-income children of color have done to make significant improvements in student performance. She describes the leadership, policies, and systems that create the gains. Prior to writing this book about what makes districts effective, Dr. Chenowith had studied dozens of high-performing and rapidly improving individual schools that served large percentages of children of color and students from low-income families. She wrote the book about districts after coming to this realization:

“I began to explore the district role is school success and in school failure. I realized that you can fix schools all you want: if the district within which they reside are dysfunctional, the schools will not stay fixed. I came to this realization after seeing highly functional schools serving children of color and children from low-income homes deteriorate when their principals were replaced by new ones who did not continue the systems and culture that had made the schools successful. Watching them fall apart has driven home to me the importance of school principals. But it also points to the fact that schools are not perpetual motion machines. They are affected by district leaders who are responsible for the hiring and placement of principals. I have noticed that when districts leaders do not understand those things, the schools fall apart.” (Chenowith, 2021)

Interestingly, the highest-performing district in the United States is one that is not under any state education department. The Defense Department schools work like an ideal school district, as described in the New York Times:

“But the schools are inherently less political — big decisions come from headquarters — and therefore less tumultuous. Case in point: An academic overhaul that began in 2015 and has stuck ever since. Defense officials attribute recent growth in test scores partly to the overhaul, which was meant to raise the level of rigor expected of students.”

“Officials described a methodical rollout, one subject area at a time: New curriculum, teacher training, global coordination, so a fifth grader at Fort Moore learns similar material as a fifth grader in Kaiserslautern, Germany.”

“It took six years to finish carrying out the changes, longer than the average public school superintendent’s tenure. Logistical planning, including a predictable budget, “isn’t very sexy,” but it is one key to success, said Thomas M. Brady, the director of Defense Department schools since 2014.”

“Such strict structure is something Cicely Abron, an eighth-grade math teacher, rarely

---


experienced in nearly 20 years in public education. At Faith Middle School, she cannot supplement curriculum and must work off an approved list she receives from the Defense Department. The goal is to raise the floor for all students...something that Jason Dougal, president of the National Center on Education and the Economy, sees in top-performing countries like Finland and Singapore."

“American school districts often have an ‘all-star team mentality,’ Mr. Dougal said, relying on exceptional teachers and principals to get results.... But the most effective jurisdictions, he said, have a “systemic way of improving everybody on the team.” (Mervosh, 2023)"

In 2005, a large-scale study in California of elementary schools serving low-income students asked the question, “Why do some schools with similar students produce consistently better results?” In a follow-up study, the same question was asked about elementary schools serving low-income and English learner students. The findings were similar:

- **Prioritizing student achievement.** These schools had high expectations with clearly defined plans for improvement.
- **Implementing a coherent standards-based curriculum and instructional program.** These high-performing schools had curriculum alignment from grade to grade, prioritized and protected instructional time for math and language arts, ensured that the curriculum aligned to standards, including resources for English learners, especially for English language development, and implemented the curriculum faithfully. In addition, professional development and coaching supported the teachers in the implementation of their curriculum.
- **Using assessment data to improve student achievement and instruction.** Principals and district leaders used assessment data from multiple sources to evaluate teachers’ practices and design improvement plans. In addition, the principals in these schools used the data to follow up on students in need of additional support, and district administrators used the data to evaluate principals and assist their schools to improve.
- **Ensuring availability of instructional resources.** The principal in schools with consistently high achievement provided effective curriculum materials for core classes as well as supplemental materials and instruction for struggling students.

In addition to the above interrelated practices, the study findings indicated the key role played by principals and district leadership:

- **Principal leadership focused on improvement within an accountability-based reform.** Principals in these schools managed the improvement processes, cultivated and articulated the vision, engaged all stakeholders, valued the teachers, and regularly used assessment data to evaluate teacher practice, provide assistance to struggling students, and to provide added support to teachers.
- **District leadership, from the superintendent down, provided accountability and support.** They set clear expectations, including for subgroups, provided achievement data, and used the data to evaluate principal and teacher practices. They engaged all stakeholders in the work. They also made sure schools had math and language arts curricula aligned to the standards and that the resources were adequate for struggling students as well. This included adequate resources and practices for English learners.

---


Finally, in the later study specifically targeting schools with large proportions of English learners, these same findings were found to differentiate schools with consistently better performance than other schools with similar students. One additional finding emerged, however, the involvement of parents and families. In the schools with strong parent involvement, especially among the families of English learners, better results were seen.⁹

So, what can we learn from High-Performing Educators and their schools? We learn that certain practices lead to better achievement results even with the same combination of research-based components and even with the same populations. HPEs provide the evidence we need to actually implement the scientific evidence, whether for literacy, math, or other subjects. More studies of these HPEs will continue to inform educator knowledge of sound instructional practices. These studies must also take into consideration the population being taught by the educator:

- Educators who produce high performance with students from a high socio-economic background
- Educators who produce high performance with students from a low socio-economic background
- Educators who produce high performance with students who enter school with limited English proficiency
- Educators who produce high performance with students with disabilities

With such studies, we will continually enrich our understanding of teaching and learning.

---