

# **How Evidence-Based Reform Will Transform Research and Practice in Education**

**Robert E. Slavin**  
**Johns Hopkins University**  
300 E. Joppa Rd., #500  
Baltimore, MD 21286  
[rslavin@jhu.edu](mailto:rslavin@jhu.edu)  
410-616-2310

Adapted from an E. L. Thorndike Award address to Division 15 of the American Psychological Association, August 7, 2018

Published in the *Educational Psychologist*

### **Abstract**

Evidence-based reform in education refers to policies that enable or encourage the use of programs and practices proven to be effective in rigorous research. This article discusses the increasing role of evidence in educational policy, rapid growth in availability of proven approaches, and development of reviews of research to summarize the evidence. A highlight of evidence-based reform was the 2015 passage of the Every Student Succeeds Act (ESSA), which defines strong, moderate, and promising levels of evidence for educational programs and ties certain federal funding to use of proven approaches. To illustrate how coordinated use of proven approaches could substantially improve educational outcomes, the article proposes use of proven programs to populate each of Tiers 1, 2, and 3 in response to intervention (RTI) policies. This article is adapted from an address for the E.L. Thorndike Award for Distinguished Psychological Contributions to Education, August 7, 2018.

Why is it that educational research so rarely has an impact on the practice of education? In many areas of educational policy and practice, it is clear why evidence plays such a limited role, because politics, financial considerations, or long-standing traditions usually have more influence than evidence. Many areas of educational policy, such as governance, funding, and learning standards and assessments, will always remain in the realm of politics, finance, and tradition. However, there is one area of educational policy and practice in which educational research can have much more of an impact. This relates to decisions school, district, and state leaders make about the use of instructional programs, materials, technology, and professional development. Few decisions in education are made entirely independently of politics or ideology, but when a principal or superintendent wants a better elementary reading program or secondary math program, it is possible that they will want to know which, among various plausible alternatives, has the best evidence of effectiveness for the purposes they have in mind, the students they serve, the resources they have available, and the political and social context in which they operate.

As an analogy, consider the health system. There are ferocious political and economic forces driving state and federal leaders to favor or oppose the Affordable Care Act or single-payer healthcare plans, so research can play a limited role. However, the evidence for a new heart valve or breast cancer treatment or migraine medication is of value to practitioners and patients no matter where they are and no matter what their political or ideological beliefs.

Yet in education, even in the domain of educational innovations, materials, software, and professional development, in which evidence *could* play a major role, it has rarely done so, at least until recently. Studies find that principals and teachers pay little attention to research and rarely consult it to improve their practices (Dagenais et al., 2012; Helmsley-Brown & Sharp,

2003; Morrison, Ross, Corcoran, & Reid, 2014). Studies of how principals and districts make choices among programs and services find that rather than consulting evidence, most educators seek advice from friends, people they trust in similar schools or districts, or sales representatives, long before they consult evidence (Honig & Coburn, 2008; Morrison, Ross, Corcoran, & Reid, 2014). If they do consider evidence, it is often to ask whether a given program is *based on accepted principles* rather than whether the program itself has been *evaluated in comparison to a control group*.

In recent years, this evidence-free adoption process is starting to change. Recent policy changes have increased interest in research among educators, and our profession has an unprecedented opportunity to contribute to this movement. This article discusses these changes, their potential to enhance the practice of education and the stature of educational psychology, and some of the problems yet to be solved.

### **Evidence-Based Reform**

Why should evidence of effectiveness be a major criterion in the selection of educational products and services? The most important answer is that programs with a strong evidence base that are implemented as they were in the validating research are likely to produce better outcomes for children. Further, making evidence a basis for program adoption would put education into a virtuous cycle of innovation, evaluation, and progressive improvement like that which has transformed fields such as medicine, agriculture, and technology (Haskins, 2014; Kolada, 2013).

Evidence-based reform (Bridgeland & Orszag, 2014; Gueron & Rolston, 2013; Slavin, 2008, 2013, 2017; see H.R. 4174, the Evidence-Based Policy Act) refers to policies in which

educators and policymakers use evidence of effectiveness as a criterion for choosing educational programs, products, and practices. Evidence of effectiveness is defined as evidence from rigorous experiments in which students experiencing experimental programs are compared over significant periods of time (say, a semester or more) to those using traditional control methods in terms of gains on valid measures of achievement or other outcomes. Ideally, students, teachers, and/or schools are assigned at random to experimental or control treatments (Bucks & McGee, 2015; Gueron & Rolston, 2013), but at a minimum, experimental and control students are well matched at pretest on measures such as achievement and demographic variables. Note that this definition does not refer to simply disseminating information about generic principles of effective practice, which has not generally been found to make much of a difference in practices or outcomes (e.g., Hemsley-Brown & Sharp, 2003; Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008), but instead refers to reforms that support use of specific programs evaluated in comparison to control groups and found to be effective and replicable. This is an important distinction. Many school leaders and politicians shun specific programs, no matter how strong their evidence base, preferring to promote generic approaches. Yet research routinely finds that teachers have great difficulty translating generic principles into effective practice, and all too often fall back on the teaching methods they have always used unless they receive specific professional development and materials to help them implement proven models.

### **ESSA Evidence Standards**

Recently, a major federal policy initiative has created new possibilities for evidence-based reform in education. The Every Student Succeeds Act (ESSA), which replaced the No Child Left Behind Act of 2001 (NCLB) as the main federal education law in December,

2015, included definitions of strong, moderate, and promising levels of evidence supporting education programs. The definitions are as follows:

“(i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on—

“(I) **strong evidence** from at least one well-designed and well-implemented experimental study;

“(II) **moderate evidence** from at least one well-designed and well-implemented quasi-experimental study; or

“(III) **promising evidence** from at least one well-designed and well-implemented correlational study with statistical controls for selection bias.”

ESSA requires that low-achieving schools seeking school improvement grants adopt programs that meet one of the three definitions, and it provides other funding to use proven programs. This has led to enhanced interest in evidence throughout the U.S.

The ESSA evidence standards, as interpreted by Evidence for ESSA (2019) and the What Works Clearinghouse (2019; discussed later in this article), are both tough and lenient. The standards for the individual studies, especially in the top two categories, are tough. They require random assignment or careful matching of samples identified in advance, measures independent of developers or researchers, adequate sample sizes and durations, and appropriate analyses. However, since a program only needs a single study to qualify, the standards are not as tough as they may be in the future, when many more rigorous studies are likely to exist.

The evidence standards are not self-enforcing, but must be broadly supported if they are to have any impact. This article discusses what has to be done to progressively increase the role of evidence in education policy and practice, where we are in fall, 2019, on that agenda, and what remains to be done.

### **What Does it Take for Evidence to Make a Difference in Education?**

For evidence-based reform to prevail, three conditions must exist (see Slavin, 2017):

1. There must be a broad range of proven programs in key areas of education, at every grade level.
2. Trusted, impartial, educator-friendly reviews of research must be available, to enable educators and policymakers to know which specific programs and practices have been proven to work in rigorous evaluations.
3. Governmental policies must provide encouragement or incentives to schools and districts to adopt proven programs.

Over the past 15 years, there has been extraordinary progress in each of these areas. These changes have created a situation in which the potential for evidence-based reform in education, though still less than guaranteed, is greater than it has ever been.

### **Building the Research Base for Effective Programs**

Perhaps the most important requirement for evidence-based reform is a substantial set of programs and practices with clear evidence of effectiveness and replicability. Educators and policymakers must have a variety of programs they can choose among with confidence.

Fortunately, there is much progress in this area.

**Investing in Innovation (i3) and Education and Innovation Research (EIR).** A major advance in the creation of a large set of proven programs was the Investing in Innovation (i3) grant program

(<https://innovation.ed.gov/what-we-do/innovation/investing-in-innovation-i3/>), established in 2009 to fund the development, evaluation, and scale-up of proven programs at all levels of education, pre-K to 12, using a “tiered funding” strategy. Programs that already had strong evidence of effectiveness could qualify for large scale-up grants, those with some evidence could qualify for smaller validation grants, and those with a good but untested idea could receive development grants. By its final year in 2016, i3 had funded 171 projects and spent more than \$1.4 billion, much of it on third-party, usually randomized evaluations of these programs. Among the 171 projects, 11 were scale-up, 45 validation, and 115 development. i3 has now been effectively replaced by the Education and Innovation Research (EIR) grant program (<https://innovation.ed.gov/what-we-do/innovation/education-innovation-and-research-eir/>), which also utilizes a tiered evidence structure.

**Institute for Education Sciences (IES).** IES (<https://ies.ed.gov/>) has long supported development and rigorous evaluation of scalable programs, and in the process has greatly built up the national pool of capable evaluators on which i3 and EIR rely. IES also funds research and development in a tiered structure, with more of a focus on theory-building.

**Education Endowment Foundation (EEF; England).** A major effort patterned on i3 is taking place in England, where the current government allocated substantial funding to a private foundation to commission evaluations of promising programs for primary and secondary schools. This Education Endowment Foundation (EEF) (<https://educationendowmentfoundation.org.uk/>) has so far funded more than 180 third-party, mostly randomized evaluations of a wide variety of programs.

In addition to IES, i3, and EIR funding, the U. S. Department of Education has established funding for specific priority areas, such as technology (Interagency Education



Research Initiative), preschool (Preschool Curriculum Education Research), and secondary literacy (Striving Readers). All of these federal initiatives essentially require random assignment, and at the higher funding levels they are increasingly requiring use of third-party evaluators. Both requirements are also applied in almost all EEF studies in England. Outcome measures made up by researchers are also increasingly being disallowed as confirmatory outcomes. These requirements make it very unlikely that studies can be biased toward the experimental groups. They are tough to meet, and across all funders and programs, the majority of studies do not find positive outcomes. However, educators can place considerable confidence in those that do.

Collectively, the numbers of programs evaluated in rigorous experimental and quasi-experimental studies has substantially accelerated since 2003. Figure 1 shows the numbers of studies of elementary programs for struggling readers (Inns, Lake, Pellegrini, & Slavin, 2018), secondary reading (Baye, Lake, Inns, & Slavin, 2019), and elementary mathematics (Pellegrini, Lake, Inns, & Slavin, 2018), that would meet the “strong” or “moderate” standards of ESSA, as of 2019.

=====

FIGURE 1 HERE

=====

## **Reviewing What Works**

**What Works Clearinghouse (WWC; <https://ies.ed.gov/ncee/wwc/>).** Education policies cannot support evidence-based practice unless there is some agreement on which specific programs have clear evidence of effectiveness. A flagship initiative of IES is the What Works Clearinghouse (WWC), established to review research on practical programs in many

areas. The WWC now has reports on research in reading, math, programs for English language learners, and other areas, and it produces readable “practice guides” on several topics.

**Best-Evidence Encyclopedia (BEE; [www.bestevidence.org](http://www.bestevidence.org)).** The BEE is a free website created and routinely updated by our staff at Johns Hopkins University. It primarily contains full-scale academic reviews of research on the effectiveness of programs in specific areas. For example, it has reviews on elementary and secondary reading, reading for struggling readers and for English learners, elementary and secondary math, elementary and secondary science, and prekindergarten programs.

**Evidence for ESSA ([www.evidenceforessa.org](http://www.evidenceforessa.org)).** Evidence for ESSA is another website from our group at Johns Hopkins, but unlike the BEE it is targeted solely at educators, not academics, and is aligned with the ESSA evidence standards. When the ESSA evidence standards appeared, IES announced that the What Works Clearinghouse would not be revised to make it easy for users to find programs in the WWC that would meet ESSA evidence standards. For this reason, the group that created the BEE developed a new website solely committed to communicating to educational leaders information on individual programs that do or do not meet the ESSA definitions for strong, moderate, or promising evidence of effectiveness. Users need only click on a tile to obtain a program description, research summary, populations served in studies, costs, key citations, and other information.

At this writing, Evidence for ESSA has completed reviews of programs for reading and math, grades pre-K-12. About 100 programs meet the “strong,” “moderate,” or “promising” standards, with the largest number in the “strong” category. As of January, 2019, Evidence for ESSA has more than 70,000 unique users, adding about 1000 more each week.

In addition to aligning with the ESSA evidence standards, Evidence for ESSA adds to the WWC in several ways, primarily in that it is much faster. The WWC is very slow to add additional studies to its reviews, while Evidence for ESSA constantly scans publications, technical reports, and other sources of information and generally reviews them within a month of becoming aware of them.

### **Evidence-Based Policies**

Ultimately, it is not enough to have many research-proven programs and trusted reviews of research. Education lacks a tradition of looking to evidence for program adoption decisions, and without clear support from government, marketing, politics, and other factors will usually outweigh evidence (Morrison et al., 2014). Weiss et al. (2008) described failed efforts to encourage use of proven programs until government began to identify specific programs with clear evidence.

Making educators aware of “what works” is not enough. Governments need to provide incentive funding, or funds to enable schools to adopt proven programs. These “jump-start” funds attract educators’ attention, of course, and they help schools afford innovative models, but they also reduce risks to school leaders, who may otherwise be concerned that their school boards or management will take a dim view of using ordinary school funds on programs new to the school or district (no matter how many times the programs have been used and proven effective elsewhere).

Incentives work. In the late 1990s, the federal Comprehensive School Reform (CSR) program offered high-poverty Title I schools \$30,000 per year for up to three year if they would adopt whole-school reform models. About 8000 schools did so by the end of the program

(Comprehensive School Reform Quality Center, 2017). Note that \$30,000 (\$46,386 in 2018 dollars) would be about \$93 per student in a school of 500 students, or less than one percent of the national average cost per pupil of \$12,300. Yet this small investment was enough to encourage thousands of schools to adopt complex whole-school approaches. CSR did not require evidence of effectiveness, but reviews of research on CSR-funded programs identified several that had notable and consistently positive impacts on achievement (Borman et al., 2003; CSRQ, 2017).

Today, ESSA is offering incentives to certain schools to adopt programs that meet the top three evidence standards (strong, moderate, or promising). Most clearly, schools in the lowest 5% of their states in academic achievement can qualify for federal school improvement funding only if they agree to implement proven programs. For other federal grants, such as portions of Title II (professional development), schools can receive competitive preference points on their grant applications if they propose to use the funds to implement proven programs. At this writing, it is unclear what effect this will have, but all states are working to respond to this opportunity and to make school leaders aware of how to find out whether particular programs do or do not meet ESSA evidence standards. Some states have gone further than the federal law requires, for example restricting schools to the strong and moderate standards and extending use of the ESSA standards into other federal grants that flow through the states. According to a survey by Whiteboard Advisors, some states, such as Arizona, Louisiana, Nevada, Minnesota, North Carolina, and Rhode Island, are also using the ESSA standards as a requirement for certain *state* funding.

Another case of widespread application of incentives to adopt proven programs almost took place in England in 2018. The government there announced that quite substantial funding

was going to be made available to high-poverty schools to adopt proven programs. This Strategic School Improvement Fund (SSIF) expressed a preference for programs that met evidence requirements like those of ESSA. Thousands of applications were submitted, but at the last minute the government cancelled the program to use the money for other purposes. Still, the fact that SSIF and its emphasis on evidence existed, and that it aroused a great deal of interest among schools, is another indication that incentives matter and that they can drive expanded use of proven programs.

The inclusion of evidence standards within ESSA, and incentives to applicants for proposing to use programs that meet the ESSA strong, moderate, or promising criteria, could do much to provide a rationale for states, districts, charter organizations, and school leaders to learn the evidence standards. For many, understanding the standards could be important in obtaining grants or other benefits and this could result in better grant proposals that, when awarded and implemented, could improve outcomes for large numbers of students.

However, the current provisions in ESSA might only be the beginning. As educators get used to seeking evidence to satisfy the requirements of ESSA, which currently apply formally to only a small slice of the vast ESSA law, formal and informal changes could gradually expand the areas of ESSA in which the evidence standards matter. For example, government action could encourage use of proven programs (according to ESSA criteria) in Title I, Title II (professional development), IDEA (special education), and many other areas. Even if Congress did not specifically provide incentives for schools and districts to use proven programs, the current law might familiarize school, district, charter management, and state leaders with information on proven programs, which they might decide to apply on their own, without federal government mandates or incentives. State and local governments, which have long

traditions of creating lists of acceptable programs, materials, and software, might begin to include or even emphasize evidence as a criterion for use of state or district funds for certain purposes, especially in areas where evidence is readily available. As noted earlier, this is only happening in several states.

### **What Would Evidence-Based Reform Mean for Schools?**

Under federal, state, and/or local evidence-based reform policies, school staffs seeking to improve student outcomes would have a wide array of proven options, with resources to help them make and implement wise choices among proven, replicable programs that are ready for them to use. As part of the learning and adoption process they might carry out needs assessments to identify their priorities, and then attend regional effective-methods fairs, send delegations to visit nearby schools using the programs, and view videos and websites to see what the programs look like in operation. Ultimately, school staffs should have opportunities to vote by secret ballot, perhaps by a super-majority, to adopt a given program. Our Success for All program requires a positive vote of 80% of all teachers, plus the principal, to ensure buy-in. The staff who will be expected to implement a program should play a key role in selecting it (Slavin, 2017). Following the adoption decisions, school staffs work with program providers to plan and carry out high-quality implementation of the chosen program or programs.

### **Can Proven Programs Go to Scale?**

A vigorous effort to develop, promote, and support proven interventions could lead to widespread, measurable improvements in practices and outcomes in Title I schools (Cohen & Moffitt, 2009). However, developers of proven programs have to be willing and able to replicate

programs on a broad scale. Experience is clear that with encouragement and modest resources, many schools will adopt externally developed programs. The National Diffusion Network (NDN) of the 1980s reached thousands of schools with more than 500 programs, using state facilitators to help disseminate promising models. The federal Comprehensive School Reform (CSR) program of the late 1990s (Stringfield, Ross, & Smith, 1996) enabled more than 8,000 mostly Title I schools to adopt whole-school reform models (CSRQ, 2017). What was lacking in these earlier efforts was a strong evidence base for most of the adopted models, but that limitation is being rapidly solved by the i3, EIR, IES, and EEF investments, among others.

It is clear that developers can create and successfully evaluate replicable models, and that schools will eagerly embrace them if government offers them encouragement and resources. In fact, if policies support schools in adopting proven programs, then publishers, software developers, and other developers are sure to invest in creation and evaluation of innovative and effective programs.

Federal, state, and local government can also play a useful role in ensuring that evaluations are of the highest quality; evidence of effectiveness is easily available to school leaders; and incentives exist for states, districts, and schools to adopt proven models. Government can set quality standards for program providers to see that they offer sufficient professional development and other high-quality technical assistance to assist school leaders in effectively implementing and supporting whatever models they choose and make any needed adaptations to local circumstances. In essence, program providers should be asked to ensure that schools implementing their programs receive the materials and professional development provided in the studies that validated their model. Practicing educators and leaders need to be

involved at every step of this process from helping to define what programs are needed to voting on adoption to helping each other in the dissemination process.

No aspect of the process described in this article is impossible to achieve, and many parts of it are already going into place under ESSA or have already worked in prior incarnations. For example, the National Diffusion Network of the 1980s identified proven and promising programs of all sorts, and engaged “state facilitators” to help disseminate all programs on the list throughout their state. NDN introduced innovations in tens of thousands of schools. Comprehensive School Reform (CSR) of the late 1990s set standards for whole-school reforms and provided modest grants to schools willing to implement them, reaching more than 8000 schools (Borman et al., 2003), as noted earlier. Our Success for All program (Slavin, 2013) and some other CSR models have school staffs vote on whether to adopt particular programs, and this helps school staffs take ownership (Peurach, 2011). At its peak in the early 2000s, more than 1500 schools in 48 states used Success for All, and about 1000 do so today.

### **How Evidence-Based Reform Could Work in Practice**

To illustrate the potential impact of evidence-based reform, consider Response to Intervention (RTI), a part of the 2004 IDEA federal special education law. RTI suggests that schools use a three-tiered strategy to meet the needs of struggling students within the regular classroom, without assigning them to special education: initial teaching (Tier 1), modest supportive services for students having difficulty (Tier 2), and only for students having continuing difficulties despite Tier 2 services, intensive instruction (Tier 3) to avoid special education. However, studies of RTI do not find it to be effective (Balu et al., 2015), because RTI principles are rarely fully implemented (Fuchs & Fuchs, 2017; Ruffini et al., 2016). Perhaps



outcomes of RTI could be greatly improved if proven strategies were specified for each level. In fact, the ESSA law now recommends that the practices used at each tier level should be proven to be effective, according to ESSA evidence standards.

There is perhaps no area in the practice of education in which there is stronger evidence than for programs for struggling learners. Evidence for ESSA and the WWC provide evidence to support a wide array of reading and mathematics approaches for struggling students. Most involve one-to-one or one-to-small group tutoring, provided by either certified teachers or well-qualified teaching assistants using structured approaches (Baye, Lake, Inns, & Slavin, 2019; Inns et al., 2018; Pellegrini et al., 2018; Wanzek et al., 2013, 2016, 2017). There are also whole-class and whole-school approaches with positive impacts on the reading or math achievement of disadvantaged or struggling readers (Inns et al., 2018; Pellegrini et al., 2018).

Schools implementing RTI might be encouraged to select proven approaches for each tier: proven whole class/whole school approaches for Tier 1, proven one-to-small group tutoring by trained teaching assistants for Tier 2, and proven one-to-one tutoring by teaching assistants or certified teachers for Tier 3. In this way, the students would be virtually assured of success, because each of the instructional approaches they receive is known to be effective if well implemented. The result could be significant improvement in the achievement of students at risk, reductions in special education placements and retentions, and reductions in achievement gaps among students, especially by bringing up the low end of the distribution (Slavin, Inns, Pellegrini, & Lake, 2019). Tables 1 and 2 (from Slavin, Inns, Pellegrini, & Lake, 2019) list programs that meet the “strong” or “moderate” levels of the ESSA standards and had effect sizes of +0.20 or more, for Tier 1, Tier 2, and Tier 3 for reading (Table 1) and mathematics (Table 2).

=====

## TABLES 1 AND 2 HERE

=====

**Criticisms and Problems**

As with any policy, evidence-based reform has its detractors and problems. The most important of these are discussed in this section.

**Do results of experiments generalize?** Many skeptics of evidence-based reform express concern that studies that took place in one set of schools will not generalize to others, or to other types of schools (see, for example, Bryk, Gomez, Grunow, & LeMahieu, 2015). This is a longstanding concern in experimental research design (see Cornfield & Tukey, 1956). The problem is that it is almost never practically possible to sample experimental and control groups from within an entire population of interest, so there is always a possibility that a treatment that worked in one set of schools will not work in another.

One development in modern times greatly reduces this problem (and some others). This is the universal acceptance among experimental methodologists of a requirement that experiments carried out at the cluster level (e.g., schools or classrooms) must be *analyzed* at the cluster level (for example, using hierarchical linear modeling; Raudenbush & Bryk, 2002). For adequate statistical power to detect meaningful differences, this generally requires 40 to 50 clusters, usually schools. The What Works Clearinghouse (2019) and Evidence for ESSA (2019) both require analysis at the level of clustering, for example. The clustering requirement was designed to reduce bias, for reasons of internal validity, but it also happens to greatly improve external validity (generalization) at the same time. That is because in an experiment with, say, 25 schools in the experimental group and 25 controls, it is unlikely that unusual characteristics of individual schools (such as outstanding principals or staffs, for example) will restrict

generalization to other schools. Use of random assignment of schools, required for the “strong” rating in ESSA, also diminishes the possibility that local factors will restrict generalization, because it ensures that the experimental group will be representative of the whole very large sample.

Even with large numbers of experimental and control schools, there may still be a problem with generalization to all demographic types of schools. For example, a large program evaluation that took place in Chicago and Baltimore might not generalize to Los Angeles or Phoenix, much less to suburban or rural locations. The WWC (2019) and Evidence for ESSA (2019) present information to make clear the nature of the schools and populations involved in studies, so at least readers themselves can infer where generalization might be most or least likely. For example, schools in Baltimore might be most interested in programs proven in Philadelphia, Washington, Detroit, or Atlanta, while schools in rural Appalachia might be most interested in programs proven in Appalachia, or other rural places.

Advocates of research-practice partnerships, such as Bryk et al. (2015), argue that school district staffs should develop solutions for their own particular needs and situations, perhaps working with local academics. New sources of ideas for interventions from educators working at the front lines are certainly welcome, and if such studies follow accepted standards of experimental research, they contribute to evidence-based reform to the same degree as any other high-quality studies, except that a deliberate focus on a single district may, ironically, inhibit generalization to schools and districts elsewhere.

The solution to generalization is successful replication in a wide variety of settings. A recent initiative announced by the Institute for Education Sciences proposed the creation of a new funding stream to support studies of replications of successful programs in a broader range

of schools and districts

(<https://ies.ed.gov/blogs/research/post/building-evidence-changes-to-the-ies-goal-structure-for-fy-2019>).

**Experiments report effects of programs on average. What if the average covers important differences for various subgroups?** If a program is found to have positive effects on all students in an experimental group, compared to a control group, it is possible that the effect is due to higher performance for some groups than for others. Reports of experiments often, but not always, report effects for different subgroups. When they do, it is typical that effects are greater or less for some groups than for others, but it is rare that a program found to be effective on average would have zero impact, or even a negative impact, for any subgroup with significant representation in the schools. For example, assume that an evaluation of a schoolwide intervention found an average effect size of +0.20, a possible and respectable outcome. Half of students are members of ethnic group A and half of ethnic group B. If the results for ethnic group B were zero, the effect size for ethnic group A would have to be +0.40, a highly unlikely outcome for any treatment other than tutoring.

**Will effects of proven programs replicate?** Beyond the problem of generalization, there is a broader problem that positive findings from early experiments often do not replicate in later experiments. Often, this happens because the early experiments involve small numbers of schools, teachers, or students, allowing researchers to provide exceptional levels of service to the experimental group. Early experiments are more likely to be done by the developers (rather than third party evaluators), often using measures made up by the developers or researchers. Evidence for ESSA (2019) reduces the impact of some of these factors by excluding developer-made measures and weighting by sample size, for example, but there are still

problems of non-replication. Over time, however, government and other funders are supporting third-party evaluations and replications of promising practices, using procedures that do not add bias, and the programs still found to be effective under these increasingly stringent circumstances are more likely to replicate under realistic conditions.

**Do low-resource schools have capacity to implement proven programs?** Some critics wonder whether under-resourced, high-poverty schools have the capacity to implement proven programs. This question can be divided into two. First, do such schools have the financial resources to implement proven programs? The answer is probably no. That is why evidence-based reform envisions government or other resources being granted to schools promising to implement proven models.

However, even with financial resources needed to implement proven programs, can high-poverty schools do so? Having actually implemented classroom- and whole-school reform models in high-poverty schools for more than 40 years in both research and dissemination (see Slavin, 2017; Slavin & Madden, 2013), I have no doubt that these schools can successfully implement proven programs, with the assistance of experienced and capable external professional development, well-designed materials, and strong local leadership. The successful studies of proven programs usually take place in high-poverty schools, so they would not be “proven” if they could not succeed in these schools. Over the years, our Success for All whole-school reform approach has worked with perhaps 3000 high-poverty Title I schools in almost every state and four other countries. These are difficult schools to work with, not because of their principals and teachers, who are generally capable and deeply committed to children, but to constant turmoil in central offices, funding cutbacks, and changes in state policies (Peurach, 2011). Solving these political problems and providing adequate resources would be

highly beneficial to high-poverty schools, but while waiting for this to happen, the best our field can do is to provide schools with the best, most effective programs we can.

**Might programs with positive short-term effects maintain their outcomes in the long-run?** It is true that most studies that meet ESSA evidence standards are validated in studies of one year's duration or less. Our Evidence for ESSA website requires a minimum study duration of 12 weeks, which in effect means that almost all studies have a duration of from one semester to one to three years. Research on achievement outcomes rarely documents long-term outcomes (though there are exceptions, such as Borman & Hewes, 2003, Hurry & Sylva, 2007). However, students go to school for many years, and a positive schooling experience depends not on a one-time intense intervention, whose impacts are unlikely to last, but to schools' use of well-implemented, proven instructional models year after year. There are never any guarantees, but it seems likely that students who receive proven, well-implemented instructional methods every year from preschool to 12<sup>th</sup> grade will have a considerable cumulative advantage over those who do not.

**Does evidence-based reform privilege experimental research to the detriment of other types of research?** Evidence-based reform does require use of quantitative, experimental research. Virtually all evaluations funded by the U.S. Department of Education and by the EEF in England require mixed-methods designs, with qualitative research to inform readers about what was going on in experimental and control schools. However, the emphasis in these evaluations and in policy based on this research is clearly on the experimental evidence.

This emphasis is not intended to privilege one type of research over another. It is, however, an exercise of the key principle of research design that must appear prominently in every research methods text ever written: The design should be appropriate to the question being

asked. In the case of evidence-based reform, the main question being asked is whether it is likely that if schools or teachers use a given program, their students will perform better on measurable achievement outcomes. This question demands a comparison of students whose classes or schools received a given treatment to those whose classes or schools did not. For questions of this type, experiments are clearly optimal. Experimental designs are not the only path to wisdom about teaching and learning, and our field should value many ways of understanding issues of importance to theory, practice, and policy. But when we want to know “what works,” we must ask, “Compared to what?”, and we must use valid measures of importance in educational contexts.

Over time, additional research should investigate further the degree to which various types of programs are replicable and have outcomes that generalize to other situations and populations, and which outcomes maintain over long periods of time. Development and evaluation of new programs are sure to add new ways to solve longstanding problems. The research that exists today is just an early indication of what may be possible in the future, assuming continued funding and continued efforts on the part of researchers and developers.

### **How Will Educational Psychology Change in Light of Evidence-Based Reform?**

Evidence-based reform offers educational psychologists of all backgrounds new opportunities to do consequential research likely to affect policy and benefit children. If it prevails, evidence-based reform will greatly increase interest in research among educators and policy-makers at all levels. The impact on educational researchers who do large scale evaluations of practical programs is apparent, but the impact of evidence-based reform goes further than that. For example, researchers focused on theory-building are needed to identify

variables worth incorporating in educational treatment, and they are essential in helping make sense of findings of experiments. Qualitative research is needed to describe current practices and identify problems in need of solutions, and in experiments it is needed to characterize what is happening in experimental treatments, to identify problems and suggest alternatives.

Correlational research is needed to learn from policies and practices already in existence, to suggest variables worth testing in experiments.

### **Consequences of Evidence-Based Reform**

The consequences of evidence-based reform could be profound. If federal policies began to favor specific programs with clear, replicated evidence from rigorous research, publishers, software developers, university researchers, and entrepreneurs of all kinds would have incentives to engage in serious development and evaluation efforts. A current example of this is school improvement. Low-achieving schools seeking this funding are required to select programs that meet the strong, moderate, or promising levels of evidence defined in the ESSA evidence standards, as noted earlier. Government and other funders, seeing the immediate and potential impacts of research and development, might provide substantially greater funding for these activities. Developers would have a reason to invest in more innovative strategies, knowing that if their programs turn out to be effective in rigorous evaluations, they will be likely to be successful in the marketplace. Most importantly, children would be more likely to receive proven programs, which would be more likely than those prevalent today to help them succeed in school.

Evidence-based reform could finally apply to education the process that led to dramatic developments in medicine, agriculture, and technology in the 20th century, and continues today,



where solutions that meet rigorous evidence standards supersede less effective products, and a vast R&D enterprise works to continuously improve on the best we have available today.

The winners in this would be millions of children, especially those who are least well served by the current system, the teachers and administrators who yearn for more effective tools to help them do their job well, and the whole society, which would come to expect progress in education as confidently as it currently expects progress in other fields. Education research would gain the respect and the resources it has never had.

Recent developments in research and policy make it possible to put us on the road to genuine reform. Under ESSA, the federal government has made crucial initial steps in making proven programs more widely available. Further developments at the federal level, and an active embrace of evidence-based reform at the state and local levels, as well as within educational psychology itself, will expand the impact of evidence, creating a virtuous cycle of research, dissemination, and impact for America's students.

### References

- Amendum, S. J., Vernon-Feagans, L., & Ginsberg, M. C. (2011). The effectiveness of a technologically facilitated classroom-based early reading intervention: The Targeted Reading Intervention. *The Elementary School Journal*, *112*(1), 107–131.  
<https://doi.org/10.1086/660684>.
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, *35*(4), 494–519. <http://dx.doi.org/10.1598/RRQ.35.4.3>
- Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2105). *Evaluation of response to intervention practices for elementary school reading*. Washington, DC: U.S. Department of Education, Institute for Education Sciences, NCEE 2016-400.
- Baye, A., Lake, C., Inns, A., & Slavin, R. (2019). Effective reading programs for secondary students. *Reading Research Quarterly*, *54*, 133-166.
- 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*(3), 444–461. <http://dx.doi.org/10.1037/0022-0663.96.3.444>
- Borman, G., & Hewes, G. (2003). Long-term effects and cost effectiveness of Success for All. *Educational Evaluation and Policy Analysis*, *24* (2), 243-266.  
[www.jstor.org/stable/3594118](http://www.jstor.org/stable/3594118)
- Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *13* (2), 125-230.

- Bridgeland, J., & Orszag, P. (2013). Can government play moneyball? At <http://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389/>
- Brown-Chidsey, R., & Bickford, K. (2015). *Practical handbook of multi-tiered systems of support*. New York, NY: Guilford Press.
- Bryk, A. S., Gomez, L.M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Bucks, S., & McGee, J. (2015). *Why government needs more randomized control trials: Refuting the myths*. Houston, TX: Laura and John Arnold Foundation.
- Cohen, D. K., & Moffitt, S. L. (2009). *The ordeal of equality: Did federal regulation fix the schools?* Cambridge, MA: Harvard University Press.
- Comprehensive School Reform Quality Center (CSRQ) (2017). *National Center for Comprehensive School Reform*. At [www.csrq.org](http://www.csrq.org).
- Cornfield, J., & Tukey, J. W. (1956). Average values of mean squares in factorials. *Annals of Mathematical Statistics*, 27, 907-959.
- Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., ... Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, 6(1), 1–23. <http://dx.doi.org/10.1080/19345747.2012.706694>
- Dagenais, C., Lysenko, L., Abrami, P., Bernard, R., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of

empirical research. *Evidence & Policy*, 8 (3), 285-309.

<http://dx.doi.org/10.1332/174426412X654031>

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, A. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, 44(2), 414–448.

<http://dx.doi.org/10.3102/0002831207302175>

Evidence for ESSA (2019). At [www.evidenceforessa.org](http://www.evidenceforessa.org)

Fuchs, D., & Fuchs, L.S. (2017). Critique of the National Evaluation of Response to Intervention: A case for simpler frameworks. *Exceptional Children*, 83 (3), 1-14.

<https://doi.org/10.1177/0014402917693580>

Fuchs, L. I., Schumacher, R. F., Long, J., Jessica, N., Malone, A. S., Amber, W., & ... Changas, P. (2016). Effects of intervention to improve at-risk fourth graders' understanding, calculations, and word problems with fractions. *Elementary School Journal*, 116(4), 625-651. <http://dx.doi.org/10.1086/686303>

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., ... & Bryant, J. D. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology*, 105(1), 58.

<http://dx.doi.org/10.1037/a0030127>

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties.

*Learning and Individual Differences*, 20(2), 89-100.

<http://dx.doi.org/10.1016/j.lindif.2009.09.003>

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015).

Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. *American Educational Research Journal*, 52(3), 516-546.

<http://dx.doi.org/10.3102/0002831214565787>

Gueron, J. M., & Rolston, H. (2013). *Fighting for reliable evidence*. New York, NY: Russell Sage.

Haskins, R. (2014, Dec. 31). Social programs that work. *The New York Times*.

Hemsley-Brown, J., & Sharp, C. (2003). The use of research to improve professional practice: A systematic review of literature. *Oxford Review of Education*, 29(4), 449-470.

<https://doi.org/10.1080/0305498032000153025>

Honig, M., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy*, 22 (4), 578-608.

<https://doi.org/10.1177/0895904807307067>

Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, 30(3), 227-248. <http://dx.doi.org/10.1111/j.1467-9817.2007.00338.x>

Inns, A., Lake, C., Pellegrini, M., & Slavin, R. (2018). *Effective programs for struggling readers: A best-evidence synthesis*. Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Jones, C. J. (2015). *The results of a randomized control trial evaluation of the SPARK literacy program*. Retrieved 01/02/19 from

<https://pantherfile.uwm.edu/groups/uwm/SOE/docs/SREed/second-cohort-final-report.pdf>

Kolada, G. (2013, September 2). Guesses and hype give way to data in study of education. *The New York Times*. Retrieved from

[http://www.nytimes.com/2013/09/03/science/applying-new-rigor-in-studying-education.html?\\_rD0](http://www.nytimes.com/2013/09/03/science/applying-new-rigor-in-studying-education.html?_rD0)

Lord, P., Bradshaw, S., Stevens, E. and Styles, B. (2015). *Perry Beeches coaching programme: Evaluation report and executive summary*. London, England: Education Endowment Foundation.

Madden, N. A., & Slavin, R. E. (2017). Evaluations of technology-assisted small-group tutoring for struggling readers. *Reading & Writing Quarterly*, 33 (4), 1-8.  
<http://dx.doi.org/10.1080/10573569.2016.1255577>

Mathes, P. G., Torgeson, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal*, 38(2), 371–410.  
<http://dx.doi.org/10.3102/00028312038002371>

May, H., Sirinades, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. Newark, DE: University of Delaware, Center for Research in Education and Social Policy.

Merrell, C., & Kasim, A. (2015). *Butterfly Phonics: Evaluation report and executive summary*. London, England: Educational Endowment Foundation.

Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, 92 (4), 681-693.  
<http://dx.doi.org/10.1037/0022-0663.92.4.681>

Morrison, J. R., Ross, S. M., Corcoran, R. P., & Reid, A. J. (2014). *Fostering market efficiency in K-12 ed-tech procurement*. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education.

- Pellegrini, M., Inns, A., & Slavin, R. (2018). *Effective programs in elementary mathematics: A best-evidence synthesis*. Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.
- Peurach, D.J. (2011). *Seeing complexity in public education: Problems, possibilities, and Success for All*. New York, New York: Oxford University Press.
- Ruffini, S.J., Miskell, R., Lindsay, J., McInerney, M., & Waite, W. (2016). *Measuring the implementation fidelity of the Response to Intervention framework in the Milwaukee Public Schools (Report No. REL 2017-192)*. Washington, DC: U.S. Department of Education.
- Rutt, S., Easton, C., & Stacey, O. (2014). *Catch Up<sup>®</sup> Numeracy: Evaluation report and executive summary*. London, England: Education Endowment Foundation
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods (Second Edition)*. Thousand Oaks, CA: Sage Publications, Inc.
- Sibieta, L. (2016). *REACH: Evaluation report and executive summary*. London, England: Education Endowment Foundation.
- Slavin, R.E. (2013). Overcoming the four barriers to evidence-based education. *Education Week*, 32 (29), 24.
- Slavin, R. E. (2017). Evidence-based reform in education. *Journal of Education for Students Placed at Risk*, 22(3), 178-184. <https://doi.org/10.1080/10824669.2017.1334560>
- Slavin, R.E., Inns, A., Pellegrini, M., & Lake, C. (2019). *Response to Proven Instruction (RTPI): Enabling struggling learners*. Submitted for publication.

- Slavin, R.E., & Madden, N.A. (2013). Success for All at 27: New developments in whole-school reform. *Journal of Education for Students Placed at Risk*, 18 (3), 169-176. <https://doi.org/10.1080/10824669.2013.862095>
- Smith, J. L. M., Nelson, N. J., Fien, H., Smolkowski, K., Kosty, D., & Baker, S. K. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal*, 116(4), 549-573. <http://dx.doi.org/10.1086/686249>
- Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal*, 50(2), 397-428. <http://dx.doi.org/10.3102/0002831212469045>
- Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative Integrated Reading and Composition: Two field experiments. *Reading Research Quarterly*, 22, 433–454. <http://dx.doi.org/10.2307/747701>
- Stringfield, S., Ross, S., & Smith, L. (Eds) (1996). *Bold plans for school restructuring: The New American Schools*. Mahwah, NJ: Erlbaum.
- Styers, M. & Baird-Wilkerson, S. (2011). *A final report for the evaluation of Pearson's focusMATH Program*. Charlottesville, VA: Magnolia Consulting.
- Torgerson, C. J., Wiggins, A., Torgerson, D., Ainsworth, H., & Hewitt, C. (2013). Every Child Counts: Testing policy effectiveness using a randomised controlled trial, designed, conducted and reported to CONSORT standards. *Research In Mathematics Education*, 15(2), 141–153. doi:10.1080/14794802.2013.797746
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for



- dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60(1), 40–56.  
<http://dx.doi.org/10.1007/s11881-009-0032-y>
- Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1997). The prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading*, 1, 217-234. [http://dx.doi.org/10.1207/s1532799xssr0103\\_3](http://dx.doi.org/10.1207/s1532799xssr0103_3)
- Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading*, 13(5), 383–425.  
<http://dx.doi.org/10.1080/10888430903162894>
- Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15(6), 471–497.  
<http://dx.doi.org/10.1080/10888438.2010.501091>
- Wanzek, J., Petscher, Y., Al-Otaiba, S., Rivas, B., Jones, F., Kent, S., Schatschneider, C., & Mehta, P. (2017). Effects of a year-long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology* 109 (8), 1103–19. <https://doi.org/10.1037/edu0000184>.
- Wanzek, J., Vaughn, S., Scammacca, N., Metz, K., Murray, C., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research*, 83 (2), 163-195.  
<https://doi.org/10.3102/0034654313477212>
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades k-3. *Educational Psychology Review* 28 (3), 551–76. DOI:[10.1007/s10648-015-9321-7](http://dx.doi.org/10.1007/s10648-015-9321-7)

Weiss, C., Murphy-Graham, E., Petrosino, A., & Gandhi, A. (2008). The fairy godmother—and her warts: Making the dream of evidence-based policy come true. *American Journal of Evaluation*, 29(1), 29–47. <https://doi.org/10.1177/1098214007313742>

What Works Clearinghouse (2019). Retrieved April 8, 2019 from <https://ies.ed.gov/ncee/wwc/>.

**Table 1: Reading Programs Meeting ESSA Standards for Strong and Moderate Evidence**

<b>Program</b>	<b>Grades Validated</b>	<b>Group Size</b>	<b>Number of Studies</b>	<b>Mean Effect Size</b>	<b>ESSA Rating</b>
<b>Multi-Tier Approaches (Schools/Classes Plus Tutoring)</b>					
Success for All (Slavin & Madden, 2013)	K-2	Schools	5	+0.41	Strong
Enhanced Core Reading Instruction (Smith et al., 2016)	1	Schools	1	+0.22	Strong
<b>Classroom Approaches (Tier 1)</b>					
Cooperative Reading and Composition (Stevens et al., 1987)	3-6	Classes	4	+0.19	Moderate
PALS (Mathes, Torgesen & Allor, 2001)	1	Classes	1	+0.58	Moderate
<b>One-to-Small Group Tutoring: Teachers (Tier 2)</b>					
Lindamood (LIPS) (Torgesen, Wagner, & Rashotte, 1997)	1	1-3	1	+0.64	Strong
Read, Write, & Type (Torgesen et al., 2010)	1	1-3	1	+0.42	Strong
Butterfly Phonics (Merrill & Kasim, 2015)	6	1-6/8	1	+0.30	Strong
<b>One-To-Small Group Tutoring: Teaching Assistants (Tier 2)</b>					
Quick Reads (Vadasy & Sanders, 2009)	2-3	1-3	2	+0.22	Strong
Early Reading Intervention (Coyne et al., 2013)	K	1-3/5	1	+0.31	Strong
Lightning Squad (Madden & Slavin, 2017)	1-3	1-4/6	3	+0.34	Strong
<b>One-To-One Tutoring: Teachers (Tier 3)</b>					
Reading Recovery (May et al., 2016)	1	1-1	4	+0.42	Strong
Targeted Reading Intervention (TRI) (Amendum et al., 2011)	K-1	1-1	2	+0.52	Strong

Early Steps/Next Steps (Morris, Tyner, & Perney, 2000)	1	1-1	1	+0.86	Moderate
Lindamood (LIPS) (Torgesen et al., 2010)	K-2	1-1	1	+0.69	Strong
Intensive Reading Remediation (Blachman et al., 2004)	2-3	1-1	1	+0.85	Strong
<b>One-To-One Tutoring: Teaching Assistants (Tier 3)</b>					
Sound Partners (Vadasy & Sanders, 2011)	K-1	1-1	2	+0.43	Strong
Reading Rescue (Ehri et al., 2007)	1	1-1	1	+0.81	Moderate
Reach (Sibieta, 2016)	6-7	1-1	1	+0.42	Strong
Perry Beeches (Lord et al., 2015)	6	1-1	1	+0.36	Strong
<b>One-To-One Tutoring: Paid Volunteers (Tier 3)</b>					
SPARK Literacy (Jones, 2015)	K-2	1-1	1	+0.51	Strong
SMART (Baker, Gersten, & Keating, 2000)	1-2	1-1	1	+0.42	Strong

Adapted from Slavin, Inns, Pellegrini, & Lake (2019).

**Table 2: Mathematics Programs Meeting ESSA Standards For Strong and Moderate Evidence**

<b><u>Program</u></b>	<b>Grades Validated</b>	<b>Group Size</b>	<b>Number of Studies</b>	<b>Mean Effect Size</b>	<b>ESSA Rating</b>
<b>One-To-Small Group Tutoring: Teachers (Tier 2)</b>					
Number Rockets (Gersten et al., 2015)	1	1-2/3	1	+0.34	Strong
<b>On-To-Small Group Tutoring: Teaching Assistants (Tier 2)</b>					
Fraction Face-Off! (Fuchs et al., 2016)	4	1-2	2	+0.51	Strong
ROOTS (Doabler et al., 2016)	1	1-2/5	2	+0.24	Strong
FocusMath (Styers & Baird-Wilkerson, 2011)	3,5	1-6/8	1	+0.24	Strong
<b>One-To-One Tutoring: Teachers (Tier 3)</b>					
Numbers Count (Torgerson et al., 2013)	1-2	1-1	1	+0.33	Strong
Math Recovery (Smith et al., 2013)	1	1-1	1	+0.24	Moderate
<b>One-To-One Tutoring: Teaching Assistants (Tier 3)</b>					
Catch-Up Numeracy (Rutt et al., 2014)	1-5	1-1	1	+0.21	Strong
Galaxy Math (Fuchs et al., 2013)	1	1-1	1	+0.25	Strong
Pirate Math (Fuchs et al., 2010)	3	1-1	1	+0.37	Strong

Adapted from Slavin, Inns, Pellegrini, & Lake (2019).



