Perspectives on Behavior Science
https://doi.org/10.1007/s40614-021-00287-x

SPECIAL ISSUE: DIRECT INSTRUCTION

Building a More Effective, Equitable, and Compassionate Educational System: The Role of Direct Instruction

Jean Stockard[1]

Accepted: 5 April 2021/

Abstract

In a recent book, Anthony Biglan describes how strong social research can be used to build a compassionate and more caring society that promotes the well-being of all. This article asserts that a strong educational system needs to be part of this transformation and that widespread use of Direct Instruction (DI) could be key in the process. Analysis of the underlying theory, development, and use of DI describes the way it is based on careful developmental research. It promotes effective and efficient learning while embodying respect for students and teachers. The results of a recent large meta-analysis of research on DI's effectiveness show it is more effective than other educational approaches, with effect sizes that surpass the effect associated with the difference in achievement of students from lower income and other homes. Alternative approaches to educational change are reviewed and it is suggested that DI is a more effective and efficient method of improving student success. Powerful actors within the educational establishment have expressed opposition to DI and have worked to hide evidence of its effectiveness. This paper identifies other social actors who could work together to counter the resistance to DI and build an educational system that promotes the well-being of all.

Keywords Direct Instruction · educational equality · achievement · meta-analysis

In a recent book, the behavioral psychologist Anthony Biglan calls for researchers, policy makers, businesses, and the general public to work together, guided by research findings, to build a caring, compassionate, and nurturing society—a society that promotes the well-being of all individuals as well as the nation as a whole (Biglan, 2020). He examines a wide range of issues such as health, criminal justice, and the political world, showing how strong research findings, coupled with collective action,

✉ Jean Stockard
jeans@uoregon.edu

[1] University of Oregon, Eugene, OR, USA

can counter seemingly difficult problems. In this article, I echo Biglan's call to use research to promote a better society, focusing on our educational system. Decades of research have shown how education promotes individual well-being in virtually every aspect of adult life: enhancing health and longevity, economic security, strong civic engagement and the ability to resist oppression and inequalities. On a societal level, education is key to economic development and maintaining a just democratic society. This association is strongest when educational advantages accrue equally to all. Poor education can be seen as a root cause of many social problems.

A large body of research documents low levels of academic achievement for students in the United States, both in comparison to established benchmarks of proficiency and to other countries (Programme for International Student Assessment [PISA], 2010, 2019). Results from the National Assessment of Educational Progress (NAEP), which has gathered data since 1969, indicate the extent of the problem. In 2019, the most recent year for which data are available, only 35% of fourth graders had reading scores that fell at or above the level termed "proficient," indicating "solid academic performance." A similar percentage (34%) fell below the "basic" level, with reading skills that failed to meet even "partial mastery" of essential knowledge and skills. The situation was worse for students from low-income homes, with almost half (47%) below the basic level. Similar outcomes appeared for math and in other grades (NAEP, n.d.).

In this article I assert that Direct Instruction (DI), the system of instruction developed by Siegfried Engelmann and his colleagues (see below), could—and should—be a key element in transforming our educational system to address the issue of low achievement. I also suggest that it could be a key element in building the more compassionate society that Biglan envisions. In the first section I describe the careful, research-based, manner in which DI programs are designed and implemented and the way in which they embody respect for students and teachers. In the second section I summarize research on DI's effectiveness, showing its potential for promoting real and lasting change in students' achievement and self-confidence. In the third section I briefly describe two alternative perspectives regarding educational change and explain the ways in which their concerns intersect with, and are better met by, DI. In the final section I discuss the nature and sources of opposition to DI within education. Then, building on Biglan's call for social actors to work together for change, I describe potential members of a coalition that could promote the use of DI to build a stronger and more equal educational system and society.

DI's Assumptions, Development, and Use

Direct Instruction, or DI (capitalized), refers to the system of instructional practices developed by Siegfried Engelmann and colleagues in the 1960s, coinciding with the War on Poverty and passage of the 1964 Civil Rights Act. In subsequent years they developed curricular programs that span most academic subjects and are designed for students from preschool to higher grades. DI is distinct from "direct instruction" (uncapitalized and sometimes called "little di"). The latter term was coined to describe individual characteristics of instruction found to be correlated with higher student achievement, such as teacher engagement or active student responding (Rosenshine,

1971, 1978). DI incorporates most, if not all, of the elements listed within "little di" analyses, but does so in a systematic and organized manner based on extensive research. This is what makes DI unique—its underlying theory, the careful research involved in developing the programs, and the way they are used in schools.

The paragraphs below briefly describe these elements, but readers are cautioned that this overview only scratches the surface. The literature related to DI's development and its use in classrooms is extraordinarily large, running to hundreds of publications (see National Institute for Direct Instruction [NIFDI], 2017). Accessible descriptions include Barbash (2012), Engelmann (2014a, 2014b), Stockard et al. (2020), and Wood (2014a). More detailed, technical expositions are in Engelmann (2018), Engelmann and Carnine (1991, 2011), Engelmann and Colvin (2006), and Engelmann and Steely (2004).

Theory and Program Development

The most basic assumption underlying DI is that, with appropriate instruction, all students can succeed and become confident learners. It is important to note that the onus for learning is not on the student, but on the instruction. If students do not learn it is because the instruction has not been appropriate. That is, the instruction has to be *effective*. In addition, instruction must be *efficient*. It should be designed so that time isn't wasted, so that students learn as much as they can in the shortest amount of time. And the development of the programs needs to be research-based. Each element of the programs, as well as the assumptions on which they are based, should be tested and validated through research.

DI assumes that students are inherently logical and make inferences from the world around them. They learn best when they are given examples that are as clear as possible—so clear that they imply only one possible conclusion. If the examples result in erroneous conclusions, later learning is more difficult, as students become frustrated or discouraged or lack the knowledge or skills needed for later learning. Thus, erroneous conclusions or mistakes should be quickly corrected. The examples must also be carefully sequenced, forming a "stair step" type of progression through learning. Mastery of each step is seen as essential. We learn new material more quickly when we have thoroughly mastered prerequisite knowledge and skills, when the information becomes part of what cognitive psychologists call "long-term memory" (Engelmann, 2014b; Stockard et al., 2020, pp. 10–13).

The development of DI instructional programs begins with careful, in-depth analysis of a subject, determining each element that needs to be taught and the way in which the elements are interconnected and build upon each other. The developers create carefully worded examples to teach each element. Their wording is tested in-house and then with small groups of students to make sure that only one type of interpretation is possible. The ordering of the elements is carefully designed to make sure that it is logical and that the steps of learning are as even and efficient as possible. This is not a simple task. For instance, Engelmann found that English has a total of at least 44 different sounds and 220 spelling patterns. He and his colleagues carefully analyzed how each of these elements contributed to reading skills and developed the most effective and efficient way to teach them (Barbash, 2012, p. 23; Engelmann, 2014a).

The elements, or steps, involved in teaching a given skill comprise a track and these tracks are combined into lessons. The lessons are also carefully constructed and reflect research regarding how much practice students need to learn new material and to reinforce previous learning. About 10%–15% of each lesson is new material and the remainder is devoted to "firming" material that was recently taught and reviewing material that students learned earlier and should have already mastered. The programs include clear guidelines on checking for student mastery of these elements; and teachers are given directions, based on careful research, on how best to quickly correct mistakes. When the lessons are developed they are thoroughly tested, this time in real classes with real teachers. Feedback is sent to the developers each day and the programs are revised and retested with other groups. In total it takes 3–10 years to develop a DI program (Barbash, 2012, p. 43; Engelmann, 2014a).

The theoretical development and underlying assumptions of DI are intimately associated with the field of applied behavior analysis, primarily through the work of Wesley Becker, who began a long collaboration with Engelmann in 1966. Becker, a clinical psychologist, was especially instrumental in the development of teacher training, methods of gathering formal data on student progress, and in directing research studies related to the development and testing of programs. He is perhaps best known for instructing those who work with students to "catch kids in the act of being good," a dictum that is the foundation of effective behavior management (Rasplica, 2014; Stockard et al., 2020, pp. 48–49).

DI in Classrooms

When Engelmann began his work, he and his colleagues tried to help teachers learn how to develop their own examples that were clear and explicit. But they soon found that this was extraordinarily difficult, a fact that is not at all surprising given the complex nature of all that is taught as well as the wide range of factors that teachers must contend with from minute to minute. Realizing that teachers needed more guidance, Engelmann and colleagues developed scripts with the precise wording that teachers should use. The wording is carefully developed to be simple and direct. The examples use as few words as possible. This allows students to respond quickly and teachers to easily see if students have grasped a concept and correct mistakes in a prescribed way. Research shows that when students respond quickly the concept is more quickly mastered and stored in long-term memory (Stockard et al., 2020, p. 11). This careful design of the examples, based on extensive research, is key to making sure the teaching is effective.

Some may wonder if the use of scripts, rather than teacher designed lessons, somehow hampers teachers' creativity or diminishes their responsibilities. Those who have used DI insist that it is just the opposite, freeing teachers to concentrate on their interactions with students. As Barbash (2012, p. 43) put it,

> It is unrealistic and unfair to expect teachers to be able to write their own lessons. Asking teachers to design instruction is like asking the pilot of a 747 to design the plane, or the conductor of a symphony to compose the score, or the lead in Hamlet to write the play.

Engelmann compared good teaching to good acting, encouraging teachers to "ham it up" and use their best acting skills. This helps engage students in the lesson, but also makes the process even more enjoyable for the teacher (Engelmann, 2014b).

DI in use also embodies efficiency, making sure that each moment in the classroom is devoted to making sure students are learning and being successful. The clear wording and fast paced examples contribute to this efficiency. But even more important is making sure that students are placed in the programs at a point where they will be successful. Teachers use short placement tests, again developed through careful research, to determine what skills students have already learned so that they can start a program at the point where they will be successful. This also helps ensure that students will not be bored. In other words, they begin where the material will be easy because they have the prior knowledge, but also where they will be learning new concepts and skills.

Respecting Students and Respecting Teachers

Because DI embodies respect for both students and teachers, it could be a prime ingredient in developing the more compassionate, nurturing, and equitable society that Biglan describes. Respect for students is inherent in the basic assumption that each student can succeed. If a student doesn't grasp a concept or display a needed skill, it isn't the student's fault. Instead, something is wrong with the instruction. In DI classrooms, students have continuing, concrete evidence that they are learning. Throughout each lesson the teacher tells them that they are learning new things, that they have grasped a new skill. At the end of each lesson they know that they have learned something new. They also are reminded that they learned something the day before and that they remember things they were taught a few weeks ago. DI teachers routinely put visual displays on classroom walls that show students, and classroom visitors, how much progress they have made. Each day students are reminded, with real evidence, that they are capable and successful (Engelmann, 2014b).

DI also embodies respect for teachers. All teachers want their students to succeed and DI provides the tools by which that can happen. It gives teachers the freedom to attend to the smallest elements of student understanding, see each student's response, correct when needed, provide reinforcement, and share in the pride and joy of learning. Because teachers know that the material they are presenting has been carefully designed and tested, they can devote themselves to what they signed up for—making sure that their students are successful and confident learners.

The Effectiveness of Direct Instruction: Results from a Meta-Analysis

Still, one might ask, DI programs may be well-tested during development, but do they actually work in classrooms? Over a span of almost 10 years, my colleagues and I systematically examined the research literature on DI's effectiveness to answer that question. We came to the project with backgrounds in the social sciences (sociology, history, and psychology) rather than education. However, we shared a strong concern for educational equality and the well-being of all children. We believe that a strong educational system is essential for developing a stable and inclusive democratic society.

In addition, as social scientists, we realized that good scientific judgments and effective social policy need to be based on as much information as possible. Reliable scientific conclusions must reflect multiple replications involving different samples, settings, and methodologies. Thus, we embarked on a meta-analysis to obtain quantitative estimates of DI's impact on students and, just as important, to see the extent to which these estimates were stable across different conditions. The paragraphs below summarize our results, and interested readers are referred to the full publications (Stockard et al., 2018, 2020) for details.

Methodology

Studies of DI's effectiveness began to appear in the 1960s as the programs were first developed. The literature grew rapidly. Between 1988 and 2011 10 meta-analyses and systematic reviews of this literature appeared, all of which concluded that DI was much more effective than other programs (Coughlin 2014; Stockard et al., 2020, pp. 17–19). But each of these analyses limited their review in some manner, such as looking at only certain types of outcomes, research designs, or subject areas. On average, they summarized results from 27 studies, and none summarized more than 50 research reports. Thus, it was possible that they did not include substantial portions of the available literature and it was important to know if their conclusions were supported by the full breadth of available evidence. To what extent might results vary from one study to another? Might DI benefit some students but not others, or might it be effective with only some types of outcomes or in comparison to only some types of programs?

We identified 549 reports of DI's effectiveness published over a span of 50 years, no doubt the largest body of research about any single instructional program. With the dedicated help of our university librarians we located 533 of these reports. We carefully read each one and coded details about the subject matter, methodology, sample, setting, and outcomes. In our review we wanted to be as complete as possible, looking at peer reviewed material as well as doctoral dissertations and theses, and so-called "gray literature," such as technical reports and policy papers. We were careful to note situations where one study might be described in two reports (e.g., a dissertation and an article) or where a report might describe results of two studies, and we controlled for these differences in the analysis. We also noted studies that could be seen as using methods or approaches that were less than optimal. In total, we were able to calculate valid effects from 445 research reports describing 377 studies, incorporating 4,643 effects, and involving tens of thousands of students (Stockard et al., 2020, pp. 174–175).

Having such a large group of research studies was important both to allow more precise estimates of DI's effectiveness and to test for situations or conditions in which it might be more or less effective. To quantify the results we calculated effect sizes for each comparison. We used Cohen's *d* because it is a flexible statistic and allows effects to be calculated from results presented in a range of statistical formats, such as regression coefficients, *F*-ratios, or chi-square values. (Results were, however, identical with other effect measures.) Educational researchers have traditionally used an effect size of .25 to signify results that are "educationally significant" (Tallmadge, 1977). However, based on their analysis of studies of a large number of educational programs, Lipsey et al. (2012), finding that the average effect was .28, suggested that an effect size

of .25 should be considered "large." Because effects of .50 "rarely" occur, they concluded that effects of this magnitude should be considered "huge" (pp. 4, 33–34).

We used two methods to summarize the results. The first was a relatively simple "counting" method in which we calculated the average effect size within each study. The second was a more sophisticated mixed-model regression approach that allowed us to consider differences in effects between studies and within studies. For instance, some studies reported results for students in different grade levels or from both lower-income and higher-income homes and it was important to see if these results differed. We also calculated results with what we called a "reduced sample" omitting studies or effects that had unusually large or small results (outliers), that had some type of quality issue, or that involved only maintenance periods. For the mixed models we used the xtmixed program in STATA, with study and design (nested in studies) as random effects. We used the marginals procedure in STATA to estimate effects and confidence intervals for various subgroups involved in a given analysis. To help increase accuracy, we also included a control for the number of students in a given comparison (see Stockard et al., 2020, pp. 34–42, 173–186, for additional details).

Findings

Some of our findings are summarized in Tables 1 and 2, and 3. The first column of numbers in each table reports the estimated effect for a group, the second column reports the number of effects for that group, and the third column reports the number of studies. The estimated effects were derived from the mixed model analyses using the reduced sample (n = 327 studies and 3,477 effects) and are similar to those found with the counting analysis. These estimates are, however, slightly smaller than estimates obtained with the full sample, primarily because most of the outliers were positive. Thus the data reported should be seen as a conservative estimate of DI's effect.

Across all 327 studies and almost 3,500 effects, the average effect size was .56 (the first line in Table 1). The 95% confidence interval ranged from .50 to .61. Thus, the *entire confidence band* falls within the range that Lipsey et al. (2012) indicated only rarely occurs in studies of educational programs—the level that they described as "huge." Effect sizes can be translated into an "Improvement Index," based on percentiles, and these values help depict the magnitude of the results. An average effect of .56 is equivalent to an improvement index of 21 percentile points. This indicates that, based on results from thousands of effect sizes and hundreds of studies, *an average student taught with DI would be expected to score 21 percentile points higher than an average student taught with other methods*.

Because we examined the results of hundreds of studies and thousands of effect sizes we were able to systematically look at the extent to which a wide range of variables might moderate or alter the impact of DI on student outcomes. Our results were simple, straightforward, and consistent, no matter how we "sliced" the data.

Table 1 provides examples of results related to methodological characteristics. It shows that average effects were statistically equivalent across the 5 decades of research that we examined and also equivalent to those found in the large-scale Project Follow Through conducted in the 1960s and 1970s (and discussed more thoroughly below). Estimated effects were similar with different assessments and when published in different outlets. They were similar with different types of research designs, in

Table 1 Predicted Effect (Cohen's d) of DI on Student Outcomes, Mixed Model Analyses, Baseline, by Year of Publication, and by Selected Methodological Characteristics

Comparison	Effect Size	Number of Effects	Number of Studies
Baseline (no controls)	0.56	3,477	327
Year of Publication			
Follow Through	0.63	998	20
1966–76	0.55	165	22
1977–86	0.56	295	45
1987–96	0.64	426	49
1997–06	0.48	917	108
2007–16	0.55	676	83
Maintenance Effect			
Not Single Subject	0.42	483	37
Single Subject	0.55	39	10
All Maintenance Effects	0.43	522	47
Research Design			
Random Assignment	0.52	471	64
Norm Control	0.62	1,096	78
Cohort Control	0.54	386	65
Statistical Control	0.46	496	33
Other Pre/Post	0.48	391	61
Other Post Only	0.63	285	32
Single Subj. and Other Designs**	0.95	352	37
Type of Assessment			
Normed/Published	0.57	2,492	186
CBM***	0.79	369	56
State Assessments	0.56	335	44
Experimenter Designed/Other	0.44	281	41
Type of Publication			
Article	0.56	2,329	170
Thesis	0.43	392	57
Gray Literature	0.55	756	100

Note. Effects calculated from mixed model analyses in which effects were nested within designs (n = 397), which were nested within studies. The number of effects per study ranged from 1 to 205, average 10.6. The number of effects per design ranged from 1 to 195, average 8.8. Models controlled for sample size (using ln (n) to adjust for skewedness of the variable). Estimates of maintenance effects based on 522 effects, in 47 studies with 54 designs. *** = $p < .001$; ** = $p < .01$, * = $p < .05$. Calculated from Stockard et al., 2020, pp. 36, 39–40, 65–66, 73–74, 78–80, 141–142.

randomized control trials and in more realistic field settings, although stronger in single subject designs. We also examined differences in the method of delivering instruction (an experimental teacher or a classroom teacher), the breadth and size of the sample, the type of data that were available (e.g., means and *SD*s, percentages, and counts), who wrote or sponsored the article (e.g., those associated with the publisher, ourselves), issues regarding the quality of the study, and the criteria used by the What Works

Table 2 Predicted Effect (Cohen's d) of DI on Student Achievement, Mixed Model Analyses by Characteristics of Study Setting and Students

Comparison	Effect	N of Effects	N of Studies
Locale			
Urban U.S.	0.54	1,840	166
Suburban/ Rural U.S.	0.55	687	65
Multiple and Non-U.S.	0.54	950	96
Control of School			
Public Schools	0.55	2,490	240
University Affiliated*	0.90	203	25
Private	0.77	249	26
Publicly Funded Altern.	0.62	422	24
Other	0.45	113	12
Family Income			
Not Low Income	0.52	1,947	245
Low Income	0.58	1,530	82
Minority Status			
Lower Proportion Students of Color	0.55	2,626	265
High Proportion Students of Color	0.54	851	62
History of Low Academic Achievement			
No History	0.53	2,486	176
Low Skills	0.47	390	55
Special Classroom	0.63	453	69
Low Incidence Disabilities	0.72	148	27
Grade Level			
Pre or K	0.55	453	49
Grade 1	0.57	501	39
Grade 2	0.53	425	45
Grade 3	0.56	1,069	51
Grade 4	0.61	282	49
Grades 5–8	0.55	441	56
Grade 9 and higher	0.52	306	38

Note. Source, Stockard et al. 2020, pp. 107–110; see also note to Table 1.

Clearinghouse (WWC, 2014) in its review process. In total, based on the mixed model regressions, we calculated confidence intervals around the estimated effect of DI for 53 methodological characteristics. Rules of statistical probability indicate that at least two comparisons would result in estimates that were negative and statistically significant. But *all* of the estimates were not only greater than zero, they were larger than the average value (.28) reported by Lipsey et al. (2012) and often greater than the level of .50 that they suggested occurred only rarely (Stockard et al., 2020, pp. 40, 74–83, 100).

Similar results appeared when we looked at effects in different types of settings or with students with different backgrounds (Stockard et al., 2020, pp. 105-111). Table 2 reports some of those results. It shows estimated effects for students in different

Table 3 Predicted Effect (Cohen's d) of DI from Mixed Model Analyses, by Type of Outcome and Comparison Program

Comparison	Effect Size	N Effects	N Studies
Academic Areas			
Reading	0.53	1,832	184
Math **	0.40	608	72
Language	0.55	284	34
Spelling	0.45	264	14
Other Academic Areas	0.61	168	13
All Academic Areas	0.52	3,155	317
Nonacademic Areas			
Student Affective	0.36	95	18
Teacher–Parent Attitudes	0.13	110	11
Ability	0.39	117	22
All Nonacademic	0.38	322	51
Comparison Curriculum			
Basic Skills**	0.49	545	46
Developmental, Balanced	0.57	501	52
Traditional/Usual	0.65	2,431	229

Note. The category of "developmental, balanced" also included those described as "cognitive" and teacher developed. There were no differences in effects between these categories. The category of "traditional/usual" included norm-comparisons. Categories were combined to increase degrees of freedom and thus the accuracy of estimates. * = $p < .05$, ** = $p < .01$, *** = $p < .001$. Calculated from Stockard, et al. 2020, pp. 112–116.

geographic locales, in different types of schools, with students from families with different incomes, with students of color and other students, for students with or without a history of lower achievement, and at different grade levels. In total we looked at 29 different types of settings and student characteristics and the average estimated effect was greater than .50 in all but 3. The minimum effect was .41 to .47, again substantially larger than the average results reported by Lipsey and associates (2012).

We also found similar estimates when we compared results across different types of outcomes, such as academic areas of reading and math, and nonacademic outcomes, including students' self-esteem and views of learning and teachers' opinions of the program. Effects were similar in comparison to all types of other curricula, including those that were described as oriented toward "basic skills" or incorporating the "little di" methods (Table 3). They were also similar from one type of DI program to another (Stockard et al., 2020, pp. 113–121). As a further test of these results we calculated "joint models," including any of the characteristics that were significant in individual comparisons. These analyses confirmed our results.

In general, there was no indication that DI was effective with only certain "types of students" or with "low-level" or rote skills but not with "higher-order" thinking. Nor was there any evidence that it somehow destroys students' love of learning or that teachers don't like the programs. In fact, results indicate just the opposite, students taught with DI had higher academic achievement, were more confident in their

academic abilities and had more positive attitudes toward school, just as the underlying theory would suggest. After using DI, teachers are more confident of their abilities and more positive toward the program. With each of the control variables, our estimates of the effect of DI, and the associated confidence intervals, were similar to the baseline results. Moreover, the positive impacts of DI were long-lasting. Across the 47 studies of long-term impacts, the estimated effect was .43, substantially greater than the average short-term impact of other programs (Table 1).

In 1996, Gary Adams, the author of the first meta-analysis of studies of DI's effectiveness, wrote that he was "stunned at the results" and that his analysis "revealed the largest effect sizes that I had ever seen" (Adams & Engelmann, 1996, p. vi). Writing a quarter of a century later, we made a similar observation:

> At the start of this project, even though we had read Adams' work as well as that of others, we weren't prepared for the sheer magnitude of the body of research on DI. We knew that a lot of research had accumulated over the half-century, but were stunned to find that there were hundreds of reports. As we read, summarized and coded the material, we realized that the vast majority reported positive findings. But, it was only when we actually "crunched the numbers" that we saw both the magnitude of the effects and their consistency. The results were so strong and so consistent that we checked, rechecked, and re-rechecked our findings to make sure that they were correct. Whenever there was even a hint of a problem or inconsistency, we repeated our analyses, tried other methods of checking results, and returned to the text of individual studies to make sure our coding was accurate. . . . [W]e looked, in as many ways as we could think of, for conditions under which DI might be less effective. We simply could not find any situation in which DI did not work. (Stockard et al., 2020, p. 146)

Enhancing the Effectiveness of DI

We did, however, uncover two factors that made DI even more effective. The first was exposure or dosage. We measured higher dosage in three ways: starting the programs in kindergarten or first grade, being taught with DI over multiple grades, and following the recommended length of class periods in each day. The second factor was implementation fidelity, which we measured with a scale that captured the extent to which teachers and schools followed the developers' guidelines. If a program were ineffective, a larger dose would not be beneficial. If the design were faulty it wouldn't matter if users followed the guidelines. Thus, the fact that both of these factors increased effectiveness could be seen as additional evidence of its positive impact (Stockard, 2010; Stockard et al., 2020, pp. 124–140).

Figure 1 summarizes the results of our analysis. The bars in the figure represent the estimated effect size for students with different levels of exposure and teacher fidelity. Results were calculated separately for each academic area and included any variables found to be significantly associated with an outcome. For reading, the estimated effect for students who began DI after K, had 1 year of exposure, and were in a classroom with relatively low fidelity was .38. But for students who began the program in kindergarten, were in a classroom with high fidelity, and had DI through grade 3, the

estimated effect was 1.11. A similar pattern occurred with other subjects. For math, the estimated effect was .41 for students who began DI after kindergarten, but .77 for those who started in K. With other academic subjects, such as spelling and language, the estimated effect was .38 for students who began DI after K and were in low-fidelity settings, but 1.05 for those who started in K and were in high-fidelity settings.

The two horizontal lines in Fig. 1 provide important comparisons. The lower line corresponds to an effect of .28, the average value for other studies reported by Lipsey et al. (2012). The upper line, which corresponds to an effect of .74, represents the difference in average NAEP scores in reading of students eligible and not eligible for the federal free or reduced lunch program (Stockard et al., 2020, p. 30). In other words, this line estimates the difference in NAEP scores for students from low-income homes and those from other homes. All of the estimated effects in Fig. 1, no matter what the dosage or quality of implementation, were larger than the average reported in other studies. But with more optimal conditions the effects were greater, often substantially greater, than the difference in achievement scores of those from low income and other homes.

From these data we calculated projections to show the way in which systematic use of DI could produce strong changes in NAEP scores. With simply the average results across the hundreds of studies in our analysis, the percentage of fourth grade students classified as scoring at the below basic level in reading would drop by more than half (from 34% to 15%), and the percentage classified as proficient or advanced would be projected to almost double (from 35% to 59%). Projections of change from results with optimal exposure and fidelity would, of course, result in even more dramatic changes, with only 6% projected to remain at the below basic level and over three fourths (78%) projected to be classified as proficient or advanced (Stockard et al., 2020, pp. 150–152, 200–201).

Of course, these projections are estimates, derived from results of statistical analyses. Although it would be inappropriate to attach extraordinary precision to the numbers, it would be entirely appropriate to suggest that they illustrate the type of change that

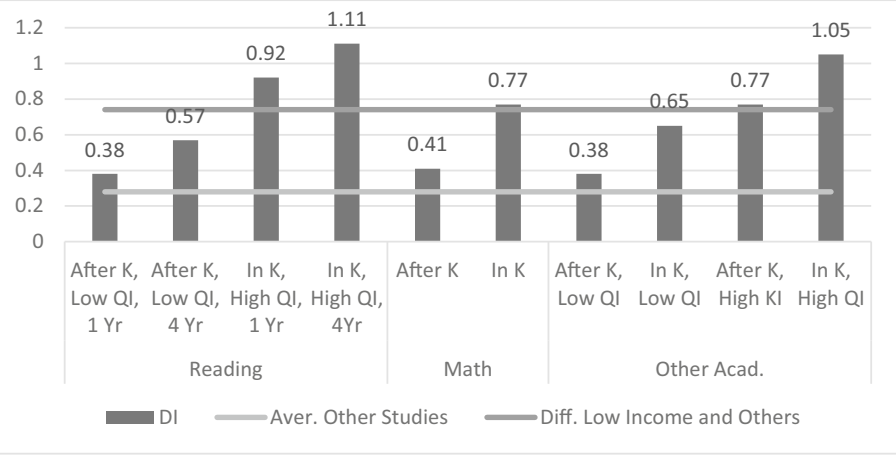

Fig 1. Estimated Effect of DI by Dosage and Fidelity Source: Calculated from mixed model regressions from final joint, reduced models within each academic area. See Stockard et al., 2020, pp. 137–140

could occur with more wide-spread use of DI. And they bolster the argument that DI could be a key ingredient in building the compassionate and equitable society envisioned by Biglan.

Alternative Views of Attaining Educational Equality

It is probably not an exaggeration to say that all within our society want students to be successful and confident. Most no doubt also want an educational system that promotes a more equitable society in which all can reach their full potential and equally participate in social, political, and cultural worlds. Yet, views about how to reach this goal can appear to contrast sharply with the assumptions that underlie DI and the research evidence. Although it is impossible to fully describe these perspectives in this article, it is important to acknowledge them and explain issues with their underlying logic, but also point to commonalities with DI. One general approach focuses on the impact of poverty and discrimination on student achievement, the other on educational processes and approaches.

Low Achievement Reflects Poverty and Discrimination

Compared to other democratic, industrialized countries, the United States has distressingly high levels of poverty and economic inequality, and provides far less support for low-income families. As a result there are large differences in the resources that individual families can devote to their children's educational endeavors. Moreover given long-standing patterns of housing segregation related to income and race-ethnicity, there are often vast differences in the educational resources allocated to students in lower- and higher-income communities. Thus, some imply that the differences in NAEP scores described above result from variations in economic resources and related factors. Differences in achievement of students from minority groups and other students reflect overrepresentation in low-income communities as well other serious problems ranging from segregated housing to demeaning and insensitive daily interactions (e.g., Darling-Hammond, 2010; Ladd, 2012).

I believe that poverty, inequality, and discrimination must be addressed if we are to have a truly democratic and just society. Moreover, differences in educational resources should and must be addressed. Yet, the *root sources of economic, social, and interactional inequality are not the same as the root sources of low educational achievement*. The former reflects deep problems in our economic and political systems as well as deep-seated cultural and social patterns of discrimination and prejudice. The latter stems from our educational system and the way in which students are taught. Both must be dealt with. But, because they have different root causes, the most efficient and effective ways to handle these problems are different.

To imply that altering patterns of low achievement depends upon eliminating poverty is not only illogical, it ignores decades of research evidence, as summarized above. The studies of DI convincingly show that children from all types of backgrounds can be successful students. Moreover, NAEP results show relatively high rates of academic problems among students from higher income backgrounds. Thus, there is no guarantee that economic changes alone would magically produce higher

achievement. Finally, because political power is concentrated among those with the greatest economic resources, policies designed to reduce poverty and inequality usually face vehement and strong opposition. Thus implying that greater student achievement can only occur when poverty and inequality are eliminated implies acceptance of an extraordinarily inefficient method of change.

The suggestion that growing up in a low-income home is causally related to low achievement can also be seen as demeaning to students from those environments. To suggest that students from low-income families cannot succeed academically until poverty and discrimination are eliminated, or to imply that their family background or community environment somehow makes learning especially hard, ignores their inherent capabilities and smacks of the worst kinds of prejudice. Such views no doubt reflect adherents' sympathies for students from these backgrounds and are not meant to be demeaning. However, I suggest, as explained above, that DI's assumption that all students can learn is inherently much more respectful of students from all backgrounds.

Alternative Views of Learning

Much of the educational establishment adheres to philosophical views of how children learn that contrast sharply with the assumptions and research findings of DI. These views take a variety of forms. For instance, those who emphasize a developmental approach suggest that lower levels of achievement reflect differences in development or prior experiences, perhaps as a result of home or community environments or some type of "disabling" condition. Thus students may have lower levels of achievement because they are not yet "ready to learn" and may need additional time or experiences. Those who advocate a constructivist view emphasize the importance of students' developing their own understandings and interpretations of a subject. They too suggest that failure to learn can reflect differences in students' backgrounds, prior learning, or their self-esteem or confidence as a student. Some criticize common definitions of literacy and other academic skills, suggesting that they are narrow and culturally biased. They believe that these biases are reflected in standardized assessments, such as the NAEP. Thus, they call for instructional methods that build on and value students' unique cultural heritages and imply that the process of education is more important than ultimate outcomes. A common thread in these views is the rejection of research conclusions regarding the effectiveness of structured and explicit instruction and the promotion of "flexible" educational approaches with a "constructivist" or "developmental" orientation (e.g., Au, 1998; Esmonde, 2009; Gee, 2015; Guiterrez et al., 2009; Jackson et al., 2008–2009; Nasir et al., 2006; Street, 2005).

Those who adopt these perspectives are well-intentioned and, I believe, sincerely want students to be successful and self-confident. But waiting for students to "develop" maturity or obtain greater knowledge is an extraordinarily inefficient approach to learning and can only exacerbate differences in achievement (Silvestri and Heward, 2016; Stone, 1994, 1996). Suggesting that students' self-confidence must be enhanced before they can learn is not only inefficient but perversely reverses the underlying causal order. In reality students develop greater self-confidence when they are successful learners (Heward 2003, pp. 193-194). The suggestion that definitions of "literacy" or "achievement" be altered for some groups of students is potentially most dangerous, in essence channeling groups of students into alternative pathways with no guarantee

that these routes will result in greater adult success or build a generation that is better equipped to counter issues of discrimination and prejudice.

The most compelling criticisms of these views have been given in recent years by civil rights leaders and organizations, many of whom have decried current educational practices. These critics have stressed the importance of "full and complete literacy," strong academic understandings by all students, as essential to maintaining democratic institutions (Spain et al., 2020; Weaver, 2019). Directly commenting on those who call for a flexible approach to education, Lane (2019) wrote,

> Just two generations ago people risked their lives to be able to read and here we are today watching the educational establishment—through its degradation of standardized assessments, emphasis on the individual over the collective whole, and dismissal of science—risk the subjugation of an entire people to second class citizenship. It is frightening and marks the gravest miscarriage of justice we have seen this side of educational history. An entire generation of children is not being taught to read. (see also McWhorter, 2009, 2011)

Similar Concerns, DI as a Common Solution

It is important to stress that adherents to both of the perspectives described above hold the same ultimate goal as those who developed and use DI. They all want students to be successful and confident learners, but they differ in views of how best to accomplish this goal. I suggest that DI could accomplish this shared goal more effectively and more efficiently. The data show that students from all backgrounds can succeed, including those who come from low income homes or communities, thus countering the fears of those who focus on the need to eliminate poverty and inequality. In addition, the development and use of DI embody the respect for students and accommodations for variations in prior learning that concern those who call for flexible approaches to instruction. Finally, both of the perspectives outlined above clearly imply a concern with building a more equitable society. A more effective educational system, with increasing numbers of knowledgeable and confident students, would almost certainly increase the numbers of people to work toward that goal. I suggest that an excellent way to increase this pool of concerned citizens is through systematic implementation of DI.

Building a More Compassionate, Caring Society

Many people care deeply about the problems of our nation's schools. So it is reasonable to ask why, given the research evidence, it has not been more widely adopted. Below I describe examples of active opposition to DI and speculate why this has occurred. Then, paralleling Biglan's (2020) analysis, I suggest groups of social actors that could partake in a coalition to promote a more effective and equitable educational system.

Roadblocks to Change

Despite the accumulated evidence of DI's effectiveness, it has faced strong opposition, almost from the moment the first studies appeared (Engelmann, 2007, p. 13; Stockard

et al., 2020, p. 47). Foremost among the critics have been faculty in teacher education programs and their colleagues within the world of educational policy. Few teacher education programs provide instruction in DI, let alone other types of structured methods, and can be quite vocal in dismissing the approach. As a result, teachers and administrators who graduate from these programs are unlikely to have learned of DI. The faculty in teacher education programs are often closely aligned with powerful actors in the policy world, including those who set requirements for teacher education and "approved" curricula. They also influence the way that research results are presented to the public (Stockard et al., 2020, 152–155, 156; Walsh, 2013).

Two examples can illustrate the powerful role of educational "gate keepers" in hiding evidence of DI's effectiveness. The first involves Project Follow Through, a field study of the relative effectiveness of educational programs conducted in the late 1960s and 1970s that included tens of thousands of students and dozens of schools throughout the country. Besides DI, the programs included others termed "basic skills" in orientation and several that embodied the developmental and constructivist approaches. Extensive support was provided for teacher training as well as nutritional meals, health care, and dental care in both the schools in the programs and those in the control group. The success of the approaches was carefully evaluated using measures recommended and accepted by sponsors of the programs, and the data were evaluated with well-regarded statistical methods and appropriate controls. The results revealed that DI was the *only* program with significantly positive outcomes on all of the outcome measures, both academic and affective. (As shown in Table 1, the associated effect size, from our analysis, across all types of outcomes and compared to all other programs was .63.) In other words, much like our meta-analysis of many more studies, the results of Follow Through indicated that DI was much more likely than the types of programs favored by the educational establishment to result in greater academic achievement and higher student self-concept. Again, parallel to the results of our meta-analysis, student achievement was higher with greater exposure to the programs and when taught with greater fidelity (Grossen, 1996, Stockard et al., 2020, pp. 49–53; Watkins, 1996, 2009).

Those who designed Project Follow Through had hoped that such an extensive field experiment would help resolve debates about the relative effectiveness of educational programs. The evidence was certainly overwhelming. But powerful actors in the world of educational policy effectively hid the results from the public. The final official report combined the results from all the programs and announced that the results of Follow Through were a failure. Because only one of the programs (DI) was effective, its potential was hidden, its positive findings overwhelmed by the many more negative results. Many scholars and representatives of the evaluators objected, but to no avail, and the results of this extraordinarily large, expensive, and carefully designed study have become virtually unknown in the world of educational research (Stockard et al., 2020, pp. 55–57; Watkins, 1996, 2009).

Although the maneuverings involving reports on Follow Through could appear designed to hide the effectiveness of DI, other actions have been less overt, but just as effective. For instance, the What Works Clearinghouse (WWC) was established by the U.S. Department of Education to develop summaries of research on educational programs and recommend those that are most effective (WWC n.d.). A key tenet of science is the notion of cumulative evidence, replicating studies in as many different settings and with as many different approaches as possible. Meta-analyses reflect this

inclusive approach. In contrast, the WWC adopted a highly "exclusive" method, choosing to examine studies only if they met each of a long list of criteria, such as published after a certain time and using certain methodological approaches. When this approach was open to public comment, over 90% of the almost 300 scholars and professional organizations who responded, including representatives of the American Evaluation Association and the American Educational Research Association, objected. Reflecting the standard view of science as cumulative in nature, they feared that the WWC procedures would provide only a limited view of the available literature. Yet, even with such concerted concerns, the WWC made no substantive changes to its plans; and, in fact, later adopted even more limitations to the studies that would be considered.

I have not been able to find any methodological or theoretical justification for the WWC's approach to summarizing research literature, but it is possible that results with their methods would not differ from those obtained with the more traditional inclusive approach. My colleagues and I tested this possibility using studies of DI. Our results were not encouraging. *None* of the hundreds of reports on DI's effectiveness would pass all of the criteria used by the WWC, and thus the WWC would conclude that there was no "acceptable" evidence regarding DI. We then examined the possibility that their screening criteria might somehow produce more accurate or precise estimates of DI's effect and looked at differences in these estimates in studies that did or did not pass each of the various criteria. But, as noted above, the effects of DI were substantial no matter which criteria would be applied (Stockard & Wood, 2017; Stockard et al., 2020, pp. 84–92).

A range of other problems with WWC policies and procedures have been documented including erroneous interpretations of the research and recommending the use of programs deemed ineffective in the original studies (Stockard et al., 2020, pp. 93–100; Wood, 2014b). The net result is that reports from the WWC, like the official reports from Project Follow Through, suggest that there are no programs that can promote strong educational success. Both Follow Through and the WWC were launched with the expectation that they would provide data-based tests of the effectiveness of educational programs. Yet they appear to have failed in that mission. Instead, their decisions appear to be closely aligned and, indeed supportive of, the interests of those within the educational establishment who have so strongly opposed DI.

Why is Opposition So Strong?

As an outsider to the world of education, it has been difficult for me to understand why there is so much resistance to DI. The research findings are extraordinarily strong. Students taught with DI have higher academic achievement and more confidence in their learning ability than those taught with other programs. In addition, when they use the programs, teachers like it; they feel good about their work and the way they can help their students. Why then have powerful forces within the educational establishment ignored the research evidence and, in fact, actively worked against its use. Part of the answer involves their deeply held beliefs about how children learn and their faith in the philosophies that underlie developmental and constructivist approaches. Yet, there are two additional possible explanations, both of which involve the benefits that can accrue from resisting the widespread use of DI.

The first involves the simple fact that abandoning opposition to DI would wound professional pride. Many academic and professional reputations are based on adherence and strong support of particular views. For academics and many of the policy establishment to abandon their current ideological postures would appear to involve changes akin to the scientific revolutions described by Thomas Kuhn (1970). Such changes have indeed happened, but they have usually taken many years.

The second possible explanation involves monetary benefits. The federal government and private foundations spend many millions of dollars each year on research to develop "effective" curricula, increase student self-confidence, or develop assessments to document student "failure." Faculty in teacher education programs and many in the educational policy world receive these funds. The profits of publishing companies are regularly increased by producing new products that are marketed as somehow "solving" learning problems. Accepting the research regarding DI's effectiveness could directly affect this flow of money or at least force researchers to find alternative issues to study.

Implications for Change

The forces opposing the use of DI are clearly powerful and numerous, yet Engelmann, Becker, their colleagues, and a substantial number of those within the educational world worked diligently for decades to demonstrate the power of DI and counter the opponents. They did so believing that the DI could help not just individual students, but dramatically improve our nation's educational system as a whole (e.g., Engelmann, 1992, 2007). Their efforts were largely unsuccessful. But the recent writings of Biglan (2020) can provide hope for change. I believe that systematic use of DI could help build the compassionate, supportive society that he envisions.

In discussing how to build this better world, Biglan stresses the importance of collective action. He asked, rhetorically,

> What would happen if all the people concerned about one or more of the problems we face as a nation formed a coalition around a clear vision of the kind of society we want to build and committed to work over an indefinite period of time to build the society we aspire to? We would finally succeed in changing the direction of the nation, our economy, and the well-being of our people. (Biglan, 2020, p. 92)

Many within the educational establishment might be reluctant to join a coalition that features DI as part of this movement for change. Yet, there are potentially several other sources of energy to challenge and change the current educational system. I do not have data to support the ideas presented below. They are clearly speculative, but presented in a spirit of hope.

One potential source is the intellectual power of researchers from multiple academic disciplines. Foremost among them could be faculty in schools of education and those in educational policy positions who promote the role of scientific evidence in decision making. They could join with scholars in empirically based social sciences, such as political science, economics, psychology, and sociology, who are concerned with student success and developing a better educational system. Together, they could provide an important counterweight to the power of the educational establishment.

A second potential source of support for change could be the business community, or at least those segments that do not profit from having a poorly educated workforce. Economists have estimated that even relatively modest increases in academic achievement could result in substantial increases in national economic productivity (PISA, 2010). Greater human capital in the workforce would also increase profits for individual businesses and corporations.

Third, all who are concerned with poverty and economic inequality should be potential allies. For just as higher achievement results in greater profitability for individual business, it is also related to the growth of democracy and to growing political pressure for change. The civil rights organizations, whose views were highlighted above, would be natural partners in such an endeavor. In addition, those who focus on ending poverty as a means of raising student achievement should be encouraged to join the efforts.

Finally, those who have used DI and seen its success could be the biggest source of energy in a coalition for change. Teachers and administrators who have used DI in their classrooms could promote its use for others. Parents whose children have benefitted could add their voices. Their students, classrooms, and schools could be a model—showing a pathway to develop a more caring and more successful schooling experience for all students.

Conclusion

It is clear that our nation is at a crossroads and that it is time to build a compassionate and more caring society that promotes the well-being of all. A strong educational system needs to be part of this transformation, and an extensive body of evidence shows that Direct Instruction could be key in the process. When taught with DI students learn more and come to see themselves as more successful. The instructional system embodies respect for students and for teachers and addresses the goals of those who advocate other paths to student success. Many powerful actors within the educational establishment have expressed opposition to DI and actively worked to hide the evidence of its effectiveness. However, others could work together to counter this power and build an educational system that promotes the well-being of all. Our nation's students deserve no less.

References

Adams, G. L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Educational Achievement Systems.

Au, K. H. (1998). Social constructivism and the school literacy learning of students of diverse backgrounds. *Journal of Literacy Research, 30*(2), 297–319.

Barbash, S. (2012). *Clear teaching: With Direct Instruction*. Education Consumers Foundation: *Siegfried Engelmann discovered a better way of teaching.*

Biglan, A. (2020). *Rebooting capitalism: How we can forge a society that works for everyone*. Values to Action.

Coughlin, C. (2014). Outcomes of Engelmann's Direct Instruction: Research syntheses. In J. Stockard (Ed.), *The science and success of Engelmann's Direct Instruction (pp. 25–54)*. NIFDI Press.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. Teachers College Press.

Engelmann, S. (1992). *War against the schools' academic child abuse*. Halcyon House.

Engelmann, S. (2007). *Teaching needy kids in our backward system: 42 years of trying*. ADI Press.

Engelmann, S. (2014a). Research from the inside: The development and testing of DI programs. In J. Stockard (Ed.), *The science and success of Engelmann's Direct Instruction (pp. 3–24)*. NIFDI Press.

Engelmann, S. (2014b). *Successful and confident studies with Direct Instruction*. NIFDI Press.

Engelmann, S. (2018). *Strategies for teaching students with severe and low-incidence disabilities*. (C. R. Khoury & J. Stockard, ed.). NIFDI Press.

Engelmann, S., & Carnine, D. W. (1991). *Theory of instruction: Principles and applications* (rev. ed.) ADI Press. (Original work published 1982)

Engelmann, S., & Carnine, D. W. (2011). *Could John Stuart Mill have saved our schools?* Full Court Press.

Engelmann, S., & Colvin, G. (2006). *Rubric for identifying authentic Direct Instruction programs*. Engelmann Foundation.

Engelmann, S., & Steely, D. (2004). *Inferred functions of performance and learning*. Lawrence Erlbaum Associates.

Esmonde, I. (2009). Ideas and identities: Supporting equity in cooperative mathematics learning. *Review of Educational Research, 79*(2), 1008–1043.

Gee, J. P. (2015). The new literacy studies. In *The Routledge handbook of literacy studies (pp. 35–48)*. Routledge. Retrieved from August 31, 2020, from https://www.routledgehandbooks.com/doi/10.4324/9781315717647.ch2

Grossen, B. (1996). The story behind project Follow Through. *Effective School Practices, 15*(Winter), 4–10.

Guiterrez, K. D., Morales, P. Z., & Martinez, D. C. (2009). Re-mediating literacy: Culture, difference, and learning for students from nondominant communities. *Review of Research in Education, 33*, 212–245.

Heward, W. L. (2003). Ten faulty notions about teaching and learning that hinder the effectiveness of special education. *The Journal of Special Education, 36*(4), 186–205.

Jackson, L. B., Ryndak, D. L., & Wehmeyer, M. L. (2008–2009). The dynamic relationship between context, curriculum, and student learning: A case for inclusive education as a research-based practice. *Research & Practice for Persons with Severe Disabilities, 33–34*(4–1), 175–195.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Ladd, H. F. (2012). Education and poverty: Confronting the evidence. *Journal of Policy Analysis and Management, 31*(2), 203–227.

Lane, J. (2019). Literacy: The forgotten social justice issue. *Project Forever Free.* Retrieved September 3, 2019, from https://projectforeverfree.org/literacy-the-most-forgotten-social-justice-issue/

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NSER 2013-3000). In *National Center for Special Education Research*. Institute of Education Sciences: U.S. Department of Education.

McWhorter, J. (2009) Why aren't we taking advantage of an old solution to the reading-gap problem. *The New Republic*. Retrieved from https://newrepublic.com/article/47597/why-arent-we-taking-advantage-old-solution-the-reading-gap-problem

McWhorter, J. (2011). Color blind: The misguided liberal obsession with school integration. *The New Republic*. Retrieved April 21, 2011, from https://newrepublic.com/article/87207/education-segregation-integration-liberals-herbert

Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2006). Learning as a cultural process: Achieving equity through diversity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences (pp. 489–504)*. Cambridge University Press.

National Assessment of Educational Progress (NAEP) (n.d.). NAEP: National Assessment of Educational Progress. *National Center for Education Statistics.* Retrieved from https://nces.ed.gov/nationsreportcard/

National Institute for Direct Instruction (NIFDI). (2017). Comprehensive bibliography on Direct Instruction. Retrieved March 7, 2021, from https://www.nifdi.org/research/di-bibliographies/comprehensive-bibliography

Programme for International Student Assessment (PISA) (2010). The high cost of low educational performance: The long-run economic impact of improving PISA outcomes. Organisation of Economic Co-operation & Development

Programme for International Student Assessment (PISA) (2019). PISA 2018 results. *Organisation of Economic Co-operation & Development.* Retrieved March 7, 2021, from https://www.oecd.org/pisa/publications/pisa-2018-results.htm

Rasplica, C. (2014). Direct Instruction and behavior support in schools. In J. Stockard (Ed.), *The science and success of Engelmann's Direct Instruction (pp. 124–140)*. NIFDI Press.

Rosenshine, B. (1971). *Teaching behaviors and student achievement*. National Foundation for Educational Research in England & Wales.

Rosenshine, B. (1978). Academic engaged time, content covered, and direct instruction. *Journal of Education, 160*(3), 38–66.

Silvestri, S. M., & Heward, W. L. (2016). The neutralization of special education, revisted. In R. M. Foxx, & J. A. Mulick (Eds.). *Controversial therapies for autism and intellectual disabilities: Fad, fashion, and science in professional practice* (2nd ed.) (pp. 136-153). Routledge/Taylor & Francis.

Spain, J. D., Sr., Walker, S. S., & Kerns, S. L. (2020). Providing evidence-based literacy instruction in our schools: A letter from the Arlington County NAACP. *Reading Rockets.* Retrieved from https://www.readingrockets.org/article/providing-evidence-based-literacy-instruction-our-schools-letter-arlington-county-naacp

Stockard, J. (2010). An analysis of the Fidelity implementation policies of the What Works Clearinghouse. *Current Issues in Education, 13*(4).

Stockard, J., & Wood, T. W. (2017). The threshold and inclusive approaches to determining "best available evidence": An empirical analysis. *American Journal of Evaluation, 38*(December), 471–492.

Stockard, J., Wood, T. W., Coughlin, C., & Khoury, C. R. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research, 88*(4), 479–507. https://doi.org/10.3102/0034654317751919.

Stockard, J., Wood, T. W., Coughlin, C., & Khoury, C. R. (2020). *All students can succeed: A half century of research on the effectiveness of Direct Instruction*. Lexington Books/Rowman & Littlefield.

Street, B. V. (2005). At last: Recent applications of new literacy studies in educational contexts. *Research in the Teaching of English, 39*(4), 417–423.

Stone, J. E. (1994). Developmentalism's impediments to school reform: Three recommendations for overcoming them. In R. Gardner III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 57–72). Monterey, CA: Brooks/Cole.

Stone, J.E. (1996). Developmentalism: An obscure but pervasive restriction. *Education Policy Analysis Archives, 4*(8). Retrieved from https://usf.sobek.ufl.edu/content/SF/S0/02/45/11/00057/E11-00057.pdf.

Tallmadge, G. K. (1977). *The Joint Dissemination Review Panel idea book*. Office of Education: National Institute of Education and U.S.

Walsh, K. (2013). 21st-century teacher education. *Education Next, 13*(3). Retrieved March 7, 2021, from https://www.educationnext.org/21st-century-teacher-education/

Watkins, C. L. (1996). Follow Through: Why didn't we? *Effective School Practices, 15*(Winter), 57–66.

Watkins, C. L. (2009). *Project Follow Through: A case study of contingencies influencing instructional practices of the educational establishment*. Cambridge Center for Behavioral Studies.

Weaver, K. (2019). The biggest threat to our democracy: Illiteracy and the science-deniers who contribute to it. *The Latest News from Core*. Retrieved from https://www.corelearn.com/the-biggest-threat-to-our-democracy-illiteracy-and-the-science-deniers-who-contribute-to-it/

What Works Clearinghouse. (n.d.) What Works Clearinghouse. https://www.ies.ed.gov/ncee/wwc/. Retrieved 8 Apr 2021.

What Works Clearinghouse. (2014). *WWC procedures and standards handbook* (Version 3.0). Washington, DC: Institute of Education Sciences. Retrieved from https://www.ies.ed.gov/ncee/wwc/.

Wood, T. W. (2014a). *Engelmann's Direct Instruction: Selected writings from the past half century*. NIFDI Press.

Wood, T. W. (2014b). *Examining the inaccuracies and mystifying policies and standards of the What Works Clearinghouse: Findings from a Freedom of Information Act request*. NIFDI Technical Report 2014-5. NIFDI Press.