A·B·A·I
Association for Behavior Analysis International

Check for updates

Just How Effective is Direct Instruction?

Lee Mason[1,2] • Maria Otero[1,3]

Abstract

Despite overwhelming evidence in support of Direct Instruction, this research-validated curriculum has not been widely embraced by teachers or school administrators. The Direct Instruction model, developed and refined by Engelmann and colleagues over the past 50 years, has been the focus of numerous research studies, systematic reviews, and meta-analyses. Although its efficacy cannot be doubted, the significance of Direct Instruction's impact may be misunderstood. We attempt to clarify the importance of Direct Instruction with help from the binomial effect-size display. Binomial effect-size displays allow for intuitive and informative data-based decision making by clearly conveying the real-world importance of treatment outcomes through a juxtaposition of the relative proportions of success. The limitations of analyzing effect sizes in absolute terms are discussed. Using the binomial effect-size display as a framework, we present a series of dichotomies in an attempt to answer the question: Just how effective is Direct Instruction?

Keywords Direct instruction · Binomial effect size display · Effective schools · Effect size

You would be hard-pressed to find an educator who does not have an opinion on Direct Instruction (DI). Whether they tried it in their classroom and found that previously "uneducable" students were now learning or heard about it as a strawman for "drill and kill" techniques that diminish the "love of learning," everyone seems to love DI or hate DI (Pondiscio, 2018). Despite the philosophical arguments, there can be little doubt that DI works.

Direct Instruction is perhaps the most heavily researched educational model (Borman et al., 2003; Hattie, 2009). More than 500 individual research reports have been

✉ Lee Mason
  lee.mason@cookchildrens.org

[1] Child Study Center, Cook Children's Health Care System, 1300 West Lancaster Avenue, Fort Worth, TX 76102, USA

[2] ANSERS Institute, Texas Christian University, Fort Worth, USA

[3] Department of Behavior Analysis, University of North Texas, Fort Worth, USA

identified on DI, and an entire journal was once dedicated to its exploration (Stockard et al., 2020). This mountain of data has been analyzed, summarized, and scrutinized elsewhere, and it is not our intent to provide another systematic review. Rather, we provide some commentary on the vast amount of research on DI and offer some perspective on the interpretation of their results. The love/hate divergence is the first in a series of dichotomies that we present to help clarify some of the ambiguities that are inherent with complex statistical analyses. Because, regardless of which side of the aisle you stand, everyone should have access to the basic facts.

Direct Instruction has been lauded for its careful instructional programming to condition abstract stimulus control (i.e., teaching "big ideas"; Slocum, 2004). It has also been criticized for leading to student overdependence, the use of rigid and inflexible teaching methods, and emphasizing fact accumulation at the expense of critical thinking skills (Edwards, 1981). The polarizing qualities of DI cannot merely stem from its efficacy, because it would be difficult to refute the data demonstrating the variety of skills that can be taught through DI. Table 1 provides a sample of the available data demonstrating the effectiveness of DI across students and academic content.

Direct Instruction—as developed by Siegfried Engelmann, Carl Bereiter, Douglas Carnine, Jean Osborn, Wesley Becker, and colleagues—was initially a preschool program for children from communities of low socioeconomic status. Instructional periods for these children were only 20–30 min long and focused exclusively on math and reading. The teacher's instructions were carefully programed and sequenced, ensuring mastery of prerequisites skills to eliminate ambiguity when introducing more advanced concepts. Children in these early studies showed marked improvements, encouraging further research with this mode of instruction (Bereiter & Engelmann, 1966).

Direct Instruction was then expanded to include math, reading, and language for populations outside of preschools. This model became known as the *Direct Instruction System for Teaching Arithmetic and Reading* (*DISTAR*; Engelmann, 2007; Wood, 2014). Early success led to DI's inclusion in Project Follow Through, a longitudinal study from 1968 to 1977 that compared 20 educational programs and school-based interventions across some 700,000 children of low socioeconomic status. Project Follow Through found that DI was the only intervention with a "significant positive impact on all of the outcome measures" (Stockard et al., 2018, p. 482). Although the benefits of DI were clear, it was (and continues to be) criticized for being dogmatic, utilitarian, and authoritarian, with opponents claiming that its tightly structured scope and sequence leave little room for teacher and student creativity (Hattie, 2009).

The academic impact of DI was largely suppressed by the reviewers of Project Follow Through who recommended every model of purported benefit to students (e.g., along measures of self-concept, attitude, mental or physical health) or teachers (e.g., instructional fidelity). Watkins (1997) laments that, "it does the public little good to 'identify' little-used methods" (p. 44). Nonetheless, DI programs have been developed for teaching a wider variety of curriculum content to increasingly diverse student populations.

Direct Instruction versus direct instruction

Within the literature, discriminative control is established between Direct Instruction ("big DI") and direct instruction ("little di"). Big DI, as developed by

Table 1 Effect sizes obtained in a representative sample of direct instruction studies: 1972–2011

Chronological citations	Participants	Content	Effect size (d)
Haring and Krug (1975)	Elementary Special Education	Reading	0.64
Maggs and Morath (1976)	Elementary Special Education	Language	1.80
Summerell and Brannigan (1977)	Elementary Special Education	Reading	0.49
Richardson et al. (1978)	Elementary Special Education	Reading	0.09
Lloyd et al. (1980)	Elementary Special Education	Multiple	0.75
Stein and Goldman (1980)	Elementary Special Education	Reading	0.75
Lloyd et al. (1981)	Elementary Special Education	Multiple	0.76
Bereiter & Kurland (1981-1982)	Elementary General Education	Multiple	1.40
Beveridge and Jerrams (1981)	Elementary General Education	Language	-0.41
Robinson and Hesse (1981)	Elementary General Education	Spelling	2.00
Gregory et al. (1982)	Secondary Special Education	Reading	1.68
Lewis (1982)	Elementary Special Education	Reading	0.05
Becker and Gersten (1982)	Elementary General Education	Multiple	0.25
Branwhite (1983)	Elementary Special Education	Reading	1.65
Fielding et al. (1983)	Secondary General Education	Legal	0.35
Walker et al. (1983)	Elementary Special Education	Social Skills	0.97
Lum and Morton (1984)	Elementary General Education	Spelling	0.32
Meyer (1984)	Elementary General Education	Multiple	0.71
Cole and Dale (1986)	Elementary Special Education	Language	0.13
Kelly et al. (1986)	Secondary Special Education	Math	1.32
Woodward et al. (1986)	Secondary Special Education	Math	0.52
Darch and Kame'enui (1987)	Elementary Special Education	Reading	1.20
Darch et al. (1987)	Elementary Special Education	Multiple	0.61
Hasselbring et al. (1987–1988)	Elementary General Education	Math	1.20
Woodward et al. (1988)	Secondary Special Education	Health	1.60
Cole et al. (1989)	Elementary Special Education	Multiple	-0.14
Moore and Carnine (1989)	Secondary General Education	Math	0.65
Sexton (1989)	Elementary Special Education	Reading	0.54
Kelly et al. (1990)	Secondary Special Education	Math	1.11
Cole et al. (1991)	Elementary Special Education	Language	-0.14
Vitale and Romance (1992)	Secondary General Education	Science	2.44
Cole et al. (1993)	Elementary Special Education	Language	0.13
Kitz and Thorpe (1995)	Secondary Special Education	Math	1.65
Tarver and Jung (1995)	Elementary General Education	Math	0.78
Crawford and Snider (2000)	Elementary General Education	Math	1.15
Crowe et al. (2009)	Elementary General Education	Reading	0.44
Kamps et al. (2008)	Elementary General Education	Reading	0.59
Stockard (2010)	Elementary General Education	Math	0.29
Stockard (2011)	Elementary General Education	Reading	0.31

Note. Constructed from data provided by Adams and Engelmann (1996) and Barbash (2012). For a more comprehensive list of DI outcomes, see Stockard et al. (2020). The 39 studies have a simple mean of $d = 0.79$, ranging from $d = -0.41$ to $d = 2.44$. Participants in these studies came from both general and special education classrooms, across all grade levels. Curriculum content included reading ($n = 12$), math ($n = 9$), language ($n = 5$), spelling ($n = 2$), science ($n = 1$), health ($n = 1$), social skills ($n = 1$), legal concepts ($n = 1$), and multiple domains ($n = 7$)

Engelmann and colleagues, is centered on a carefully crafted model of instruction focused on teaching concept formation (Cooper et al., 2020; Engelmann & Carnine, 1992). These instructions have been specifically sequenced, scripted, and programmed to minimize obfuscation, and have demonstrated tremendous success at teaching specified objectives when compared to other modes of instruction (Kennedy, 1978; Stebbins et al., 1977).

Little di, on the other hand, refers to a broader repertoire of teacher behavior that incorporates elements of systematic or explicit instruction. Little di is often exemplified by the work of Rosenshine (1976, 2012) who identified a set of teacher variables that yield significant academic improvements for students, including: (1) structuring materials, (2) using clear and concise instructions, (3) promoting active student responding, (4) providing immediate feedback contingent on student answers, and (5) minimizing free time. As with big DI, little di incorporates systematically sequenced instructions along with gradual increases in rigor and difficulty. Centered on direct measures of student behavior, relevant outcome data, and systematic measurement practices, the combination of these instructional technologies have produced favorable cost-benefit ratios when compared with traditional teaching practices (Binder & Watkins, 1990; Lipsey et al., 2012).

Both big DI and little di focus on explicit and systematic instruction. The primary distinction between the two is that little di refers to a set of teacher behaviors, whereas big DI (hereafter, simply "DI") refers to curricular programs in the Engelmann tradition (e.g., *DISTAR*, *Reading Mastery*, *Connecting Math Concepts*, *Language for Learning*) and the teacher–student interactions that put the programs in play (e.g., fast paced scripted lessons, choral responding, error correction; Rolf & Slocum, 2021; Slocum & Rolf, 2021).

Effect Sizes

Effect sizes, like those listed in Table 1, measure the magnitude of a phenomenon, and are commonly reported using the indices for standardized mean difference (d), correlation coefficient (r), or coefficient of determination (r^2; Reimer & Russell, 2017). When an effect size is expressed as the difference between means in a standard score (i.e., z-value), it represents the number of standard deviations between experimental and control groups. Assuming groups of equal sample size, these various indices may be converted from one to another. For example, by coding group membership with a binary variable (e.g., "0" for the control group, and "1" for the experimental group), we can then convert d to r using the following formula:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

Likewise, we could then calculate the coefficient of determination (r^2) by squaring the correlation coefficient to measure the proportion of variance (or percentage of variance if we then multiply by 100) in the dependent variable that can be predicted by the independent variable.

The standard convention for describing d and r includes the terms "small" ($d = 0.2$, $r = .1$), "medium" ($d = 0.5$, $r = .3$), and "large" ($d = 0.8$, $r = .5$). Cohen (1988) used the average height of women as the context for this interpretation. A "small" effect might be interpreted as the difference between the heights of 15- and 16-year-old females in the United States. A "medium" effect was visible to the naked eye, and could be interpreted as the difference between the heights of 14- and 18-year-old women. A "large" effect is grossly perceptible such as the difference between the heights of 13- and 18-year-old women.

However, Cohen (1988) also acknowledged the danger of labeling a change in magnitude as "small," "medium," and "large" devoid of context, noting that we can only interpret the effectiveness of a particular intervention in relation to another intervention that seeks to produce the same effect, along with their relative costs and benefits. As a more tractable interpretation for evaluating educational interventions, Glass et al. (1981) proposed that an effect size of 1.00 corresponds to the difference of about a year of schooling on the standardized test scores of elementary school students. Coe (2002) argued that 0.60 is a more accurate estimate of this phenomenon. In other words, an average third-grade student at pretest who received an intervention with an effect size of 0.60 would perform at the average fourth-grade level at posttest.

Binomial Effect-Size Display

Statistically laden effect sizes like d and r are challenging to interpret, especially when there is no indicator of which effect-size index to use when working with multiple datasets (Randolph & Edmondson, 2005; Rosenthal, 1991). For example, in their meta-analysis of more than 500 individual studies on DI, Stockard et al. (2020) used mixed-model regressions to identify multiple effect sizes across various dimensions (e.g., grade level, subject matter, educational setting), along with more extensive analyses and estimates of effectiveness with different levels of exposure and fidelity. Answering the seemingly simple question of "Is DI effective?" required no less than 290 pages (with an additional 383 pages of supplemental material available online). The data provided by Stockard and colleagues are so immersive that the nonstatistician cannot see the wood for the trees. As Lipsey et al. (2012) note,

> The situation is not unusual—the native statistical representations of the findings of studies of intervention effects often provide little insight into the practical magnitude and meaning of those effects. To communicate that important information to researchers, practitioners, and policymakers, those statistical representations must be translated into some form that makes their practical significance easier to infer. (p. 1)

Though we cannot argue with Stockard et al.'s conclusion on the use of DI, "Our students deserve no less" (p. 165), we also cannot help but wonder whether they might have taken a more direct route to reach it.

To help clarify the real-world importance of an intervention, Rosenthal and Rubin (1982) suggested reducing continuous measures to a simple dichotomy. For example, we may select a given measure—such as the mean or median—and classify each value

that falls below that measure as a "failure" (or "0") and each value above it as a "success" (or "1"). We can then calculate r as the difference in proportions in each category (Coe, 2002). This conversion, called the binomial effect-size display (BESD) allows for more intuitive and informative data-based decision making by clearly conveying the real-world importance of treatment effects.

The BESD helps to interpret exceedingly complex data, which may be useful when evaluating DI outcomes because it is "(a) easily understood by researchers, students, and laypersons; (b) applicable in a wide variety of contexts; and (c) conveniently computed" (Rosenthal & Rubin, 1982, p. 166). In particular, the BESD may be more pragmatic than other effect-size indices, increasing the likelihood that parents and elected officials will select schools, teachers, and curriculum based on empirically supported evidence.

The BESD is premised upon three primary assumptions. Foremost, we assume that samples for both experimental and control groups follow a normal distribution. The sample that forms this normal distribution is better described as a density curve along a continuum of magnitude ranging from minimal to maximal effectiveness. The interpretation of effect sizes is sensitive to violations of this assumption, and it may be difficult to make an accurate comparison between an effect size based on a normal distribution and one derived from a nonnormal distribution (Coe, 2002).

Second, the BESD requires that we draw an imaginary line at the median value—which, in this case, is also the mean—and group our sample into one of two categories (see Fig. 1). The values that fall to the right of this imaginary line are above the mean, and therefore might be labeled "good," "effective," or a "success." The values that fall to the left of this imaginary line are below the mean, and therefore might be labeled "bad," "ineffective," or a "failure."

Finally, we must assume the use of a binary dependent variable upon which we can classify an outcome as a success or failure. That is, either the response contacted reinforcement or it did not. Our concern is only that the response meets this minimal criterion, and—as before—without regard for the magnitude of any individual effect. These overlapping dichotomies of success and failure result in a Punnett square through which we can examine a given effect-size estimate.

Life vs Death

As with other applications of human behavior, selectionism may serve as an appropriate framework for demonstrating the BESD. Table 2 shows the BESD for a hypothetical medical treatment that increases a patient's rate of survival by 1%. An $r = .10$ is the very definition of "small," but still considered important in the context of a life-saving intervention. Cohen (1988) notes that "'Death' tends to concentrate the mind" (p. 534). In absolute terms, an r^2 of .01 is undeniably "small," but when it represents a 10% increase in the rate of survival, our hypothetical patient might consider it large: "*alive*, mind you!" (p. 534).

Thus, a difference between success rates of .55 and .45 yields $r = .10$, and accounts for only 1% of the variance in the dependent variable as a function of the independent variable. However, effects that are operationally defined as "small" cannot be
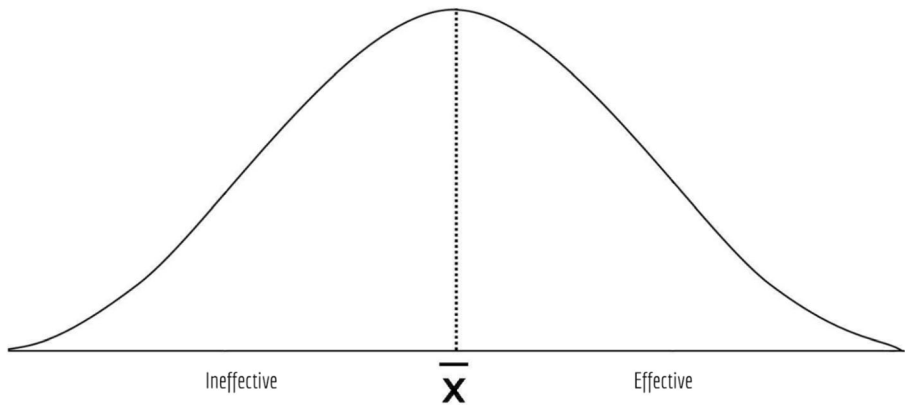
Fig. 1 A normal distribution showing the dichotomous categorization of effective (right) and ineffective (left) practices

overlooked when it marginally increases the rate of survival. Even marginal differences have the potential to shift an entire population (Baum, 2017; Rachlin, 1991).

Rosenthal (1990) described two occasions on which randomized controlled trials on the use of medications to prevent heart attacks were ended prematurely. In both cases, the results from the experimental group were so favorable that the research teams for each study determined it would be unethical to continue to withhold the life-saving drugs from the patients in the control group. The *r*s for these two independent studies were both well below the .10 threshold, with $r = .04$ (propranolol) and $r = .034$ (aspirin). As Rosenthal explained,

> Behavioral researchers are not used to thinking of *r*s of .04 as reflecting effect sizes of practical importance. But when we think of an *r* of .04 as reflecting a 4% decrease in heart attacks, the interpretation given *r* in a BESD, the *r* does not appear to be quite so small, especially if we can count ourselves among the 4 per 100 who manage to survive. (p. 775)

Table 3 provides a conversion from Cohen's *d* effect sizes ranging from zero to three in intervals of 0.10, to their corresponding *r* and r^2 values. The differences in success rates increase systematically along with the effect sizes.

Table 2 BESD for a hypothetical medical treatment that accounts for 1% of variance in survival

Effect Size	Condition	Outcome %		Total
		Alive	Dead	
$r = .10$	Treatment	55	45	100
$r^2 = .01$	Control	45	55	100
Total		100	100	200

Note. Constructed from data provided by Cohen (1988). BESD = Binomial effect-size display. "*r*" stands for the Pearson product-moment correlation coefficient. "r^2" stands for the coefficient of determination

As can be seen, the difference in success rate is always identical to *r*. Cohen (1988) explains that "the fact that the difference in proportions equals the *r* is not a coincidence, but a necessity when the table is symmetrical (i.e., when the two values are equidistant from .50"; p. 534). Even a *d* as small as 0.1, accounting for less than 1% of the variance, increases the probability of selection from 48% to 53%. As a result, the chance of survival—whether species, individual, or behavior—has increased by 5%.

Effective Schools vs Ineffective Schools

"Small" effects become even more important when they are cumulative (Cohen, 1988). More than a one-time intervention, education is designed to last 12 or more years in duration. Coe (2002) explains that few educational interventions have effects that are anything other than "small," due to the wide variation found within the population as a whole. Barbash (2012) and Tallmadge (1977) agreed that *d* = 0.25 is the conventional threshold to indicate an educationally significant outcome, though Lipsey et al. (2012) felt that even this was too high. (In contrast, recall that the average effect size for DI in Table 1 is *d* = 0.79.)

Despite the longitudinal efforts of our educational system, Coleman et al. (1966) found that schooling accounts for only 10% of student achievement, with the other 90% attributed to students' home environment. It would seem that Hart and Risley (1995) were correct! However, we can once again employ a BESD to clarify the importance of 10% variance (r^2 = .10). Table 4 shows the impact of a mere "moderate" effect size on the potential for student success.

Although Coleman et al.'s (1966) findings state that schools only account for 10% of student achievement, the BESD for this finding paints a drastically different picture. The success rate for students increases from 34.3% to 65.8% when they are placed in an effective school. In other words, effective schools benefit an additional 32 per 100 students, regardless of the background from which they step into the classroom. Almost two thirds of the students in an effective school are likely to succeed, compared to just over one third of students in an ineffective school.

In contrast to the findings by Coleman et al. (1966), more recent findings conclude that schools may actually account for as much as 20% of variance in student achievement, with the remaining 80% attributed to their background environment. According to a meta-analysis by Marzano (2003), the use of verbal acuity as Coleman et al.'s primary dependent measure underestimates the impact of schools. Table 5 shows the revised impact of schools on the potential for student success.

Although Marzano's (2003) findings suggest that schools account for an adjusted 20% of student achievement, the BESD for this finding demonstrates that a student's success rate increases from 27.6% to 72.4% when placed in an effective school. In other words, effective schools benefit an additional 45 per 100 students, regardless of the background from which they step into the classroom. Almost three fourths of the students in an effective school are likely to succeed, compared to just over one quarter of students in an ineffective school.

Table 3 BESD corresponding to various values of d, r, and r²

d	r	r²	Success rate increased		Differences in success rates
			To	From	
0.00	0.00	0.00	0.50	0.50	0.00
0.10	0.05	0.00	0.48	0.53	0.05
0.20	0.10	0.01	0.45	0.55	0.10
0.30	0.15	0.02	0.43	0.57	0.15
0.40	0.20	0.04	0.40	0.60	0.20
0.50	0.24	0.06	0.38	0.62	0.24
0.60	0.29	0.08	0.36	0.64	0.29
0.70	0.33	0.11	0.34	0.67	0.33
0.80	0.37	0.14	0.31	0.69	0.37
0.90	0.41	0.17	0.30	0.71	0.41
1.00	0.45	0.20	0.28	0.72	0.45
1.10	0.48	0.23	0.26	0.74	0.48
1.20	0.51	0.27	0.24	0.76	0.51
1.30	0.55	0.30	0.23	0.77	0.55
1.40	0.57	0.33	0.21	0.79	0.57
1.50	0.60	0.36	0.20	0.80	0.60
1.60	0.63	0.39	0.19	0.81	0.63
1.70	0.65	0.42	0.18	0.82	0.65
1.80	0.67	0.45	0.17	0.83	0.67
1.90	0.69	0.47	0.16	0.84	0.69
2.00	0.71	0.50	0.15	0.85	0.71
2.10	0.72	0.52	0.14	0.86	0.72
2.20	0.74	0.55	0.13	0.87	0.74
2.30	0.76	0.57	0.12	0.88	0.76
2.40	0.77	0.59	0.12	0.88	0.77
2.50	0.78	0.61	0.11	0.89	0.78
2.60	0.79	0.63	0.10	0.90	0.79
2.70	0.80	0.65	0.10	0.90	0.80
2.80	0.81	0.66	0.09	0.91	0.81
2.90	0.82	0.68	0.09	0.91	0.82
3.00	0.83	0.69	0.08	0.92	0.83

Note. BESD = Binomial effect-size display. "*d*" = standardized mean difference. "*r*" = Pearson product-moment correlation coefficient. "*r²*" = coefficient of determination

Direct Instruction versus Other Models

The model provided by Marzano (2003) represents the absolute best that schools can offer to students. Effective schools can increase the academic achievement for almost half of their student population, regardless of socioeconomic status, English proficiency, and other risk factors.

Table 4 BESD that accounts for 10% of variance in student achievement

Effect Size	Condition	Outcome %		Total
		Success	Failure	
$r = .316$	Effective schools	65.8	34.2	100
$r^2 = .10$	Ineffective schools	34.2	65.8	100
Total		100	100	200

Note. Constructed from data provided by Coleman et al. (1966). BESD = Binomial effect-size display. "*r*" = Pearson product-moment correlation coefficient. "r^2" = coefficient of determination

Since its inception, DI has focused on teaching students from academically disadvantaged backgrounds. By providing teachers with an organized curriculum and explicit instruction, DI has been shown to effectively control for the outside variables that have traditionally prohibited academic achievement. This includes: (1) student-level factors, such as frequent tardiness or absences; (2) parent-level factors, such as little interest in child's educational outcomes; and (3) teacher-level factors, including lack of focus on basic skills (Butler, 2020).

Adams and Engelmann (1996) reviewed the then current literature on DI, and found an overall effect size per variable of $d = 0.97$, across 173 total comparisons (72 general education and 101 special education). By conventional standards, this would be considered a "large" magnitude of effect. Using the formula for converting d to r above, we can then create a BESD for DI.

Table 6 shows that $d = 0.97$ converts to $r = .436$, which falls within the "medium" range between .3 and .5. Already we note a discrepancy in the interpretation of effect size based on the different indices. However, the $r^2 = .19$ tells us that DI accounts for 19% of the variance in student achievement.

Adams and Engelmann's (1996) 19% of variance in student achievement approximates Marzano's (2003) 20% model of maximum school effectiveness. In other words, DI accounts for almost all of the difference in schooling. The BESD for this finding demonstrates that with the use of DI student success rates increase by 43.6%. In other words, DI benefits an additional 44 per 100 students, regardless of background. Compared to the other curricula that were examined in the meta-analysis, DI increased academic achievement from 28% to 72%.

Table 5 BESD that accounts for 20% of variance in student achievement

Effect Size	Condition	Outcome %		Total
		Success	Failure	
$r = .447$	Effective schools	72.4	27.6	100
$r^2 = .20$	Ineffective schools	27.6	72.4	100
Total		100	100	200

Note. Constructed from data provided by Marzano (2003). BESD = Binomial effect-size display. "*r*" = Pearson product-moment correlation coefficient. "r^2" = coefficient of determination

The conventional use of effect sizes provides thresholds by which we can refer to the size of a phenomenon as "small," "medium," "large," etc. As previously stated, Cohen (1988) acknowledged the danger of describing effect-size estimates in such terms, stressing that the effectiveness of a particular intervention can only be interpreted within the context of another intervention that seeks to produce the same effect. The use of absolute values to define "small" and "large" stems from a perspective of realism indicative of methodological behaviorism (Baum, 2017). However, such absolute thresholds force conceptual arguments about whether to interpret the same phenomenon as "large" (i.e., $d = 0.97$) or "medium" (i.e., $r = .436$).

In contrast, BESD provides a more pragmatic view of the magnitude of an experimental effect by placing the data in context. It is one thing to say that DI has a "large" effect. It is another to say that DI increases the rate of success by 44%.

Curriculum versus Instruction

Allow us to present one last dichotomy to clarify the distinction between curriculum and instruction. Although it is superficial, we can call this the same distinction between DI, the published curricula materials, and di, the process of systematic and explicit instruction. To demonstrate the importance of this distinction, we pose the following hypothetical scenario.

Imagine that we rank ordered all of the teachers in the United States according to effectiveness. Given that there are about 3.7 million teachers in the United States (National Center for Educational Statistics, 2019), we can assume that they approximate a normal distribution. Let us also imagine that we rank ordered all of the various curricula according to effectiveness, which approximates another normal distribution. As before, we will draw an imaginary line at the mean of each distribution, and call everything to the right of this imaginary line "good," and everything to the left of this imaginary line "bad." The result is a Punnett square showing the various combinations of teachers and curricula (see Table 7).

It is safe to assume that everyone wants their child in a classroom with a good teacher using a good curriculum, and that nobody wants their child in a classroom with a bad teacher using a bad curriculum. This leads us to the following question (visualized in Fig. 2): Would you rather place your child in a classroom with a good teacher using a bad curriculum, or a bad teacher using a good curriculum?

Table 6 BESD that accounts for 19% of variance in student achievement

Effect Size	Condition	Outcome %		Total
		Pass	Fail	
$r = .436$	Direct Instruction	71.8	28.2	100
$r^2 = .19$	Control	28.2	71.8	100
Total		100	100	200

Note. Constructed from data provided by Adams and Engelmann (1996). BESD = Binomial effect-size display. "r" = Pearson product-moment correlation coefficient. "r^2" = coefficient of determination

Can effective instruction overcome the deficits of a poor curriculum? Can a systematic curriculum make up for the faults of poor instruction? To answer these questions, Marzano et al. (2003) conducted a meta-analysis to tease out these effects. The results of this analysis are displayed in Fig. 3.

According to Marzano and colleagues, the student who begins at the 50th percentile and has an average teacher who uses an average curriculum, will still be at the 50th percentile at the end of 2 years. This student has learned enough to keep pace with other students in the same grade.

But what happens to the student who has an ineffective teacher that uses an ineffective curriculum? After 2 years of school, this student has dropped from the 50th percentile to the 3rd percentile. Although this student may have acquired some basic skills, their learning is so sporadic and unorganized that they have lost considerable ground compared to other students in the same grade.

Suppose that the student who enters at the 50th percentile is given an ineffective teacher who uses a curriculum with a strong evidence base. Despite the class-wide implementation of a research-validated program, after 2 years this student has fallen to the 37th percentile when measured against other students in the same grade.

In contrast to the previous scenarios, which featured poor instruction from the classroom teacher, the next two highlight the benefits of effective instruction. Consider the student who begins at the 50th percentile, and is fortunate enough to be assigned to an effective teacher who uses an evidence-based curriculum. After 2 years, this student now performs at the 96th percentile.

Finally, let us suppose the student who enters school at the 50th percentile is given an effective teacher who uses a curriculum without empirical support. After 2 years, the student would still have gained 13 percentile points, and rank at the 63rd percentile. The importance of effective instruction can easily be seen when we compare it against a poor curriculum. Marzano et al. (2003) found that even when using ineffective curricula, effective teachers can produce meaningful gains in student achievement (Table 8).

By juxtaposing the effects of curriculum and instruction, we see that instruction accounts for 6.7% of variance in student success. The BESD for this finding demonstrates that a student's success rate increases from 37% to 63% when placed in a classroom with effective instruction. In other words, effective teachers benefit an additional 26 per 100 students, compared to the use of effective curricula alone.

Understanding the substantial effect schools can have on student achievement highlights the importance of evidence-based classroom practices (Slocum et al., 2014; Smith, 2013). When displayed in a BESD, Marzano et al.'s (2003) data force us to reconsider the idea of evidence-based practices as off-the-shelf curricula.

Table 7 Punnett square for evaluating the intersection of curriculum and instruction

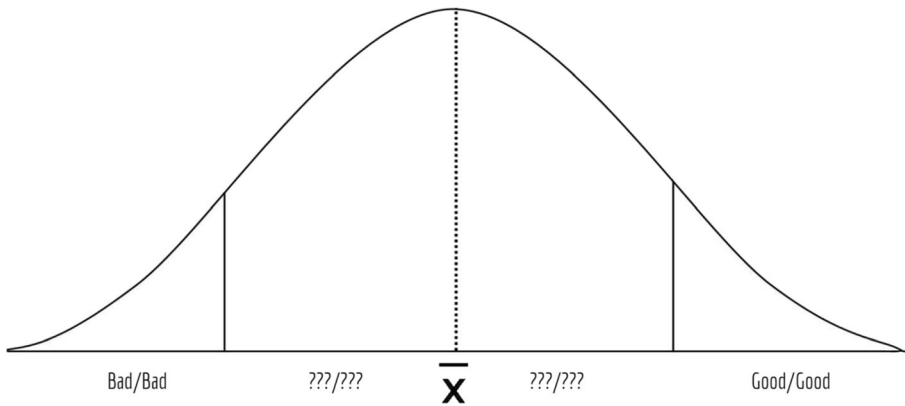| Curriculum	Instruction	
	Good	Bad
Good	Good/Good	Good/Bad
Bad	Bad/Good	Bad/Bad

Fig. 2 Would you rather your child were in a classroom with a good teacher using a bad curriculum, or a bad teacher using a good curriculum?

Government mandates for the use of evidence-based practices seemingly emphasize purchasing the right curricular products. As we can see from Table 8, this focus on curriculum alone decreases student success rates by 26%.

The term "evidence-based" is more often explanatory rather than descriptive, as in, the curriculum is evidence-based, and therefore effective. Selecting a program because it's marketed as research based and then implementing it without ongoing progress is at best the result of good marketing, and at worst malpractice. To paraphrase the tautological reasoning described by Vargas (2013), "How do you know the curriculum works? Because it is evidence based. How do you know it is evidence based? Because
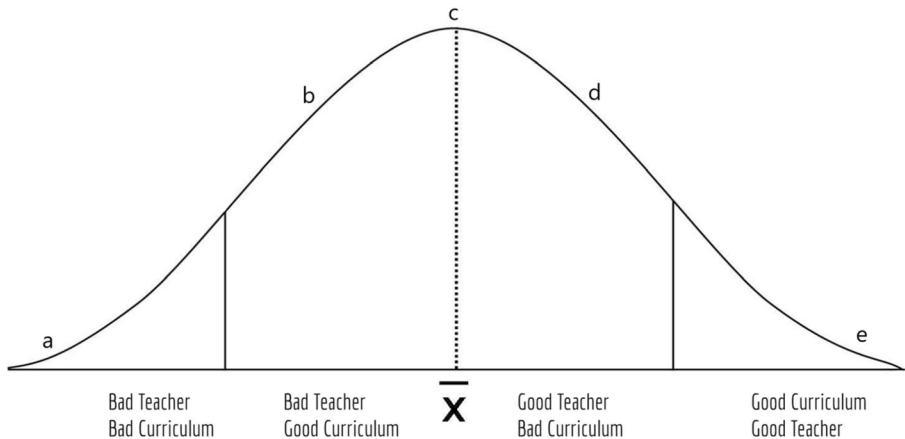

Fig. 3 A normal distribution showing the interaction of curriculum and instruction. *Note.* Constructed from data provided by Marzano et al. (2003). (1) A student who begins at the 50th percentile and receives poor instruction from a poor curriculum will rank at the 3rd percentile after 2 years. (2) A student who begins at the 50th percentile and receives poor instruction from a good curriculum will rank at the 37th percentile after 2 years. (3) A student who begins at the 50th percentile and receives average instruction from an average curriculum will remain at the 50th percentile after 2 years. (4) A student who begins at the 50th percentile and receives good instruction from a poor curriculum will rank at the 63rd percentile after 2 years. (5) A student who begins at the 50th percentile and receives good instruction from a good curriculum will rank at the 96th percentile after 2 years.

Table 8 BESD that accounts for 6.7% of variance in student achievement

Effect Size	Condition	Outcome %		Total
		Pass	Fail	
$r = .260$	Instruction	63	37	100
$r^2 = .07$	Curriculum	37	63	100
Total		100	100	200

Note. Constructed from data provided by Marzano et al. (2003). BESD = Binomial effect-size display. "*r*" = Pearson product-moment correlation coefficient. "*r²* " = coefficient of determination

it works." Moreover, research on fidelity of implementation tells us that maintaining treatment integrity is critical to student achievement (Mathews et al., 2018; Travers et al., 2016). What is "instruction" if not the implementation of a teaching program as designed?

Stockard et al. (2018) conducted a meta-regression of research on DI curricular materials that emphasized "a more precise measure of teacher preparation, including fidelity to all the various technical elements of the programs and training specific to the programs taught" (p. 501). The estimated effect size for the total sample was $d = 0.60$, which, by conventional standards, would be a "medium" effect. Using the formula for converting d to r above, we can then create a BESD for DI.

Table 9 shows that $d = 0.60$ converts to $r = .287$, accounting for 8.2% of the variance in student achievement. Stockard et al.'s (2018) findings on DI curricula surpass Marzano et al.'s (2003) 6.7% of variance resulting from effective instruction. Once again, DI bridges the gap between effective and ineffective instruction. The BESD for this finding demonstrates that the use of DI curricular materials increased student success rates from 35.6% to 64.4%. Compared to the other curricula that were examined in the meta-regression, DI increased academic achievement by 28.7%.

How Effective is DI?

Chubb and Moe (1990) assert that,

> All things being equal, a student in an effectively organized school achieves at least a half-year more than a student in an ineffectively organized school over the last two years of high school. If this difference can be extrapolated to the normal four-year high school experience, an effectively organized school may increase the achievement of its students by more than one full year. That is a substantial school effect indeed. (p. 140)

Over the past half-century, no other curriculum model has been more heavily researched than DI. Syntheses of this research tell us that by equipping teachers with DI curricular materials, and emphasizing high fidelity of implementation, the effects of schools can be substantial.

Table 9 BESD that accounts for 8.2% of variance in student achievement

Effect Size	Condition	Outcome %		Total
		Pass	Fail	
$r = .287$	Direct Instruction	64.4	35.6	100
$r^2 = .08$	Control	35.6	64.4	100
Total		100	100	200

Note. Constructed from data provided by Stockard et al. (2018). BESD = Binomial effect-size display. "r" = Pearson product-moment correlation coefficient. "r^2" = coefficient of determination

Glass et al. (1981) stated that an effective size of $d = 1.00$ equates to the difference of about a year of schooling. In their meta-analysis of DI, Adams and Engelmann (1996) found a similar overall effect size of $d = .97$. The BESD for Adams and Engelmann, which showed that DI accounted for 19% of the variance in student achievement, is consistent with Marzano's (2003) model showing that schools can account for as much as 20% of student success.

Coe (2002) argued that the difference of a year of school is closer to an effect size of $d = 0.60$. In their meta-regression on DI, Stockard et al. (2018) found an effect-size of $d = 0.60$. The BESD for Stockard et al., which showed that DI accounted for 8.2% of variance in student achievement, surpasses Marzano et al.'s (2003) finding that effective instruction can account for up to 6.7% of student success. Stockard et al.'s data also fit the model of effective schooling determined by Coleman et al. (1966).

Regardless of whether you want to describe the effect as "medium" ($d = 0.60$) or "large" ($d = 0.97$), the research shows that DI can effectively double the academic success rate of students: 2 years of gains across 1 year of instruction. Hunter and Schmidt (1990) describe the limitations of various effect-size indices such as d, r, and r^2, noting that,

> The percent of variance accounted for is statistically correct, but substantively erroneous. It leads to severe underestimates of the practical and theoretical significance of relationships between variables. . . . The problem with all percent variance accounted for indices of effect size is that variables that account for small percentages of the variance often have very important effects on the dependent variable. (pp. 199–200)

Although there are multiple ways to talk about effect sizes, any scientific approach to educational research is premised upon the identification of small, yet meaningful, variables. More than 50 years of research on DI have illustrated that rather than immediately solving all of our education needs, research-validated instructional practices point to the ongoing and systematic development of effective teacher behavior. In explaining longitudinal results, Ferster et al. (1975) declared that "The successful results of [education], even though we sometimes think of them in terms of a final dramatic outcome, actually occur in small increments of behavior" (p. 89). However, the nomenclature used to describe the magnitude of change fails to describe the pragmatic difference in student achievement resulting from DI.

In contrast to other effect-size indices, the BESD provides a real-world context for interpretation. Dichotomizing the independent variable in terms of experimental or control and dichotomizing the dependent variable in terms of effective or ineffective eliminates ambiguities to allow for easier and simpler data-based decision making. To reiterate the conclusion of Stockard et al. (2018), "Researchers and practitioners cannot afford to ignore the effectiveness [of] research on DI" (p. 503).

The BESD is not without limitations, however. McGraw (1991) argues that the BESD so distorts the original data that the exercise is misleading, whereas Strahan (1991) calls BESD a "what if" statistical technique. In addition, the assumptions of normality on which BESDs rely have been noted as a primary limitation (Hsu, 2004; Thompson & Schumacker, 1997). However, perhaps the most important limitation is raised by Skinner (1968), who warns us about the Idols of the School.

> The Idol of the Good Teacher is the belief that what a good teacher can do any teacher can do. Some teachers are, of course, unusually effective. . . . The Idol of the Good Student is the belief that what a good student can learn, any student can learn. Because they have superior ability or have been exposed to fortunate early environments, some students learn without being taught. (pp. 260–261)

The dichotomous nature of the BESD treats minimally effective teachers the same as maximally effective teachers, just as it treats minimally ineffective teachers the same as maximally ineffective teachers. One of Stockard et al. (2020) many findings was that even with limited instructional time and relatively low procedural fidelity, DI was more effective than other curricula. Student outcomes are even more dramatic with greater exposure and higher fidelity. In the classroom and other applied settings, outcomes are continuous rather than binary. Individual results may vary.

Of the many responsibilities with which the classroom teacher is charged, instructional design is not one of them (Pondiscio, 2018). Although we have posited that a good teacher can correct a bad curriculum, few teachers ever receive training in the systematic development and careful sequencing of content within the comprehensive scope of a given content area. Skinner (1968) noted that the central source error is the teacher's belief that their own personal experience is the primary basis of pedagogical wisdom. By teaching big ideas, scripting lessons, programming for generalization, and assessing mastery, DI shifts the teacher's focus from personal experience to student-generated data.

"There is no universal guideline or rule of thumb for judging the practical importance or substantive significance of a standardized effect size estimate for an intervention," argues Hill et al. (2008). "Instead one must develop empirical benchmarks of comparison that reflect the nature of the intervention being evaluated, its target population, and the outcome measure or measures being used" (p. 172). Despite having never been widely embraced or implemented, the large body of evidence supporting the effectiveness of DI has elevated it to education's empirical benchmark of comparison. Perhaps it is time to acknowledge that DI has situated itself as the Idol of the Good Curriculum.

Declarations

References

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

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

Baum, W. M. (2017). *Understanding behaviorism: Behavior, culture, and evolution* (3rd ed.). John Wiley & Sons.

Becker, W. C., & Gersten, R. (1982). A follow-up of Follow Through: The later effects of the direct instruction model on children in fifth and sixth grades. *American Educational Research Journal, 19*(1), 75–92. https://doi.org/10.3102/00028312019001075.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool*. Prentice-Hall.

Bereiter, C., & Kurland, M. (1981). A constructive look at follow through results. *Interchange, 12*(1), 1–22. https://doi.org/10.1007/bf01807392.

Beveridge, M., & Jerrams, A. (1981). Parental involvement in language development: An evaluation of a school-based parental assistance plan. *British Journal of Educational Psychology, 51*(3), 259–269. https://doi.org/10.1111/j.2044-8279.1981.tb02484.x.

Binder, C., & Watkins, C. L. (1990). Precision teaching and direct instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*(4), 74–96. https://doi.org/10.1111/j.1937-8327.1990.tb00478.x.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230. https://doi.org/10.3102/00346543073002125.

Branwhite, A. B. (1983). Boosting reading skills by direct instruction. *British Journal of Educational Psychology, 53*(3), 291–298. https://doi.org/10.1111/j.2044-8279.1983.tb02561.x.

Butler, K. (2020). The value of direct instruction for at-risk students. *Journal of Education & Development, 4*(2), 10–16. https://doi.org/10.20849/jed.v4i2.741.

Chubb, J. E., & Moe, T. M. (1990). *Politics, markets, and America's schools*. Brookings Institute.

Coe, R. (September, 2002). It's the effect size, stupid: What effect size is and why it is important. Paper presented at the *British Educational Research Association* annual conference: Exeter, England.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cole, K. N., & Dale, P. S. (1986). Direct language instruction and interactive language instruction with language delayed preschool children. *Journal of Speech, Language, & Hearing Research, 29*(2), 206–217. https://doi.org/10.1044/jshr.2902.206.

Cole, K. N., Dale, P. S., & Mills, P. E. (1991). Individual differences in language delayed children's responses to direct and interactive preschool instruction. *Topics in Early Childhood Special Education, 11*(1), 99–124. https://doi.org/10.1177/027112149101100110.

Cole, K. N., Dale, P. S., Mills, P. E., & Jenkins, J. R. (1993). Interaction between early intervention curricula and student characteristics. *Exceptional Children, 60*(1), 17–28. https://doi.org/10.1177/001440299306000103.

Cole, K. N., Mills, P. E., & Dale, P. S. (1989). A comparison of the effects of academic and cognitive curricula for young handicapped children one and two years post program. *Topics in Early Childhood Special Education, 9*(3), 110–127. https://doi.org/10.1177/027112148900900309.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F., & York, R. L. (1966). *Equality of educational opportunity* (Report No. OE-3800). National Center for Educational Statistics.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.

Crawford, D. B., & Snider, V. E. (2000). Effective mathematics instruction: The importance of curriculum. *Education and Treatment of Children, 23*(2), 122–142 https://www.jstor.org/stable/42940521?seq=1.

Crowe, E. C., Connor, C. M., & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47*(3), 187–214. https://doi.org/10.1016/j.jsp.2009.02.002.

Darch, C., Gersten, R., & Taylor, R. (1987). Evaluation of the Williamsburg County direct instruction program: Factors leading to success in rural elementary programs. *Research in Rural Education, 4*, 111–118 https://jrre.psu.edu/sites/default/files/2019-07/4-3_2.pdf.

Darch, C., & Kame'enui, E. J. (1987). Teaching LD students critical reading skills: A systematic replication. *Learning Disability Quarterly, 10*(2), 82–91. https://doi.org/10.2307/1510215.

Edwards, C. H. (1981). A second look at direct instruction. *High School Journal, 64*(4), 166–169 https://www.jstor.org/stable/40365849.

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

Engelmann, S., & Carnine, D. (1992). *Theory of instructional design: Principles and applications*. Irvington Press.

Ferster, C. B., Culbertson, S., & Boren, M. C. (1975). *Behavior principles* (2nd ed.). Prentice-Hall.

Fielding, G. D., Kame'enui, E., & Gersten, R. (1983). A comparison of an inquiry and a direct instruction approach to teaching legal concepts and applications to secondary school students. *Journal of Educational Research, 76*(5), 287–293. https://doi.org/10.1080/00220671.1983.10885468.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Sage.

Gregory, R. P., Hackney, C., & Gregory, N. M. (1982). Corrective Reading programme: An evaluation. *British Journal of Educational Psychology, 52*(1), 33–50. https://doi.org/10.1111/j.2044-8279.1982.tb02501.x.

Haring, N. G., & Krug, D. A. (1975). Evaluation of a program of systematic instructional procedures for extremely poor retarded children. *American Journal of Mental Deficiency, 79*(6), 627–631 https://pubmed.ncbi.nlm.nih.gov/1146852/.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes.

Hasselbring, T., Sherwood, R., Bransford, J., Fleenor, K., Griffith, D., & Goin, L. (1987). An evaluation of a level-one instructional videodisc program. *Journal of Educational Technology Systems, 16*(2), 151–169. https://doi.org/10.2190/br31-j510-cxm4-k41e.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x.

Hsu, L. M. (2004). Biases of success rate differences shown in binomial effect size displays. *Psychological Methods, 9*(2), 183–197. https://doi.org/10.1037/1082-989x.9.2.183.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications, Inc.

Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences for students most at risk in kindergarten. *Journal of Learning Disabilities, 41*(2), 101–114. https://doi.org/10.1177/0022219407313412.

Kelly, B., Carnine, D., Gersten, R., & Grossen, B. (1986). The effectiveness of videodisc instruction in teaching fractions to learning-disabled and remedial high school students. *Journal of Special Education Technology, 8*(2), 5–17. https://doi.org/10.1177/016264348600800202.

Kelly, B., Gersten, R., & Carnine, D. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students and high school students with learning disabilities. *Journal of Learning Disabilities, 23*(1), 23–29. https://doi.org/10.1177/002221949002300108.

Kennedy, M. M. (1978). *Findings from the Follow Through planned variation study*. U.S. Office of Education.

Kitz, W. R., & Thorpe, H. W. (1995). A comparison of the effectiveness of videodisc and traditional algebra instruction for college-age students with learning disabilities. *Remedial & Special Education, 16*(5), 295–306. https://doi.org/10.1177/074193259501600506.

Lewis, A. (1982). An experimental evaluation of a direct instruction programme (corrective reading) with remedial readers in a comprehensive school. *Educational Psychology, 2*(2), 121–135. https://doi.org/10.1080/0144341820020202.

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.* (NCSER 2013-3000). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncser/

Lloyd, J., Cullinan, D., Heins, E. D., & Epstein, M. H. (1980). Direct instruction: Effects on oral and written language comprehension. *Learning Disability Quarterly, 3*(4), 70–76. https://doi.org/10.2307/1510677.

Lloyd, J., Epstein, M. H., & Cullinan, D. (1981). Direct Instruction for learning disabilities. In J. Gottlieb & S. S. Strichart (Eds.), *Developmental theory and research in learning disabilities (pp. 41–45)*. University Park Press.

Lum, T., & Morton, L. L. (1984). Direct Instruction in spelling increases gain in spelling and reading skills. *Special Education in Canada, 58*(2), 41–45 https://eric.ed.gov/?id=EJ299608.

Maggs, A., & Morath, P. (1976). Effects of direct verbal instruction on intellectual development of institutionalized moderately retarded children: A 2-year study. *Journal of Special Education, 10*(4), 357–363. https://doi.org/10.1177/002246697601000404.

Marzano, R. J. (2003). *What works in schools: Translating research into action.* Association for Supervision and Curriculum Development.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher.* Association for Supervision & Curriculum Development.

Mathews, H. M., Hirsch, S. E., & Therrien, W. J. (2018). Becoming critical consumers of research: Understanding replication. *Intervention in School & Clinic, 53*(5), 267–275. https://doi.org/10.1177/1053451217736863.

McGraw, K. O. (1991). Problems with the BESD: A comment on Rosenthal's "How are we doing in soft psychology?". *American Psychologist, 46*(10), 1084–1086. https://doi.org/10.1037/0003-066x.46.10.1084.

Meyer, L. A. (1984). Long-term academic effects of the Direct Instruction Project Follow Through. *Elementary School Journal, 84*(4), 380–394. https://doi.org/10.1086/461371.

Moore, L. J., & Carnine, D. (1989). Evaluating curriculum design in the context of active teaching. *Remedial & Special Education, 10*(4), 28–37. https://doi.org/10.1177/074193258901000406.

National Center for Educational Statistics. (2019). *Digest of education statistics: 2018.* https://nces.ed.gov/programs/digest/d18/

Pondiscio, R. (2018). Meta-analysis confirms effectiveness of an old school approach: Direct Instruction. *Flypaper.* https://fordhaminstitute.org/national/commentary/direct-instruction-rodney-dangerfield-curriculum

Rachlin, H. (1991). *Introduction to modern behaviorism* (3rd ed.). W. H. Freeman.

Randolph, J. J., & Edmondson, R. S. (2005). Using the binomial effect size display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical Assessment, Research, & Evaluation, 10*, 1–7. https://doi.org/10.7275/zqwr-mx46.

Reimer, T., & Russell, T. (2017). Binomial effect size display. In M. Allen (Ed.), *The Sage encyclopedia of communication research methods* (p. 96). Sage. https://doi.org/10.4135/9781483381411.n38.

Richardson, E., DiBenedetto, B., Christ, A., Press, M., & Winsberg, B. G. (1978). An assessment of two methods for remediating reading deficiencies. *Reading Improvement, 15*(2), 82–95.

Robinson, J. W., & Hesse, K. D. (1981). A morphemically based spelling program' effect on spelling skills and spelling performance of seventh grade students. *Journal of Educational Research, 75*(1), 56–62. https://doi.org/10.1080/00220671.1981.10885356.

Rosenshine, B. (1976). Recent research on teaching behaviors and student achievement. *Journal of Teacher Education, 27*(1), 61–64. https://doi.org/10.1177/002248717602700115.

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*(1), 12–19 https://files.eric.ed.gov/fulltext/EJ971753.pdf.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775–777. https://doi.org/10.1037/0003-066x.45.6.775.

Rosenthal, R. (1991). Effect sizes: Pearson's correlation, its display via the BESD, and alternative indices. *American Psychologist, 46*(10), 1086–1087. https://doi.org/10.1037/0003-066x.46.10.1086.

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166–169. https://doi.org/10.1037/0022-0663.74.2.166.

Sexton, C. W. (1989). Effectiveness of the DISTAR Reading I program in developing first graders' language skills. *Journal of Educational Research, 82*(5), 289–293. https://doi.org/10.1080/00220671.1989.10885908.

Skinner, B. F. (1968). *The technology of teaching.* Appleton-Century-Crofts.

Rolf, K. R. & Slocum, T. (2021). Features of Direct Instruction: Interactive Lessons. *Behavior Analysis in Practice.*

Slocum, T. & Rolf, K. R. (2021). Features of Direct Instruction: Content analysis. *Behavior Analysis in Practice.*

Slocum, T. A. (2004). Direct instruction: The big ideas. In D. J. Moran & R. W. Malott (Eds.), *Evidence-based educational methods* (pp. 81–94). Academic Press. https://doi.org/10.1016/B978-012506041-7/50007-3.

Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. *The Behavior Analyst, 37*(1), 41–56. https://doi.org/10.1007/s40614-014-0005-2.

Smith, T. (2013). What is evidence-based behavior analysis? *The Behavior Analyst, 36*(1), 7–33. https://doi.org/10.1007/bf03392290.

Stebbins, L. B., St Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. (1977). *Education as experimentation: A planned variation model* (Vol. IV-A). ABT Associates.

Stein, C. L., & Goldman, J. (1980). Beginning reading instruction for children with minimal brain dysfunction. *Journal of Learning Disabilities, 13*(4), 52–55. https://doi.org/10.1177/002221948001300409.

Stockard, J. (2010). Promoting reading achievement and countering the "fourth-grade slump": The impact of direct instruction on reading achievement in fifth grade. *Journal of Education for Students Placed at Risk, 15*(3), 218–240. https://doi.org/10.1080/10824669.2010.495687.

Stockard, J. (2011). Increasing reading skills in rural areas: An analysis of three school districts. *Journal of Research in Rural Education, 26*(8), 1–19 https://jrre.psu.edu/sites/default/files/2019-08/26-8.pdf.

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

Strahan, R. F. (1991). Remarks on the binary effect size display. *American Psychologist, 46*(10), 1083–1084. https://doi.org/10.1037/0003-066x.46.10.1083.b.

Summerell, S., & Brannigan, G. G. (1977). Comparison of reading programs for children with low levels of reading readiness. *Perceptual & Motor Skills, 44*(3), 743–746. https://doi.org/10.2466/pms.1977.44.3.743.

Tallmadge, G. K. (1977). *The Joint Dissemination Review Panel ideabook.* National Institute of Education.

Tarver, S. G., & Jung, J. S. (1995). A comparison of mathematics achievement and mathematics attitudes of first and second graders instructed with either a discovery-learning mathematics curriculum or a direct instruction curriculum. *Effective School Practices, 14*(1), 49–57.

Thompson, K. N., & Schumacker, R. E. (1997). An evaluation of Rosenthal and Rubin's binomial effect size display. *Journal of Educational & Behavioral Statistics, 22*(1), 109–117. https://doi.org/10.3102/10769986022001109.

Travers, J. C., Cook, B. G., Therrien, W. J., & Coyne, M. D. (2016). Replication and special education. *Remedial & Special Education, 37*, 195–204. https://doi.org/10.1177/0741932516648462.

Vargas, J. S. (2013). *Behavior analysis for effective teaching* (2nd ed.). Routledge.

Vitale, M. R., & Romance, N. R. (1992). Using videodisk instruction in an elementary science methods course: Remediating science knowledge deficiencies and facilitating science teaching attitudes. *Journal of Research in Science Teaching, 29*(9), 915–928. https://doi.org/10.1002/tea.3660290903.

Walker, H. M., McConnell, S., Walker, J. L., Clarke, J. Y., Todis, B., Cohen, G., & Rankin, R. (1983). Initial analysis of the ACCEPTS curriculum: Efficacy of instructional and behavior management procedures for improving the social adjustment of handicapped children. *Analysis & Intervention in Developmental Disabilities, 3*(1), 105–127. https://doi.org/10.1016/0270-4684(83)90029-0.

Watkins, C. L. (1997). *Project Follow Through: A case study of contingencies influencing instructional practices of the educational establishment.* Cambridge Center for Behavioral Studies.

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

Woodward, J., Carnine, D., & Gersten, R. (1988). Teaching problem solving through computer simulations. *American Educational Research Journal, 25*(1), 72–86. https://doi.org/10.3102/00028312025001072.

Woodward, J., Carnine, D., Gersten, R., Gleason, M., Johnson, G., & Collins, M. (1986). Applying instructional design principles to CAI for mildly handicapped students: Four recently conducted studies. *Journal of Special Education Technology, 8*(1), 13–26. https://doi.org/10.1177/016264348600800103.