



Faultless Communication: The Heart and Soul of DI

Janet S. Twyman¹ 

Accepted: 16 July 2021/ Published online: 18 August 2021
© Association for Behavior Analysis International 2021

Abstract

We are in the midst of a global learning crisis. The National Center for Education Statistics (2019) reports that 65% of fourth- and 66% of eighth-grade students in the United States did *not* meet proficient standards for reading. A 2017 report from UNESCO reports that 6 out of 10 children worldwide do not achieve minimum proficiency in reading and mathematics. For far too many learners, instruction is riddled with confusion and ambiguity. Engelmann and Carnine's (1991) approach to improving learning is to design instruction that communicates one (and only one) logical interpretation by the learner. Called “faultless communication” this method can be used to teach learners a wide variety of concepts or skills and underpins all Direct Instruction programs. By reducing errors and misinterpretation, it maximizes learning for all students. To ensure effectiveness, the learner's performance is observed, and if necessary, the communication is continually redesigned until faultless (i.e., the learner learns). This “Theory of Instruction” is harmonious with behavior analysis and beneficial to anyone concerned with improving student learning—the heart and soul of good instruction.

Keywords education · instruction · instructional design · faultless communication · Direct Instruction

Let's begin with some data. In the latest “Nation's Report Card,” the National Center for Education Statistics (NCES, 2019) reports that 65% of fourth- and 66% of eighth-grade students in the United States did *not* meet proficient standards for reading. More concretely, two out of three fourth-graders were unable to integrate, interpret, or evaluate texts or apply their understanding of the text to draw conclusions. Two out of three eighth-graders could not reliably summarize main ideas and themes, make and support inferences about a text, connect parts of a text, or analyze text features. Proficiency matters. When proficient at a fourth-grade reading level, students should understand and glean information from text features as well as explain and draw

✉ Janet S. Twyman
jstwyman@gmail.com

¹ blast, 3075 South Morning Sky Drive, Cottonwood, AZ 86326, USA

conclusions from simple cause-and-effect relationships. When proficient at an eighth-grade reading level, students should be able to fully validate conclusions about content and support their judgment about the text (see National Assessment Governing Board, n.d., for a description of achievement levels and how they are determined).

The “Nation's Report Card,” also known as the National Assessment of Educational Progress (NAEP), is the largest national ongoing assessment of U.S. student knowledge across subject matter (in particular reading and math). NAEP assessments occur every two years, with 2019 as the most current at the time of this writing. Average national scores of student performance on the NAEP typically vary by only a point or two between test administrations, as they did in 2019 for fourth-grade reading (decreasing 1 point), fourth-grade math (increasing 1 point), and eighth-grade math (decreasing 1 point) compared to 2017. The eighth-grade reading scores are different. Not only were 66% of students not proficient, the national average dropped by 3 points, revealing the 2019 reading ability of a majority of fourth-grade students to be akin to that of students 10–20 years ago.

NAEP reading results have been generally stable across testing years, which could be a good thing if stable were not another word for stagnant. The 2019 average reading score for both fourth- and eighth-grade levels is not significantly different than a decade ago. Perhaps even worse, when compared to 2009, 2019 reading scores decreased for low-performing students whereas scores for high-performing students increased for both grade levels. In fact, NAEP fourth- and eighth-grade reading scores for low performing students (those in the bottom 10th percentile, arguably the ones that need the most help) have not consistently improved since 1992—when on average, 74% of all students were *below* proficient. Although that number decreased to 63% in 2017, an inexcusable number of students in the U.S. cannot read (NCES, 2019).

Likewise, NAEP average math scores have remained flat for the past decade. In comparison to 2009, 2019 scores for both fourth and eighth grades have either not significantly changed, or worse, have decreased for middle- and lower-performing students. At the same time, the average math scores between 2009 and 2019 have increased for high-performing students (those at the 90th percentile; NCES, 2019), showing that the achievement gap continues to widen. Let's again translate overall NAEP scores into a more salient narrative: For almost 30 years, we have tolerated at least two thirds of U.S. school-aged children to perform at below proficient standards, demonstrating only partial mastery of prerequisite knowledge and skills that are central to academic performance and success.

Education Globally

The United States is not alone in failing students. A 2017 report from UNESCO depicts a global learning crisis: Six out of 10 children worldwide do not achieve minimum proficiency in reading and mathematics. With global estimates of roughly 741 million elementary students, these data indicate that 56% of children of primary school age (6–11 years old) were not able to read or understand mathematics with proficiency. These percentages do not just represent children without access to schooling. Of the 387 million primary-age children unable to read proficiently, 262 million (68%) attended school. Of the 588 million total secondary students, 230 million adolescents of lower

secondary school age (12–14 years old) did not meet proficiency standards; of these 137 million (60%) attended school (De Brey et al., 2021; UNESCO, 2017).

A well-known international assessment administered of global student achievement is the PISA (Program for International Student Assessment, <https://www.oecd.org/pisa>), administered by the Organization for Economic Cooperation and Development (OECD). Every 3 years, PISA measures math, reading, and science skills of 15-year-old students and reports the results across six main levels of proficiency. Higher levels indicate involve tasks of increasing complexity with results reported as the share of the student population that reached each level. Level 2 is considered minimally proficient.

In 2018, 600,000 15-year-olds in the seventh grade or higher from 79 countries completed the PISA exam. Most countries saw little or no improvement in student performance. Only seven countries demonstrated significant score improvement since the first PISA administration in 2000 (Camera, 2019; also see OECD, 2019, for an in-depth analysis of the PISA assessment and the 2018 results). The percent of students demonstrating at least a minimum level of proficiency in reading ranged from close to 90% in parts of China, Estonia, and Singapore, to less than 10% in Cambodia, Senegal, and Zambia. Data for the minimum level of proficiency in mathematics reflected an even wider range of outcomes: 98% in parts of China and 2% in Zambia. Using averaged data across all countries, one in four 15-year-olds tested did not demonstrate the minimum level of proficiency (Level 2) in reading or mathematics. What does it mean for a 15-year-old to not meet the level of proficiency? According to OCED's director for education and skills, "When you don't reach Level 2 on the PISA test, that's a pretty dark subject for your educational future. That's the kind of reading skills you'd expect from a 10-year-old child" (cited in Camera, 2019, n.p.).

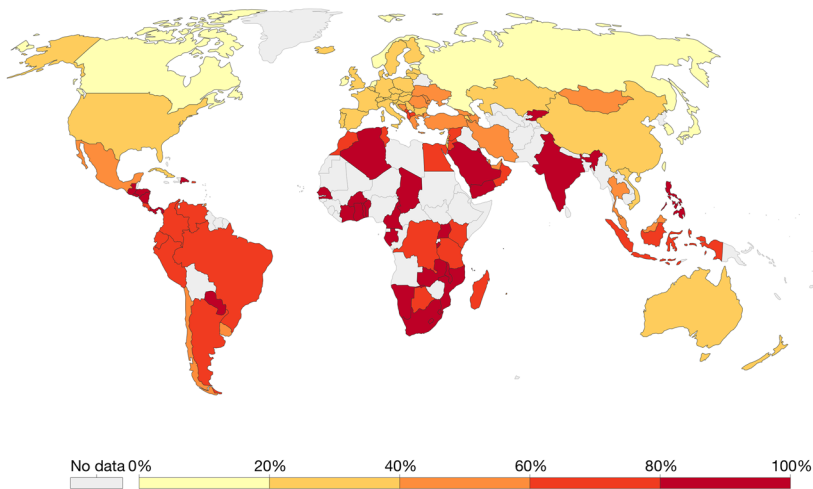
Using data from UNESCO Institute of Statistics (<http://data.uis.unesco.org>), World Bank EdStats (<http://datatopics.worldbank.org/education>), and OECD (<http://www.oecd.org>), Figure 1 represents, by country, the percentage of students with little or no learning outcomes across a 10-year period (2005–2015). Learning outcomes are defined by the percentage, or share, of students either achieving or not achieving a *minimum* proficiency benchmark (as was determined individually for each assessment). These outcomes were derived from standardized international and regional psychometric assessments aggregated across subjects (e.g., math, reading, science) and primary and secondary education levels (Roser et al., 2013; also see Koretz, 2008, for a detailed discussion of the assessment of learning outcomes). In Figure 1, darker shading indicates a greater percentage of students not meeting the minimum proficiency benchmark. Out of the 127 countries for which data are available, over half of the students in 71 of these countries do not meet minimum proficiency standards. Using these data, only 13 countries meet the barely laudable level of 20% or fewer students not meeting proficiency standards.

The Right to Effective Education

Education is basic human right. In Article 26 of the Universal Declaration of Human Rights, the United Nations (UN) declared education a fundamental right, suggesting that education is both important in itself and vital to the attainment of other human rights such as the right to work, fair wages, and an adequate standard of living including food, clothing, housing (Committee on Economic, Social and Cultural

Share of students achieving no or minimum learning outcomes (2005–2015)

Achievement outcomes come from standardised, psychometrically-robust international and regional tests. In order to maximise coverage, the most recent available data were used from the years 2005, 2010 and 2015.



Source: Angrist et al. (2019)

CC BY

Fig. 1 Poor Schooling is Pandemic. *Note:* This dataset was compiled to compare the government expenditure per primary student (\$PPP) with learning outcomes. Because of the patchy nature of the government expenditure data as obtained from the World Bank EdStats Dataset, the most recent available expenditure data were used whereas 2006 was used as a cut-off point. Higher proficiency benchmarks that represent the missing share of 100% are students that reached intermediate or advanced proficiency levels. <https://ourworldindata.org/grapher/share-of-students-achieving-no-or-minimum-learning-outcomes>. Image Source Our World in Data. CC BY 4.0

Rights, 1999; United Nations, 1948). Since the early 2000s, UNESCO has been advocating free primary education for all, based on the premise that even a basic education can provide children with the ability to improve life outcomes. As noted by Lee, “[I]f children receive basic primary education, they will likely be literate and numerate and will have the basic social and life skills necessary to secure a job, to be an active member of a peaceful community, and to have a fulfilling life” (Lee, 2013, p. 1). Even closer to home is the Association for Behavior Analysis International (ABAI) (1990) statement on Student Right to Effective Education (<https://www.abainternational.org/about-us/policies-and-positions/students-rights-to-effective-education>) which covers six areas educational entitlement for all students, including the recommendation that curriculum and instructional objectives be based on empirically validated sequences of instruction and measurable performance criteria that promote cumulative mastery and are of long-term value in the culture.

Nations around the world have made tremendous headway in making education available. In the 1970s, over a quarter of the world’s young children were not in school, yet by 2013 just one in ten children did not attend school (Roser & Ortiz-Ospina, 2013). Lee and Lee (2016) report that the percentage of primary-aged children enrolled in school across 111 countries has grown from less than 10% in 1870 to almost 96% in 2010. Primary education is now mandatory and freely provided most countries around the world.

This is progress. The increasing availability of schooling is a substantial achievement. However, schooling is not the same as learning. All children deserve an education that prepares them for decent employment and a fulfilling life in the 21st Century.

The Effects of Ineffective Schooling

Not learning to read in the early primary grades has at least two associated affects. Because learning is cumulative, without learning to read students will not be able to read to learn. Mastery of foundational skills in the early grades is strongly positively associated with later school performance. Data clearly show that children who cannot read by 3rd grade struggle to catch up, many continue to fall further behind (Moats, 2001). Once an achievement gap (the disparity in academic performance between groups of students) occurs, research shows it continues to widen (Reardon et al., 2014). In a longitudinal study of nearly 4,000 U.S. students (Hernandez, 2012) found that students who did not read proficiently by third grade were four times more likely to drop out of school than their proficient peers. For students who did not master reading skills at the basic level, the rate is almost six times greater.

Second, struggling students don't often get a chance to catch up. Often the pace of classroom instruction is determined curricula or unit, semester, or grade level expectations rather than by the pace of student learning. In many systems worldwide, teachers have no choice but to let some students fall behind. In many developing countries, pedagogical practices often consist of teachers using a "lecture" style, with relatively little opportunity for or expertise in differentiating instruction to support for the variation of student ability within a classroom (Damon et al., 2016).

The personal, societal, and global costs of a failing educational system are huge. Lower education levels reduce human capital, productivity, incomes, employability, and economic growth. Individuals without a basic education also tend to be less healthy and have less control over the trajectory of their lives (World Bank, 2018). Conversely, the benefits of a basic education for all are vast (UNESCO, 2017). If all children in school acquired basic reading skills, 171 million people could be lifted out of extreme poverty, equivalent to a 12% reduction in world impoverishment. Each additional year of schooling increases future earnings by approximately 10%; if all adults completed secondary education, world poverty would be cut in half.

Why Haven't Reforms Worked?

Governments and educational systems strive to raise the education level of their citizens, yet for the most part we have not seen consistent, broadly applicable transformations that produce success for all students (OECD, 2019). Why do so many countries find it so difficult to substantially raise the achievement level of the majority, if not all, of their students? Numerous factors are frequently cited: lack of resources, inadequate or misplaced funding, overcrowded classrooms, untested or ineffective curricula, ill-equipped, unqualified, or underpaid teachers, unprepared or unmotivated students, and a disconnect between what is learned in school and what is needed in life, are just a few (Furlong & Phillips, 2001).

The World Bank (2018) attributes the difficulty of improving education to 3 factors: (1) the poor learning outcomes themselves; (2) their related proximate or immediate

causes (such as student unpreparedness to learn, untrained or ineffectual teachers, and poor management or leadership); and (3) the more distal, deeper systemic causes (such as a lack of coherence within or across systems, disparate stakeholder interests, or the politicization of education).¹ Viewed from a systems perspective, failures of schooling should not be a surprise. It is inevitable that holders of a range of special interests influence what happens in schools.

Freire (1985) emphasized that every educational decision, at every level, is rooted within a particular ideological framework. This includes decisions as simple as whether students should sit in neat, linear rows facing the teacher or scattered around the room in small groups facing each, or whether they should be seated at all. Larger decisions allow ideologies to become reforms, sometimes with wide dissemination, such as English-only teaching, policies on student suspension or retention, adoption of particular curricula, or the interpretation of research results. Special interests are inherently at the heart of school reform² (Rossi et al., 1979). Cultural, societal, political, ideological, and other contextual forces continuously exert control that draws behavior away from student learning.

Many reforms have come, and many reforms have failed (Kliebard, 1988; Kozol, 1988; Payne, 2008). The most likely cause of disappointing results from the various reforms is...poorly conceived reforms (Hirsch, 2016a). Hirsch views these as ones that have been primarily structural in character such as smaller class size, longer school days, or school choice. Intangible should be added as another characteristic (e.g., resilience, grit, or sense of agency). Too often reform recommendations are subject to interpretation, such as “high expectations” and “positive culture,” and do not specify the research-based ways that “motivate” and “support” actually get done. Too often the lens of school improvement focuses on a variety of issues that, although important, seem to have little to do with *actual* learning (Watkins, 1995) [such as teacher collaboration (Leana, 2011) or multiculturalism (Nieto & Bode, 2009)]. And ultimately, all too often evidence-based instructional strategies coupled with specific actions are absent in analyses of school failure; what we know works is rarely broadly and systematically applied to educational curricula, instructional sequences, and teacher practices (Twyman, 2021; Watkins, 1995).

Ideological Roadblocks

We know effective teaching methods exist (Embry & Biglan, 2008), however they are not commonly used to improve education. Why is this? Obviously we want educational approaches that have empirically-proven positive outcomes related to what it is they are trying to teach. Shouldn't instructional programs be adopted based on the extent with

¹ The World Bank's analysis of immediate versus deeper systemic causes is suggestive of the last two levels in Skinner's (1981) analysis of selection: operant selection at the level of individual behavior and operant cultural selection at the level of practice; as well as Glenn's (1998) elaboration of the distinction between contingencies at the behavioral level (contingencies of reinforcement) and contingencies at the cultural level (called metacontingencies). In cultural selection individual organisms interact with the environment and what is replicated (over time or space) is a cultural practice. While not the focus of the current paper, an analysis of distal or deep system practices in education from an operant cultural selection/metacontingency framework would seem to be an interesting endeavor.

² The creation-evolution debate (Larson, 2010), Common Core Standards, or more currently, critical race theory are but a few examples.

which they improve educational achievement? They aren't (Detrich et al., 2007; Slavin, 2008; Watkins, 1995). Perhaps much of the problem lies in how the educational establishment views learning, and by default, instruction (Apple, 2007; Payne, 2008; Watkins, 1995). Some of the most harmful established views of learning include yoking education with "developmental ages" and thus supplying instruction appropriate to a child's age and disposition (Driver, 1978), emphasizing a child-centered, individualistic view of learning (Siegler et al., 2003), and the emphasis on learning styles (Willingham, 2005). These are harmful (the author acknowledges this is a strong word) because strong evidence indicates that they don't work (Hattie, 2009; Hirsch, 2016b; World Economic Forum, 2014). Yet these "theories" of educational still endure. As noted by Yoak (2008, p. 64): "Failure is not rooted in a lack of knowledge, but in the unwillingness or inability of people to see beyond the ideological divide." Skinner (1984) knew this as well. In "*The Shame of American Education*," he lamented the education establishment's inability to embrace the effective technology of Programmed Instruction, writing: "The theories of human behavior most often taught in schools of education—particularly cognitive psychology—stand in the way of this solution to the problem of American education..." (p. 947).

It is not that 60% of children worldwide cannot learn. Simply put, our schools, our curricula, our "established" practices aren't teaching them. We are not dealing with a failure to learn but with two failures. The first is that we are not broadly and consistently designing instruction that works for all learners all the time (or at least most learners most of the time), and the second is that when we do, few care (Hirsch, 2016b; Slavin, 2008; Watkins, 1995). Walker (2011) summarizes the first problem quite well:

In order to create the context for the change(s) in our educational system necessary to improve student outcomes on a broad scale, we are in need of a science of education as well as a science of instruction... What we do not need are more distal school reform initiatives (e.g., improving school climate, new models of deploying school resources, denying teachers tenure, and so forth). Instead, we desperately need more proximal solutions that impact the teaching-learning process such as the logical analysis and empirical evaluation of learning outcomes... Ultimately, we have to directly influence what transpires between teacher and student in the classroom on a daily basis if we wish to produce better student outcomes. (n.p.)

What follows connects the problem of students not learning well, enough, or at all, to a solution offered by Engelmann and colleagues. That connection is made refreshing clear in Engelmann's own words, in a paper by Heward et al. (2021). The need for proximal solutions, a logical analysis, and a way to directly influence what transpires between teacher and student in the classroom leads us to Direct Instruction (DI) and theory of instruction behind it (and meticulously described in a book by the same name). This paper takes the view that the research-informed content analysis and design of instruction, followed by the development of curricula through scientific formative, coupled with implementation fidelity, offers us our best chance at ameliorating our collective history of dismal outcomes in education. DI, and within it a particular design philosophy, gives us that chance.

Direct Instruction

It was within the context of national and worldwide sustained and pervasive failures in instruction, as well as the civil rights and other momentous social and political movements of the early 1960's (see Kantor, 1991), that Siegfried Engelmann turned to education. His early work testing advertising's influence on behavior as well as success teaching his own young sons' linear equations, ignited his passion to improve instruction. He was convinced that given the right instruction, all children could learn. Engelmann's premise was that learning is logical and that what or how much is learned depends on how logically it is taught. For a child to understand a triangle, a non-fiction essay, or the process of photosynthesis, they needed clear information about these concepts. Regardless of their age, starting point, or ability, students need to know what are the unchanging features that make something a thing. If a teacher's presentation generates only one meaning, all children will learn that meaning. Failures to learn could not be attributed to socio-economic status, gender, race, or peculiarities within a particular child, but they could be attributed to instruction. Reasons for failure, such as lack of or inadequate preparatory knowledge or insufficient opportunities to practice to mastery, became things a teacher could simply correct (Barbash, 2012; Engelmann, 2007).

With early colleagues Wes Becker, Carl Bereiter, and others, Engelmann created the explicit, carefully sequenced, scripted model of teaching known as Direct Instruction or "DI." DI emphasizes judiciously designed lessons built around a "big idea" that has been analyzed and organized into manageable chunks which are directly taught using prescribed tasks and involving high rates of student responding. Everything they designed was rigorously tested in the classroom, what worked was used to create a series of programs from which any child could learn, and any teacher could teach (Becker, 1992).

Regardless of subject matter, DI programs share recognizable features such as scripted lessons presented to small groups, clear instructional language, signals for student responding, frequent questioning and checks for mastery, specific feedback, and targeted error correction (Binder & Watkins, 1990; Stein et al., 1998; Watkins & Slocum, 2004). Figure 2 is sample script from the DI program Academic Core Level A. Unit 6. This early exercise is designed to teach students time and calendars. Note how the activity is broken into manageable steps and the clear instructions for the teacher and the learner, as well as the signaling, frequent opportunities for response, and repetition till firm (no student errors).

DI programs are thoroughly evaluated for effectiveness during and after design and development (see National Institute for Direct Instruction [NIFDI], n.d.). Using an iterative testing model (see Markle, 1967, Twyman et al., 2005), the logical and structural elements are tested with students to make sure they work. If students don't understand something, it flags a problem in the program. The problem is identified, the instruction altered, and the sequence is tested again. When a full draft of an instructional program is ready, it is field tested in schools representing diverse student demographics and geographical settings. Student performance data, teacher observations, and feedback are collected to revise the programs once again as needed. With some notable exceptions (e.g., the *Headspout* Reading Programs, see Layng et al., 2003, and Twyman et al., 2005) few other broadly available and widely used

Lesson 1

- a. Here's a fact about time: There are 12 months in a year. Say that fact. (Signal.) *There are 12 months in a year.*
 - Everybody, say the months of the year. (Signal.) *January, February, March, April, May, June, July, August, September, October, November, December.*
 - Everybody, how many months in a year? (Signal.) *12.*
 - Remember, 12 months in a year.
- b. There are 7 days in a week. Say that fact. (Signal.) *There are 7 days in a week.*
 - Everybody, start with Sunday and say the days of the week. (Signal.) *Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday.*
 - How many days are in a week? (Signal.) *7.*
 - How many months are in a year? (Signal.) *12.*
- c. Listen: There are 365 days in a year. Say that fact. (Signal.) *There are 365 days in a year.*
 - Here's a fact: There are 24 hours in a day. Say that fact. (Signal.) *There are 24 hours in a day.*

Fig. 2 Sample Direct Instruction Teaching Script for Time and Calendar Concepts. Source: Direct Instruction Academic Core Level A. Unit 6 Time Calendar. Lesson 1 (p.1). Courtesy of the National Institute for Direct Instruction.

instructional programs have been developed with this type of analysis and with a set of procedures so thoroughly tested.

Research does not end when a DI program is published. Hundreds of peer-reviewed studies have evaluated the efficacy DI programs over the last 50 years. These studies have addressed curriculum domains (e.g., reading, math, language), student age and ability (e.g., preschool to adulthood; typically developing students as well as those varying abilities), demographics (e.g., socio economic status, race and ethnicity, rural, suburban, urban locales, as well as students from the United States and in other countries). Study designs include randomized control trials, multi-site or other large-scale multi-site implementations, longitudinal studies, and single subject designs. Research has consistently shown strong evidence that the achievement level of students in DI programs is higher than for those students using other programs (see Mason & Otero, this issue; Stockard, this issue).

Evaluating DI

An early and important test of Engelmann's methodology and the resulting programs just happens to be the largest controlled educational experiment in history. Sponsored by the U.S. federal government, Project Follow Through was a longitudinal study designed to identify effective methods for teaching disadvantaged children in kindergarten through grade 3. From 1968 to 1976 over 200,000 children from over 170 different communities representing a range of demographic variables (community size,

location, ethnicity, and family income) were included in this comparison of 22 different models of instruction (Adams & Engelmann, 1996; Becker, 1978; Becker & Engelmann, 1996; Engelmann, 2007). In 1978, shortly after the conclusion of Follow Through, the U.S. Education Commissioner at the time of time wrote the following about the project:

...Since the beginning of Follow Through in 1968, the central emphasis has been on models. A principal purpose of the program has been to identify and develop alternative models or approaches to compensatory education and assess their relative effectiveness through a major evaluation study which compare the performance of Follow Through children with comparable children in non-Follow Through projects over a period of several years. That study had just been completed...The evaluation found that only one of the 22 models which were assessed in the evaluation consistently produced positive outcomes [The Becker-England Direct Instruction model]... (Engelmann, 2007, p. 248.)

To reiterate: Only one out of the 22 models consistently produced positive outcomes³. Official reports of Follow Through results (Stebbins et al., 1977) and the numerous subsequent analyses all converge on this fact: Regardless of prior achievement level and other demographics, students participating the Direct Instruction model attained the highest achievement scores across all three categories: basic skills, cognitive (higher-order thinking) skills, and affective responses (Becker, 1978; Watkins, 1995). The Direct Instruction model and, to a lesser extent the Behavior Analysis model, not only showed that tangible educational outcomes for disadvantaged children were possible, but also demonstrated the existence of instructional technologies that actually raised the academic achievement of children (cited in Watkins, 1995).

Engelmann (2007) summarized the results:

For the first time in the history of compensatory education, DI showed that long-range, stable, replicable, and highly positive results were possible with at-risk children of different types and in different settings...DI showed that relatively strong performance outcomes are achievable in all subject areas (not just reading) if the program is designed to effectively address the content issues of these areas. (pp. 229-230).

Here was the instructional antidote to the vast educational failures noted previously. Even beyond the outcomes of Project Follow Through, this remedy is widely accessible. Additionally, it is well-aligned with Hart and Risley's (1995) trailblazing work showing that the achievement gap between children in professional families and those in welfare families is primarily a language and knowledge gap; a gap that can be closed with good instruction in critical knowledge.

Engelmann and Carnine's *Theory of Instruction* (1982, 1991) provides exacting detail of the instructional principles underlying DI. are discussed in at least four other

³ In a review of Project Follow Through, Watkins (1995) provides an excellent examination of why the DI model was not more widely disseminated and embraced.

books by Engelmann (1969, 1980, 1997, 2007) and numerous publications (e.g., Barbash, 2012; Moore, 1986, Stein et al., 1998; Magliaro et al., 2005; Twyman, this issue; Twyman & Hockman, 2021). The principles have been validated in at least five meta-analyses (Adams & Engelmann, 1996; Borman et al., 2003, Hattie, 2009; Stockard et al., 2018) and across over 400 publications [see NIFDI's searchable database of DI research (<https://www.nifdi.org/component/jresearch/?view=publications&layout=tabular&Itemid=769>)].

Clearly Engelmann and colleagues were on to something important. They designed extremely effective instructional programs, to which many children and adults owe their literacy, math, and academic success. They demonstrated, in one of the most extensive educational experiment ever conducted, that DI produced significantly higher academic achievement, and greater self-esteem and self-confidence than any of the other programs (Stockard, this issue; Watkins, 1995). And they provided, to anyone and everyone interested, a framework that supported the teaching of any subject matter to any student. All one has to do is apply the principles.

Why is DI So Effective?

Engelmann and Carnine (1991) were steadfast in the notion that a theory of instruction must be based upon a scientific analysis. Using upon the premise that learning is an interaction between the learner and the environment, they focused on an overarching empirical question: How do we investigate the relationship between the environment and the learner? They recognized the impossibility of scientifically analyzing the environment-learner relationship without holding one factor constant and systematically varying the other. Because learners, with all their idiosyncrasies, could not be held constant they turned to a careful experimental control of the environment, which for them, was instruction (Engelmann et al., 1979). Thus, the focus on carefully specified and controlled instruction is fundamental to DI's design and instructional sequences.

Consider the effects of designing instruction in such a way that there would be no misunderstanding, no misinterpretation, and no confusion. This is the goal of DI programs, to reduce errors and bewilderment by providing one and only one logical interpretation of what was being taught. This unwavering focus on effective communication with the learner, which Engelmann and Carnine termed "faultless," epitomizes the core of logic of DI. It is the antidote to the ineffective principles of instruction embraced by the education establishment: ones that are rife with unfounded assumptions, knowledge gaps, generalities, inconsistencies, loose associations, and worse, that lay the failure to learn on the shoulders of the child (Heward, 2003). DI flips those faulty notions. Learners are not to blame for failures to learn, the failure lies in the instruction. Given the right instruction, all children will learn. This is why "faultless communication" the heart and soul of DI. It is one of the strongest tools we have in our fight against a failure to teach.

Faultless Communication

Ponder this potential headline: "Patient at Death's Door - Doctors Pull Him Through." Did you assume the patient was saved due to the efforts of his hardworking physicians? Or perhaps you had thoughts of the patient being brought "to the other side" by some

very callous doctors? If this were a real headline, we can use experience and expectation to assume the patient was saved by his doctors. However the ambiguity of the communication allows for multiple interpretations. The fact is, the English language (perhaps, all natural languages) is ripe for misinterpretation (Joseph & Liversedge, 2013; Tabossi & Zardon, 1993). This is unfortunate for learners and their teachers, as all too often instruction is riddled with confusion and ambiguity.

Imagine that one could design instruction in such a way that it would communicate one and only one interpretation to the learner. There would be no false inferences, no misunderstandings, no confusion. Concepts would be learned perfectly. The principles of DI, embodied in faultless communication, bring us closer to that goal (Engelmann et al., 1988). To address ambiguity and improve learning, Engelmann and Carnine contend that both the instructional content and how that content is communicated must be logically designed and research tested. To reduce misinterpretation and maximize learning, instruction must be designed and arranged to communicate a single logical interpretation (Adams & Carnine, 2003; Barbash, 2012). The effects on the learner's performance are then observed, and the instructional communication redesigned until faultless. This process of making a communication crystal clear and then observing its effects on performance of the learner is essential to the effectiveness of DI. It is used to design the instructional sequences presented by the teacher. *Ideally* the instruction would work for *all* learners. When it did, it was considered “faultless communication.” It is based the analytical, systematic, and precise arrangement of instructional stimuli such that they communicate without misunderstanding.

Noting a lack of systematic effort to develop precise principles of communications to use in instruction or to analyze knowledge systems, Engelmann and Carnine devoted *Theory of Instruction* (1991) to that task. Their logical premise: Communications so well designed as to be faultless will convey only one interpretation and thus would be capable of teaching any learner the intended concept or skill. The premise drives the construction and evaluations of all Direct Instruction programs.

Although the term faultless communication had a literal meaning for Engelmann and Carnine (1991), behavior analysts might see it as a metaphor for carefully arranged instructional sequences. Processes underlying the term include concept analysis (Keller & Schoenfeld, 2014; Skinner, 1957; Layng, 2018), stimulus control (Dietz & Malone, 1985; McIlvane & Dube, 1992), general case instruction (Horner et al., 2005) and multiple exemplar training (Holth, 2017), active student responding (Heward, 1994), reduced error and errorless learning (Touchette & Howard, 1984), immediate feedback (Mangiapanello & Hemmes, 2015; Tosti, 1978), implicit generalization (Stokes, 1992; Stokes & Baer, 1977), and explicit teaching (see Rosenshine, 1986). In fact, behavior analytic instruction and DI share so many design and delivery characteristics that teaching approaches featuring these characteristics are often referred to as “little di” (Rosenhine, 1986).

The features of faultless communication critical to DI also share many features with early programmed instruction (PI; Skinner, 1986; Vargas & Vargas, 1991). After an analysis of the subject matter, content is broken down into pieces of instruction that contain information and opportunities for learner response, called frames. The frames are carefully constructed to emphasize critical information and prime correct responding. Learners read the small bit of information within the frame, answer a question, and access feedback about the accuracy of their response before moving on to

the next frame. The instruction is cumulative; the frames lead the learner from simple to complex information in a carefully ordered sequence. Instruction occurs individually, is self-paced, and, because there are frequent opportunities to respond, active.

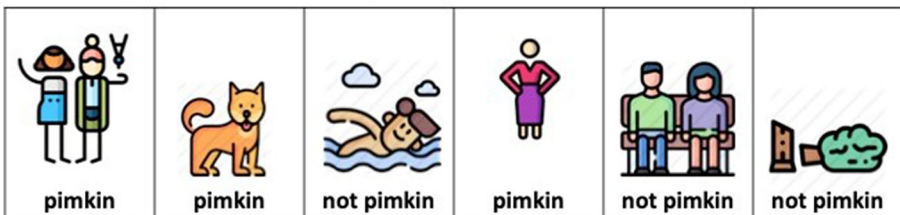
The autonomous nature of PI is one major difference between it and DI. PI frames are presented via a book, machine, or software, whereas teachers typically deliver DI programs (the computer-based phonics program *Funnix*; Engelmann & Owen, n.d.) program being an exception]. However one major similarity among DI, PI, and behavior analytic instructional is that the only measure of quality is the method’s success in meeting the needs of the learner.

Designing Faultless Communication

A modified activity from Parsons and Polson (n.d.) may illustrate the establishment of stimulus control using faultless communication. Figure 3 presents a range of stimuli used to communicate the concept of “pimkin.” Stimuli are presented that are instances of pimkin (i.e., positive examples) and that are not instances of pimkin (i.e., negative examples). For teaching, six examples are presented with text to indicate whether the example is positive (an instance of pimkin) or negative (not an example of pimkin). After viewing the teaching set, completing a test will determine if you have learned the concept. (Readers are encouraged to engage in the activity presented in Figure 3.)

Let’s review your responses for the testing set. The numbers 1, 2, 3, and 6 are each positive examples of “pimkin.” Numbers 4 and 5 are negative examples. If you got all six correct you’ve likely learned the concept of “pimkin” (or at least are off to a good start with only six teaching examples provided). If we translate the nonsense word “pimkin” into plain English, it might be “standing” or upright, or perhaps even vertical. Additional positive and negative examples would allow the concept to be narrowed to specific terms.

A. Teaching Set. Review each example of “pimkin” or “not pimkin.”



B. Testing Set. State whether each example is “pimkin” or “not pimkin.”

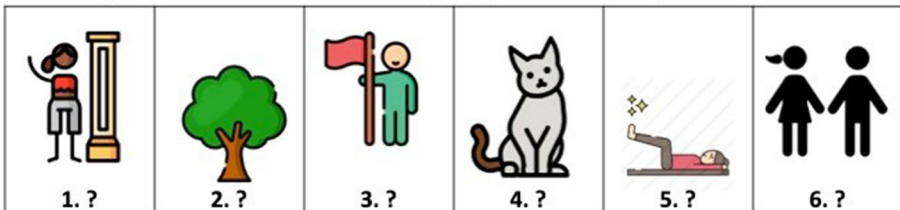


Fig. 3 Activity: Teaching the Concept of “Pimkin”

Was the communication successful (i.e., faultless)? The instructional stimuli included examples and nonexamples of “pimkin” and isolated the feature that made stimuli “pimkin” or not. Further stimuli tested generalization of the instruction to other stimuli. If faultless, logically the instruction had to succeed.

To avoid the formation of “misrules” that result in a cascade of potential errors in learning, one must be specific in how that communication is crafted. Table 1 lists five structural requirements for designing faultless communication as detailed by Engelmann and Carnine (1991, pp. 5–6). The activity in Figure 3 was designed following the rules of faultless communication.

Let’s see how the activity in Figure 3 fulfills these requirements. For clarity, “standing” (the intended meaning of the concept taught) is used instead of the nonsense word “pimkin.” The first requirement sets the stage: Present a set of positive examples that are the same with respect to one and only one distinguishing quality. All the positive examples on the top row share that same one quality, someone or something standing. Some positive examples may share other features (i.e., people) but that is ruled out, as we shall see later.

To make sense of all the stimuli we need the second requirement: Provide two signals—one for every example that possesses the quality that is to be generalized, another to signal every example that does not have this quality. The texts “pimkin” and “not pimkin” (standing or not standing) below each image in the top row serve as the two signals in our activity. They unambiguously provide the basis for classifying each example as either positive or negative. Regardless of signal modality (e.g., print, vocal-verbal, sign, or extra-stimulus enhancement such as shading or a circle), it must clearly convey to the learner that the stimuli is, or is not, an example of the quality.

The third requirement is: Demonstrate a range of variation for the positive examples (to induce a rule that is appropriate for classifying new examples on the basis of sameness), builds upon the first requirement. In our activity, each positive instance of

Table 1 Structural Conditions of Faultless Communication

The communication must:	Function
1. present a set of positive examples that are the same with respect to one and only one distinguishing quality	To identify the quality that serves as the basis for generalization
2. provide two signals: one for every example that possesses the quality that is to be generalized, another for every example that does not have this quality	To indicate to the learner all examples possess the quality, whereas all negative examples do not have the quality
3. demonstrate a range of variation for the positive examples	To induce a rule appropriate for classifying new examples on the basis of sameness
4. present a range of negative examples	To show the limits of permissible variation
5. present novel positive examples and negative examples that fall within the range of quality variation demonstrated earlier	To test for generalization

Note. Content is based on Engelmann and Carnine’s (1991) *Theory of Instruction* (pp. 5–6). In keeping with how terms are used within that book: The term “quality” in the Table refers to any irreducible feature of the example. *Positive examples* are stimuli that contain that feature. *Negative examples* are stimuli that do not possess that feature.

“standing” is the same with respect to someone or something standing, but different in many other ways (i.e., number or types of people, animals, plants, additional details, colors, image size, or location within the presentation box). If every positive example were a stick figure standing, a learner might infer “line drawings” is the concept being learned, and thus not generalize “standing” to a photograph of woman standing. Likewise, if every positive example were a dog or cat standing, a learner might infer the concept was animals or pets, and not extend the class to humans or objects.

Presenting a wide range of carefully selected positive examples is essential. To further indicate how the range of positive examples can be extended, in our activity images of two people and of one person standing are both presented as positive examples, indicating that the number is not relevant to the concept being learned. The use of people, animals, and plants as positive examples demonstrates that “standing” is not dependent upon what is standing.

In the fourth requirement, Show the limits of permissible variation by presenting negative examples, the negative examples should be precise in demonstrating the boundaries of a permissible generalization. A range of negative examples must include ones that are only slightly different from positive examples, and then we must be clear to the learner that these are in fact negative examples. In our activity, the negative example of two people sitting shares many qualities with the positive example of two people standing. The signal (i.e., the text “not pimpin”) made it clear that although both images contained people, the image with people standing was a positive example and the image with people sitting (not standing) was not a positive example. Contrasting these highly similar positive and negative examples highlighted their difference: standing, which is exactly the quality being taught. Negative instances that share several (but not all) critical attributes of a concept are often called “close-in” nonexamples (REF). Close-in negative examples are essential in defining the boundaries of a concept.

To fulfill the fifth requirement one must provide a test of generalization that involves new examples (and nonexamples) that fall within the range of quality (attribute) variation demonstrated by the examples. A test of generalization indicates how accurately the learner responds to new positive and negative examples, and if successful, adds confirmation that the quality or concept has been taught. The lower row of our activity in Figure 3 was designed to test learning through a test for generalization. Because the example set in our activity was relatively small, the test stimuli were carefully selected to catch possible misinterpretations. For example, the text sample of a cat sitting (a negative example) was selected to contrast with the positive example of a dog standing. This begins to verify that “animal-ness” does not exert undue control over “standing.” Learners who might have responded “pimkin” (standing) to the cat sitting would be required additional instruction using animals (or whatever the potential source of faulty stimulus control may be) as negative examples (e.g., sitting, walking, sleeping, eating) as well as a range of both animals and other stimuli as positive examples. In addition, in our activity the positive example of a tree standing upright was selected to contrast with the negative teaching example of a tree cut down, to verify that “tree-ness” did not overgeneralize as a negative instance, further exemplifying the use of test examples that are novel, yet primed by the teaching examples.

The stimuli used in Figure 3 warrants some clarification. The number and selection was constrained by the medium in which they would be used, for instance only six

stimuli were used for teaching and another six for testing. Images needed to be selected for clarity at a small size and may be viewed in greyscale. In a live teaching session (in-person or online), one should include greater variation in positive and negative examples, including the range of representation, size, detail background variation, 3-d representation or movement, as well as greater liberties in layout, sequence, or type of image.

Furthering Generalization

Confirmation of what has (or has not) been learned comes from tests of generalization, as noted in the fifth requirement. This typically involves extending the response to novel stimuli. For example, extension is observed when a child, who has learned that basketballs, baseballs, and beach balls, are balls, but Frisbees, skateboards, and jump ropes are not, says “ball” in the presence of a never before seen soccer ball (based on Fraley, 2004). Although the soccer ball has some properties not previously encountered in other balls, it shares the defining qualities of all balls. Calling this novel item a ball was occasioned by the defining qualities of balls, not the novel features of this particular soccer ball. This is called generic extension (Skinner, 1957) and is likely the most ubiquitous form of extension. However, there are others. Layng et al. (2011) suggest that we may test for extension using specific changes in referent behaviors. Their examples are derived from the concept of “cat”:

We can test for generic extension by providing multiple examples that include a range of different cats (Skinner, 1957, p. 91). We can test for abstract tacts by including foxes of similar size and color to see if they are called “cat” (after Skinner, 1957, p. 107). We can show that the learner “understands” the concept of “cat” by providing a juxtaposed sequence of examples and non-examples of cat such that the learner points to all the cats and none of the other similar creatures, such as lemurs or skunks. . . . We can test for metonymical extension by asking for all things that go with Halloween and seeing whether a black cat is selected. . . . We can test for such abstract tact “inheritance” by saying, “choose the mammal with big eyes relative to its head size and whiskers and upright ears,” and seeing if the learner selects a cat from among other mammals. (n.p.)

Relevance for Behavior Analysts

To improve student learning, Engelmann and Carnine (1991) considered three separate analyses—the analysis of behavior, the analysis of stimuli used as teaching communications, and the analysis of knowledge systems. The analysis of behavior provides empirically based principles regarding the ways in which the environment influences behavior for different learners. The analysis of communications supports the logical design of effective transmission of knowledge based on ways in which examples are the same and how they differ, setting the occasion for the logical range of generalizations that should occur upon learning a specific set of examples. The analysis of knowledge systems assists with the logical organization of content so that relatively

efficient communications are possible. Figure 4 depicts the interplay between a logical analysis of information and an analysis of the behavior of the learner.

Even when instruction is faultless, it may not always achieve the designed results. However, the combination of faultless communication and an analysis of the behavior of the learner can reveal important information. When we observe the effects of a faultless communication on a learner, because the communication was same for all learners we can rule out instructional variables and look at other factors for any differences in learning. Each learner's response to the instruction provides precise information about that learner (Parsons & Polson, n.d.). If “faultless” instruction fails, then an analysis of the learner’s prerequisite repertoire and current responding and the conditions of learning environment (the contingencies) should indicate what to do next. Supplemental instruction or environmental arrangements may be necessary to increase the learner’s ability to respond appropriately to the logically faultless presentation (Brown & Tarver, 1985). An applied behavior analyst can assess the learner’s behavior in relation to the instruction, and as necessary, teach any missing prerequisite skills, increase schedules of reinforcement, or incorporate other environmental contingencies to support learning.

There are many reasons for applied behavior analysts to be interested in DI programs and the underlying theory of instruction. Access to efficient, effective instruction for more learners may seem the most relevant. In addition, because applied behavior analysts, experimental behavior analysts, and all behavior analysts who teach or train others likely spend a good deal of time designing instruction, faultless communication and the principles described in *Theory of Instruction* could be very useful in designing one’s own instructional sequences and programs. When coupled with the *Structural Conditions of Faultless Communication* presented in Table 1, a behavior analyst’s

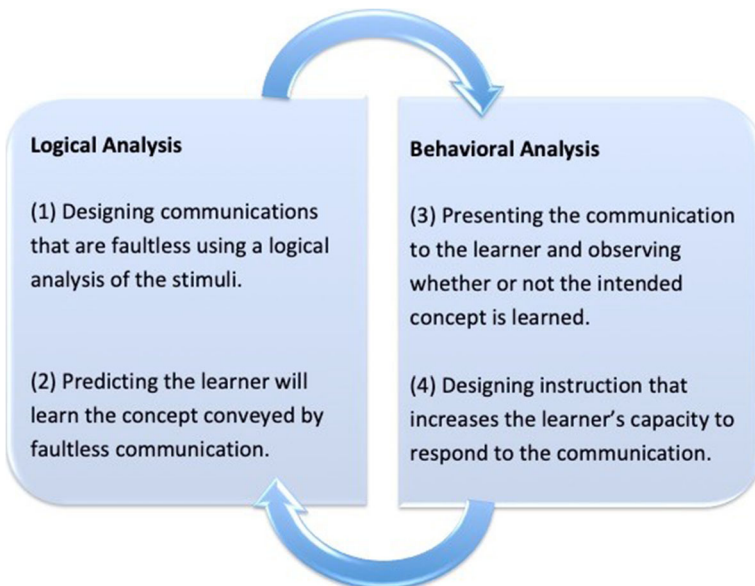


Fig. 4 Interplay Between a Logical Analysis of Information and a Behavioral Analysis of the Learner

instructional design repertoire is vastly increased. The special section on Direct Instruction in *Behavior Analysis in Practice* also may be of interest to those looking for applied examples of content analysis, content teaching and other how-to's of effective instruction.

Another reason DI's theory of instruction, in particular faultless communication, is relevant to behavior analysis has to do with the synergies between subject matter and analysis. Evidence from the fields of psychology, education, and instructional design all converge on the importance of incorporating examples to facilitate concept learning and skill acquisition (e.g., Clark, 1971; Markle & Tiemann, 1970; Merrill et al., 1992). Engelmann and colleagues conducted their concept analyses from a logical perspective; behavior analysis looks at how the organism responds to the stimuli. Various stimuli "belong" to the same group because they all evoke a common response to a common property (Keller & Schoenfeld, 2014). The field of behavior analysis has its own rich history in designing instruction (see Johnson & Chase, 1981; Markle, 1975, 1983/1990; Skinner, 1954) and although terminology may differ, many of the approaches to analyzing instructional content are the same. As described by Johnson (2014), behavior analytic or behavior-based instructional design is:

. . . a systematic approach to identifying the critical variables capable of manipulation to produce efficient learning and the process of continual refinement of these instructional variables to improve environmental contingencies. Put differently, it is how designers can best establish different types of stimulus control for different types of performance outcomes. In contrast to other forms of instructional design, it is not rooted in learner traits or cognitions beyond the designer's influence and it does not postulate unnecessary hypothetical processes or inferred structures as the primary explanatory models. (p. 63)

Both basic and applied behavior analysts may be interested in DI's theory of instruction and faultless communications for reasons beyond the (real) need to arrange environmental conditions to obtain better learning. Basic behavior analytic research has contributed greatly to the understanding of the relevancy of antecedent stimuli and controlling variables (Touchette & Howard, 1984). The experimental analysis of behavior has provided some "rules" for how to select, arrange, and present stimuli (see Stromer et al., 1992). The work of Terrace (1963) studying the differential effects of stimuli complexity and the transfer of stimulus control also has tremendous relevance to the design of faultless communication. Schilmoeller et al.'s (1979) work distinguishing stimulus shaping and fading, as well as earlier work by Reid et al. (1969) on stimulus priming (as opposed to prompting) could offer insight on strengthening faultless communications. Moreover, faultless communications may be an effective vehicle for the further investigations of stimulus shaping and fading (Schilmoeller et al., 1979), stimulus priming (Reid et al., 1969), or stimulus equivalence methods to increase instructional efficiencies (Sidman, 1984). Fifty years ago, Ray and Sidman (1970, p. 199) wrote:

All stimuli are complex in that they have more than one element, or aspect, to which a subject might attend [and thus] . . . we may never have a generalizable formula for forcing subjects to discriminate a specific stimulus aspect. We may

have to settle, instead, for a combination of techniques, each of which is known to encourage stimulus control.

Perhaps that “generalizable formula” is a three-legged stool made sturdy by an analysis of faultless communication, the behavior analytic design of instruction, and the experimental analysis of dimensions related to stimulus control.

Relevance for Learners

Both DI and applied behavior analysis (ABA) are grounded in the foundational premise that when children are not learning, the fault lies not with them, but with the instruction (ABAI, 1990; Engelmann & Carnine, 1982). Each has its own variation of the credo “the rat is always right” (attributed to Skinner; see Davidson, 1999; Lindsley, 1990), restated by Fred S. Keller as “the student is always right” (see Sidman, 2006), and Engelmann’s “whatever the kid does is the truth” (see Heward et al., 2021).

Similar to the ABAI’s statement on “Student Right to Effective Education,” NIFDI (n.d.) identifies five key philosophical principles:

- All children can be taught.
- All children can improve academically and in terms of self-image.
- All teachers can succeed if provided with adequate training and materials.
- Low performers and disadvantaged learners must be taught at a faster rate than typically occurs if they are to catch up to their higher-performing peers.
- All details of instruction must be controlled to minimize the chance of students’ misinterpreting the information being taught and to maximize the reinforcing effect of instruction.

Walker (2011) sees “[the role of] teaching is to present information efficiently and clearly so students do not make needless errors and draw incorrect inferences regarding key discriminations” (n.p.). When based on a science of instruction, this process leads to critically important student outcomes. Thus, to be effective, education must embrace two critical actions: (1) the scientific design and logical analysis of instruction that allows students to learn with minimal errors and (2) the careful monitoring and evaluation of student responses to instruction (Walker, 2011). Bushell and Baer (1994) also encouraged the careful monitoring in education, stating “close continual contract with the relevant outcome data is a fundamental, distinguishing feature of applied behavior analysis . . . [it] ought to be a fundamental feature of classroom teaching as well” (p. 7). With *Theory of Instruction* and its emphasis on faultless communication, Engelmann and Carnine (1982, 1991) provide a logical scientific method to design instruction, whereas behavior analysis gives us a mechanism to improve instruction and precisely monitor and analyze its impact. Together they bring us closer to a much-needed science of education as well as a science of instruction (Kauffman, 2011). Engelmann’s (2007) words illustrate why something as logical as faultless communication is the heart and soul of Direct Instruction:

We’re not going to fail you. We’re not going to discriminate against you, or give up on you, regardless of how unready you may be according to traditional

standards. We are not going to label you with a handle, such as dyslexic or brain-damaged, and feel that we have now exonerated ourselves from the responsibility of teaching you. We're not going to punish you by requiring you to do things you can't do. We're not going to talk about your difficulties to learn. Rather, we will take you where you are, and we'll teach you. And the extent to which you fail is our failure, not yours. We will not cop out by saying, "He can't learn." Rather, we will say, "I failed to teach him. So I better take a good look at what I did and try to figure out a better way." (n.p.)

Conclusion

Faultless communication seems more imperative today than ever. Pandemic-driven school closures worldwide have led to losses in learning that will not be quickly or easily be made up. It will be hard for schools to catch up to their prior performance levels, let alone move ahead. These losses will have lasting economic impacts both on the affected students and on each nation unless they are effectively remediated (Hanushek & Woessmann, 2020). No single intervention will eliminate the education inequities between low-income countries and high-income countries, close the achievement gap between low-performing and high-performing students, guarantee that every student, in every nation worldwide graduates secondary school fully ready and able to be successful in life, or ensure that every third grader, irrespective of circumstances, socioeconomic status, race, or ethnicity, will be more than proficient at reading, math, and other fundamental academic skills. However, coupled with all that we know about behavior analysis in education, one model could go a long way towards getting us there: Direct Instruction. Perhaps this article and others in the special sections of *Perspectives on Behavior Science* and *Behavior Analysis in Practice* will make it more likely that more readers will study, implement, research, and disseminate the theories that underlie DI and the programs it produces.

References

- Adams, G., & Carnine, D. (2003). Direct instruction. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 403–416). Guilford.
- Adams, G., & Engelmann, S. (1996). Project Follow Through: In-depth and beyond. *Effective School Practices*, 15(1), 43–56.
- Apple, M. W. (2007). Ideological success, educational failure? On the politics of No Child Left Behind. *Journal of Teacher Education*, 58(2), 108–116.
- Association for Behavior Analysis International (ABAI). (1990). *Statement on students' right to effective education*. <https://www.abainternational.org/about-us/policies-and-positions/students-rights-to-effective-education,-1990.aspx>.
- Barbash, S. (2012). *Clear teaching: With Direct Instruction, Siegfried Engelmann discovered a better way of teaching*. Education Consumers Foundation.
- Becker, W. C. (1978). National evaluation of Follow Through: Behavior-theory-based programs come out on top. *Education & Urban Society*, 10, 431–458.

- Becker, W. C. (1992). Direct Instruction: A twenty-year review. In R. P. West & L. A. Hamerlynck (Eds.), *Designs for excellence in education: The legacy of B. F. Skinner*. Sopris West.
- Becker, W., & Engelmann, S. (1996). Sponsor findings from Project Follow Through. *Effective School Practices*, 15(1), 33–42.
- 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.
- Brown, V. L., & Tarver, S. G. (1985). Book reviews: Two perspectives on Engelmann and Carnine's Theory of Instruction. *Remedial & Special Education*, 6(2), 56–60. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.872.7992&rep=rep1&type=pdf>.
- Bushell Jr., D., & Baer, D. M. (1994). Measurably superior instruction means close, continual contact with the relevant outcome data: Revolutionary! 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. 3–10) Brooks/Cole.
- Camera, L. (2019). U.S. students show no improvement in math, reading, science on international exam. *U.S. News & World Report*. <https://www.usnews.com/news/education-news/articles/2019-12-03/us-students-show-no-improvement-in-math-reading-science-on-international-exam>
- Clark, D. C. (1971). Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. *Journal of Educational Psychology*, 62(3), 253–278. <https://doi.org/10.1037/h0031149>.
- Damon, A., Glewwe, P., Wisniewski, S., & Sun, B. (2016). Education in developing countries: What policies and programmes affect learning and time in school? OECD.org. Retrieved July 15, 2021 from <https://www.oecd.org/derec/sweden/Rapport-Education-developing-countries.pdf>
- Davidson, J. (Director). (1999). *B.F. Skinner: A fresh appraisal* [Video]. Davidson Films
- De Brey, C., Snyder, T. D., Zhang, A., & Dillow, S. A. (2021). *Digest of Education Statistics 2019* (NCES 2021-009). National Center for Education Statistics.
- Detrich, R., Keyworth, R., & States, J. (2007). A roadmap to evidence-based education: Building an evidence-based culture. *Journal of Evidence-Based Practices for Schools*, 8(1), 26.
- Dietz, S. M., & Malone, L. W. (1985). Stimulus control terminology. *The Behavior Analyst*, 8(2), 259–264. <https://doi.org/10.1007/bf03393157>.
- Driver, R. (1978). When is a stage not a stage? A critique of Piaget's theory of cognitive development and its application to science education. *Educational Research*, 21(1), 54–61.
- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child & Family Psychology Review*, 11(3), 75–113. <https://doi.org/10.1007/s10567-008-0036-x>.
- Engelmann, S. (1969). *Conceptual learning*. Dimensions Publishing.
- Engelmann, S. (1980). *Direct instruction*. Educational Technology.
- Engelmann, S. (1997). *Preventing failure in the primary grades*. Association for Direct Instruction.
- Engelmann, S. (2007). *Teaching needy kids in our backward system: 42 years of trying*. ADI Press.
- Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. Irvington Publishers.
- Engelmann, S., & Carnine, D. (1991). *Theory of instruction: Principles and applications* (Rev. ed.). ADI Press.
- Engelmann, S., & Owen, D. (n.d.) Funnix [computer software]. <https://www.funnix.com>
- Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The Direct Instruction follow through model: Design and outcomes. *Education & Treatment of Children*, 11(4), 303–317.
- Engelmann, S., Granzin, A., & Severson, H. (1979). Diagnosing instruction. *Journal of Special Education*, 13(4), 355–363.
- Fraleigh, L. E. (2004). On verbal behavior: The second of four parts. *Behaviorology Today*, 7(2), 9–36.
- Furlong, J., & Phillips, R. (Eds.). (2001). *Education, reform and the state: Twenty-five years of politics, policy and practice*. Routledge. <https://doi.org/10.4324/9780203469552>
- Freire, P. (1985). *The politics of education: Culture, power, and liberation*. Greenwood Publishing Group.
- Glenn, S. S. (1988). Contingencies and metacontingencies: Toward a synthesis of behavior analysis and cultural materialism. *The Behavior Analyst*, 11(2), 161–179. <https://doi.org/10.1007/BF03392470>.
- Goldiamond, I. (1964). A research and demonstration procedure in stimulus control, abstraction, and environmental programming. *Journal of the Experimental Analysis of Behavior*, 7, 216.

- Hanushek, E. A., & Woessmann, L. (2020, September). The economic impacts of learning losses. OECD Education Working Papers, No. 225. *OECD Publishing*. <https://www.oecd.org/education/The-economic-impacts-of-coronavirus-covid-19-learning-losses.pdf>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hernandez, D. J. (2012). Double jeopardy: How third-grade reading skills and poverty influence high school graduation. The Anne E. Casey Foundation. <https://gradelevelreading.net/wp-content/uploads/2012/01/Double-Jeopardy-Report-030812-for-web1.pdf>
- Heward, W. L. (1994). Three "low-tech" strategies for increasing the frequency of active student response during group instruction. 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. 283–320) Brooks/Cole.
- Heward, W. L. (2003). Ten faulty notions about teaching and learning that hinder the effectiveness of special education. *Journal of Special Education*, 36(4), 186–205.
- Heward, W. L., Kimball, J. W., Heckaman, K. A., & Dunne, J. D. (2021). In his own words: Siegfried Engelmann talks about what's wrong with education and how to fix it. *Behavior Analysis in Practice*.
- Hirsch, E. D., Jr. (2016a). Don't blame the teachers. *The Atlantic*. <https://www.theatlantic.com/education/archive/2016/09/dont-blame-the-teachers/500552/>
- Hirsch, E. D., Jr. (2016b). *Why knowledge matters: Rescuing our children from failed educational theories*. Harvard Education Press.
- Holth, P. (2017). Multiple exemplar training: Some strengths and limitations. *The Behavior Analyst*, 40(1), 225–241.
- Horner, R., Sprague, J., & Wilcox, B. (2005). General case programming. *Encyclopedia of behavior modification and cognitive behavior therapy*, 1, 1343–1348.
- Johnson, K. R., & Chase, P. N. (1981). Behavior analysis in instructional design: A functional typology of verbal tasks. *The Behavior Analyst*, 4(2), 103–121.
- Johnson, D. A. (2014). The need for an integration of technology, behavior-based Instructional design, and contingency management: An opportunity for behavior analysis. *Revista Mexicana de Análisis de La Conducta*, 40(2), 58–72. <https://doi.org/10.5514/rmac.v40.i2.63665>
- Joseph, H. S., & Liversedge, S. P. (2013). Children's and adults' on-line processing of syntactically ambiguous sentences during reading. *PLoS ONE*, 8(1), e54141. <https://doi.org/10.1371/journal.pone.0054141>.
- Kantor, H. (1991). Education, social reform, and the state: ESEA and federal education policy in the 1960s. *American Journal of Education*, 100(1), 47–83.
- Kauffman, J. M. (2011). *Toward a science of education: The battle between rogue and real science*. Full Court Press.
- Keller, F. S., & Schoenfeld, W. N. (2014). *Principles of psychology: A systematic text in the science of behavior* (Vol. 2). B. F. Skinner Foundation.
- Klausmeier, H. J., Ghatala, E. S., & Frayer, D. A. (1974). *Conceptual learning and development: A cognitive view*. Academic Press.
- Kliebard, H. M. (1988). Success and failure in educational reform: Are there historical "lessons?". *Peabody Journal of Education*, 65(2), 143–157.
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Harvard University Press. <http://www.hup.harvard.edu/catalog.php?isbn=9780674035218&content=reviews>
- Kozol, J. (1988). *Savage inequalities*. Crown.
- Larson, E. J. (2010). *The creation-evolution debate: Historical perspectives* (Vol. 3). University of Georgia Press.
- Layng, T. V. J. (2018). Tutorial: Understanding Concepts: Implications for Behavior Analysts and Educators. *Perspectives on behavior science*, 42(2), 345–363.
- Layng, T. V. J. (2016). Thirty million words—and even more functional relations: A review of Suskind's *Thirty Million Words*. *The Behavior Analyst*, 39, 339–350.
- Layng, T. V. J., Sota, M., & Leon, M. (2011). Thinking through text comprehension I: Foundation and guiding relations. *The Behavior Analyst Today*, 12(1), 3–11. <https://doi.org/10.1037/h0100706>.
- Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2003). *Headsprout Early Reading™*: Reliably teaching children to read. *Behavioral Technology Today*, 3, 7–20.
- Leana, C. R. (2011, Fall). The missing link in school reform. *Stanford Social Innovation Review*. Retrieved July 15, 2021 from https://ssir.org/articles/entry/the_missing_link_in_school_reform

- Lee, J. W., & Lee, H. (2016). Human capital in the long run. *Journal of Development Economics*, 147-169.
- Lee Dr, S. E. (2013). Education as a Human Right in the 21st Century. *Democracy and Education*, 21(1), 1-9. Retrieved July 15, 2021 from <https://democracyeducationjournal.org/cgi/viewcontent.cgi?article=1074&context=home>
- Leon, M., Layng, T. V. J., & Sota, M. (2011). Thinking through text comprehension III: The programing of verbal and investigative repertoires. *The Behavior Analyst Today*, 12(1), 22–33. <https://doi.org/10.1037/h0100708>.
- Levin, J. R., & Allen, V. L. (1976). *Cognitive learning in children: Theories and strategies*. Academic Press.
- Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children*, 22(3), 10–15.
- Magliaro, S. G., Lockee, B. B., & Burton, J. K. (2005). Direct instruction revisited: A key model for instructional technology. *Educational Technology Research & Development*, 53(4), 41–55.
- Mangiapanello, K. A., & Hemmes, N. S. (2015). An analysis of feedback from a behavior analytic perspective. *The Behavior Analyst*, 38(1), 51–75. <https://doi.org/10.1007/s40614-014-0026-x>.
- Marchand-Martella, N. E., Kinder, D., & Kubina, R. (2005). Special education and direct instruction: An effective combination. *Journal of Direct Instruction*, 5(1), 1–36.
- Markle, S. M. (1967). Empirical testing of programs. In P. C. Lange (Ed.), *Programmed instruction: Sixty-sixth yearbook of the National Society for the Study of Education: 2* (pp. 104–138). University of Chicago Press.
- Markle, S. M. (1975). They teach concepts, don't they? *Educational Researcher*, 4(6), 3–9.
- Markle, S. M. (1983/1990). *Designs for instructional designers*. Morningside Press.
- Markle, S. M., & Tiemann, P. W. (1970). Behavioral analysis of cognitive content. *Educational Technology*, 10(1), 41–45.
- McIlvane, W. J., & Dube, W. V. (1992). Stimulus control shaping and stimulus control topographies. *The Behavior Analyst*, 15(1), 89-94.
- Merrill, M. D., Tennyson, R. D., & Posey, L. O. (1992). *Teaching concepts: An instructional design guide*. (2nd ed.). Educational Technology Publications.
- Moats, L. C. (2001). When older students can't read. *Educational Leadership*, 58(6), 36-41.
- National Assessment Governing Board. (n.d.). What are NAEP achievement levels and how are they determined? <https://www.nagb.gov/content/dam/nagb/en/documents/naep/achievement-level-one-pager-4.6.pdf>
- Moore, J. (1986). Direct instruction: A model of instructional design. *Educational Psychology*, 6(3), 201-229. <https://doi.org/10.1080/0144341860060301>
- National Assessment of Educational Progress (NAEP). (2019). Nations report card (2019). <https://nces.ed.gov/nationsreportcard/>
- National Center for Education Statistics. (2019). *Digest of education statistics, 2018* (NCES 2020-009), Chapter 1: All levels of education. Institute for Education Sciences. https://nces.ed.gov/programs/digest/d18/ch_1.asp
- National Institute for Direct Instruction. (n.d.). *Basic philosophy of Direct Instruction (DI)*. <https://www.nifdi.org/what-is-di/basic-philosophy>
- Nieto, S., & Bode, P. (2009). School reform and student learning: A multicultural perspective. In J. A. Banks & C. A. McGee Banks (Eds.), *Multicultural education: Issues and perspectives* (pp. 395-410). Wiley.
- Organisation for Economic Co-operation & Development (OECD). (2019). *PISA 2018 Results (Vol. I): What students know and can do*. <https://doi.org/10.1787/5f07c754-en>.
- Parsons, J., & Polson, D. (n.d.). *Siegfried Engelmann and Direct Instruction*. Psychology Learning Resources. <https://psych.athabascau.ca/open/engelmann/index.php>
- Payne, C. M. (2008). *So much reform, so little change: The persistence of failure in urban schools*. Harvard Education Press.
- Ray, B. A., & Sidman, M. (1970). Reinforcement schedules and stimulus control. In W. N. Schoenfeld (Ed.), *The theory of reinforcement schedules* (pp. 187–214). Appleton-Century-Crofts.
- Reardon, S. F., Robinson-Cimpian, J. P., & Weathers, E. S. (2014). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In *Handbook of research in education finance and policy* (pp. 507-525). Routledge.
- Reid, R. L., French, A., & Pollard, J. S. (1969). Priming the pecking response in pigeons. *Psychonomic Science*, 14(5), 227–227.
- Rosenhine, B. (1986). Synthesis of research on explicit teaching. *Educational leadership*, 43(7), 60–69.
- Roser, M., & Ortiz-Ospina, E. (2013). Primary and secondary education. *OurWorldInData.org*. <https://ourworldindata.org/primary-and-secondary-education>

- Roser, M., Nagdy, M., & Ortiz-Ospina, E. (2013). Quality of education. *Our World in Data*. <https://ourworldindata.org/quality-of-education>
- Rossi, P. H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. Sage.
- Schieffer, C., Marchand-Martella, N. E., Martella, R. C., Simonsen, F. L., & Waldron-Soler, K. M. (2002). An analysis of the reading mastery program: Effective components and research review. *Journal of Direct Instruction*, 2(2), 87–119.
- Schilmoeller, G. L., Schilmoeller, K. J., Etzel, B. C., & Leblanc, J. M. (1979). Conditional discrimination after errorless and trial-and-error training. *Journal of the Experimental Analysis of Behavior*, 31(3), 405–420.
- Schleicher, A. (2019). PISA 2018: Insights and interpretations. OECD. <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- Sidman, M. (1979). Remarks. *Behaviorism*, 7(2), 123–126.
- Sidman M. (1984). *Equivalence relations and behavior: A research story*. Authors Cooperative.
- Sidman, M. (2006). Fred S. Keller, a generalized conditioned reinforcer. *The Behavior Analyst*, 29, 235–242.
- Siegler, R. S., DeLoache, J. S., & Eisenberg, N. (2003). *How children develop*. Worth.
- Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24, 86–97.
- Skinner, B. F. (1957). *Verbal behavior*. Prentice Hall.
- Skinner, B. F. (1984). The shame of American education. *American Psychologist*, 39(9), 947–954. <https://doi.org/10.1037/0003-066X.39.9.947>.
- Skinner, B. F. (1986). Programmed instruction revisited. *Phi Delta Kappan*, 68(2), 103–110.
- Slavin, R. E. (2008). Evidence-based reform in education: What will it take? *European Educational Research Journal*, 7(1), 124–128.
- Slocum, T. A., & Rolf, K. R. (2021). Features of direct instruction: Content analysis. *Behavior Analysis in Practice*.
- Stein, M., Camine, D., & Dixon, R. (1998). Direct instruction: Integrating curriculum design and effective teaching practice. *Intervention in School & Clinic*, 33(4), 227–233.
- 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.
- 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), an evaluation of Project Follow Through. Abt Associates.
- Stokes, T. (1992). Discrimination and generalization. *Journal of Applied Behavior Analysis*, 25(2), 429–432. <https://doi.org/10.1901/jaba.1992.25-429>.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization 1. *Journal of Applied Behavior Analysis*, 10(2), 349–367. <https://doi.org/10.1901/jaba.1977.10-349>.
- Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education*, 2(3), 225–256.
- Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory & Language*, 32(3), 359–372.
- Terrace, H. S. (1963). Errorless transfer of a discrimination across two continua. *Journal of the Experimental Analysis of Behavior*, 6(2), 223–232.
- Tosti, D. T. (1978). Formative feedback. *NSPI Journal*, 13, 19–21. <https://doi.org/10.1002/pfi.4180170816>.
- Touchette, P. E., & Howard, J. S. (1984). Errorless learning: Reinforcement contingencies and stimulus control transfer in delayed prompting. *Journal of Applied Behavior Analysis*, 17(2), 175–188. <https://doi.org/10.1901/jaba.1984.17-175>.
- Twyman, J. S. (2021). The problem. In S. Redding (Ed.), *Opportunity and performance: Equity for children from poverty (from state policy to classroom practice)* (pp. 105–108). Information Publishing.
- Twyman, J. S., & Hockman, A. (2021). You have the big idea, concept, and some examples. . . . Now what? *Behavior Analysis in Practice*.
- Twyman, J. S., Layng, T. V. J., Stikeleather, G., & Hobbins, K. A. (2005). A non-linear approach to curriculum design: The role of behavior analysis in building an effective reading program. In W. L. Heward, T. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Y. Cartledge, R. Gardner, L. D. Peterson, S. B. Hersh, & J. C. Dardig (Eds.), *Focus on behavior analysis in education (Vol. 3; pp. 55–68)* Merrill/Prentice Hall.
- UNESCO. (n.d.). Learning: Achieving quality for all.
- UNESCO. (2017). More than one-half of children and adolescents are not learning worldwide Fact Sheet No. 46). <http://uis.unesco.org/sites/default/files/documents/fs46-more-than-half-children-not-learning-en-2017.pdf>

- United Nations. (1948). *Universal declaration of human rights*. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
- Vargas, E. A., & Vargas, J. S. (1991). Programmed instruction: What it is and how to do it. *Journal of Behavioral Education, 1*(2), 235–251. <https://doi.org/10.1007/bf00957006>.
- Walker, H. M. (2011, August 18). Making scientific sense of schooling: A path to improved student outcomes [Review of the book *Could John Stuart Mill have saved our schools?*, by S. Engelmann]. Amazon. <https://www.amazon.com/Could-John-Stuart-Saved-Schools-ebook/product-reviews/B006YY1WCQ>
- Watkins, C. (1995). Follow Through: Why didn't we? *Effective School Practices, 15*(1), 57–66 <https://pages.uoregon.edu/adiep/ft/watkins.htm>.
- Watkins, C., & Slocum, T. (2004). The components of direct instruction. In N. E. Marchand-Martella, T. A. Slocum, & R. C. Martella (Eds.). *Introduction to direct instruction* (pp. 28–65). Boston: Allyn & Bacon.
- Willingham, D. T. (2005). Ask the cognitive scientist: Do visual, auditory, and kinesthetic learners need visual, auditory, and kinesthetic instruction? *American Educator, 29*(2), 31–35.
- World Bank. (2018). *World Development Report 2018: Learning to realize education's promise*. World Bank. <https://doi.org/10.1596/978-1-4648-1096>.
- World Economic Forum (2014). 7 things that don't work in the classroom. <https://www.weforum.org/agenda/2014/10/7-things-dont-work-in-classroom/>
- Yoak, E. (2008). Review of *So Much Reform, So Little Change: The Persistence of Failure in Urban Schools* by Charles M. Payne. *Critical Questions in Education, 2*(1), 63–65. Retrieved July 15, 2021 from <https://academyedstudies.files.wordpress.com/2015/02/yoakvol2no1.pdf>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.