J Educ Change
DOI 10.1007/s10833-016-9294-4

What works may hurt: Side effects in education

Yong Zhao[1]

Abstract Medical research is held as a field for education to emulate. Education researchers have been urged to adopt randomized controlled trials, a more "scientific" research method believed to have resulted in the advances in medicine. But a much more important lesson education needs to borrow from medicine has been ignored. That is the study of side effects. Medical research is required to investigate both the intended effects of any medical interventions and their unintended adverse effects, or side effects. In contrast, educational research tends to focus only on proving the effectiveness of practices and policies in pursuit of "what works." It has generally ignored the potential harms that can result from what works. This article presents evidence that shows side effects are inseparable from effects. Both are the outcomes of the same intervention. This article further argues that studying and reporting side effects as part of studying effects will help advance education by settling long fought battles over practices and policies and move beyond the vicious cycle of pendulum swings in education.

Keywords Educational research · Methodology · RCT · Direct instruction · International assessment · Side effects · PISA · Educational policy · Educational reform

Medical research is held as a field for education to emulate (Bryk 2015; Slavin 2002). Education researchers have been urged to adopt randomized controlled trials (RCT), a more "scientific" research method believed to have resulted in the advances in medicine (Slavin 2002). The hope is that if educational research adopts this scientific method, it will lead to evidence-based (better) policies and practices and ultimately better education (Slavin 2002).

✉ Yong Zhao
yonhzhao@ku.edu

[1] 419 JRP, School of Education, University of Kansas, Lawrence, KS 66049, USA

Published online: 04 February 2017

Despite differing opinions that argue for the merits of other methods (Berliner 2002; Shavelson and Towne 2002; Slavin 2002, 2008), proponents of RCTs have won the day, at least for now. Over the past decade or so, the funding and desire for randomized controlled experiments in educational research has increased significantly. In essence, RCT is now the gold standard in educational research. For example, the *What Works Clearinghouse* (WWC) of the Institute of Educational Sciences (IES), which aims to "promote informed education decision making by providing educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence about "what works" in education" (What Works Clearinghouse 2015), accepts only studies using RCT as meeting "WWC Group Design Standards without Reservations." Quasi-experimental design can be accepted, but it only meets the standards with reservations (What Works Clearinghouse 2014).

It is uncertain whether RCTs should be the gold standard for educational research and whether they improve educational policies and practices. But there is another, more important lesson education should learn from medicine: a prevalent acute concern over side effects resulting from intervention or treatment. In the development and research of medical interventions, be they new drugs or medical procedures, it is essential to weigh the risks against their effectiveness. For example, the U.S. Food and Drug Administration (FDA) requires prior to research on a new drug, or "investigating a new drug (IND)", sponsors must "notify FDA and all participating investigators, in a written IND safety report, of any adverse experience associated with the use of the drug that was both serious and unexpected, and any finding from tests in laboratory animals that suggested a significant risk for human subjects." [U.S. Department of Health and Human Services (Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)), 2012, p. 2]. Clinical trials, the next phase of development, have similar requirements. Even after a drug is approved, research on side effects continues and the FDA continues to watch and actively encourage the reporting of previously unidentified side effects.

Why study side effects in education?

"Side effect" is defined in the online *Cambridge English Dictionary*[1] as "an unwanted or unexpected result or condition that comes along with the desired effects of something." It is universally accepted in medicine that any intervention can have unwanted effects that can damage the person it intends to heal. Medical research thus considers benefits and risks to be fundamentally inseparable: they are two sides of the same coin. When studying and reporting the benefits or effectiveness, one must also study and report the risks.

Studying and reporting side effects in trials can save lives by preventing products that may be effective but pose high health risks from entering the market. For example, the global pharmaceutical giant Pfizer terminated the development of

[1] http://dictionary.cambridge.org/us/dictionary/english/side-effect.

Torcetrapib, a drug that early studies showed to be effective in preventing heart attacks and strokes, because it triggered a higher rate of chest pains and heart failure than the control group during clinical trials (Ginsberg and Kingston 2014). Although the termination meant that Pfizer lost its more than $800 million-dollar investment, it likely saved many lives.

Moreover, considering side effects also helps advance the field. Medical research is not always about finding more effective treatment; it is also about minimizing side effects. Discovering new treatments that may be equally or less effective but cause less severe or fewer side effects have also motivated medical improvements. For example, aminopterin was used to treat children with leukemia but it had toxic effects, including "a troublesome stomatitis affecting the rapidly dividing lining cells of the mouth, leading to painful ulceration" (Sneader 2005, p. 251). A new drug, methotrezate replaced amintopterin because it caused less severe side effects. Just recently, a cooling cap system was developed and approved to minimize the side effect of hair loss for breast cancer patients undergoing chemotherapy (Kaplan 2016).

Studying and reporting both effects and side effects helps consumers make informed decisions. Doctors and patients can weigh the benefits and risks with the information. Some treatment may be more effective, but so are their side effects. While a treatment may be less effective, but so is its adverse effect. Thus in some cases, making choices about a medical treatment is extremely difficult. But being fully informed of known risks associated with the treatment definitely helps in these difficult times. This is why medical products are required to disclose their side effects and information about their effectiveness. This is true even for common over the counter drugs such as cold medicine. When you buy a bottle of Ibuprofen, for example, the label clearly indicates its effectiveness in relieving pain and reducing fever, but also its potential to cause severe allergic reactions and stomach bleeding.

Educational researchers neither share this belief, nor are they required to adopt similar practices. Much research in education focuses exclusively on proving or disproving the effectiveness of an intervention. It is extremely rare to find a study that evaluates both the effectiveness and adverse effects of a product, teaching method, or policy in education. I have not yet found an educational product that comes with a warning label carrying information such as "this program works in raising your students' test scores in reading, but may make them hate reading forever." The *What Works Clearing House* does not make it a requirement to report side effects. It only reports to what degree the intervention was effective, backed up with scientifically collected evidence. In other words, educational research seems to be exclusively interested in what works, but ignores the possibility that what works may hurt at the same time.

There are probably many reasons for the apparent lack of concern for risks of intervention in education. One of them is the positive perception of "education." Education is universally perceived to be good, as such very few people automatically associate education with any adverse effects. Thus, when we consider educational interventions, we only want to know if they are effective. Another possible reason is that damages caused by education may take a long time to be observed or felt, quite unlike negative reactions to medicine (although even in

medicine negative effects on health may only appear years later in some cases). It is thus rather difficult to study or find out about education's side effects. It is also possible that the narrow definition of educational outcomes makes it hard to observe adverse effects. The dominant definition of educational outcomes today is cognitive abilities in a few subjects, measured by standardized tests. Unsurprisingly, this means most educational studies use standardized test scores as the measure of effectiveness. But there are other outcomes that matter but are rarely measured, such as personal qualities and motivation (Duckworth and Yeager 2015; Zhao 2016a). Thus we would not know if an intervention found to be effective in promoting cognitive abilities might adversely affect personal qualities and motivation. Furthermore, there may be political, commercial, or other reasons to deliberately overlook side effects, just like some pharmaceutical companies might overlook or suppress the reporting of side effects in pursuit of profits.

Whatever the reason, the lack of concern about side effects has resulted in a number of serious problems in education. First, there have been long, unproductive, yet fierce battles over the good or bad of pedagogical approaches (e.g. direct instruction), and policies (e.g. charter schools). Their proponents can put up abundant evidence to prove their effectiveness, while their opponents have equally abundant evidence to prove their negative effects. More evidence, however scientifically collected, is unlikely to end the battles because both sides are right and wrong at the same time. They are simply looking at two different sides of the same coin. One side is only looking at the effects and the other, the side effects.

Second, policies and practices have been advocated and implemented without adequately studying and reporting their risks, thus resulting in more damages than benefits. These damaging outcomes, when later discovered and reported, are often cast away as unintended consequences. However, they could be prevented or minimized in advance if educational research had a similar concern about side effects as medicine.

Third, education seems to be stuck in perpetuating pendulum swings (Barker 2010; Cuban 1990; Kaestle 1985; Slavin 1989). Despite the many reforms over the years, researchers, practitioners, and the public feel there are no new ideas in education. Old ideas are recycled every few years, perhaps with new names. As a result, there are really no new ideas, no advancement, and no progress. There are many reasons for pendulum swings in education (Cuban 1990), but the lack of concern for both effects and side effects is one of the major contributors. When an idea is implemented without considering its adverse effects, it will be eventually abandoned when its negative effects eventually materialize. But when an idea is abandoned due to its side effects, it does not mean it is not effective, but that the risks outweigh the benefits. Thus when the succeeding idea's negative effects emerge, the old idea's positive effects are good reasons to bring it back. A way forward is to consider both positive and negative effects of an idea, work on maximizing its positive effects while minimizing its negative impact, instead of completely replacing the idea; an important lesson to learn from the emerging field of improvement science (Bryk 2015).

Finally, the idolization of education systems, institutions, and approaches often leads to ultimate disappointment without considering side effects. In recent years,

international organizations have been busy with identifying educational systems worth learning from. Singapore, Finland, Shanghai, and Korea have been promoted as education systems that others should look up to and learn from (Tucker 2011; Sahlberg 2011) but whatever is in these systems that resulted in their top performance may have negative side effects. Upon examining the side effects, these systems may look much less glorious and thus causes disappointment (Zhao 2014, 2015). Likewise, institutions and individuals who have garnered great honor for raising test scores turned out to have done so while causing significant collateral damages (Nichols and Berliner 2007).

These problems are good reasons for starting to treat side effects in educational research the same way they are treated in the field of medicine. In the remainder of this article, I discuss three examples of side effects in education and how considering side effects can help address some of the persistent problems in education and advance education.

Direct instruction: Instruction that stifles creativity

With hundreds, if not thousands, of studies, direct instruction[2] (DI) is perhaps one of the most researched pedagogical approaches[3] (e.g. Adams and Engelmann 1996; Becker and Gersten 1982; Brent and DiObilda 1993; Dean and Kuhn 2007; Gunn et al. 2000; Meyer 1984; Peterson 1979; Roehler and Duffy 1982; Schwerdt and Wuppermann 2011; Swanson and Sachse-Lee 2000). But despite the vast amount of research, there is no general agreement whether direct instruction is an effective approach. Essentially, the lack of a tradition of considering side effects as an integral part of effectiveness in educational research has resulted in two irreconcilable bodies of literature: one proving the effectiveness of direct instruction and the other condemning direct instruction for its potential negative side effects. As a result, the debate among proponents and opponents of the approach has degenerated into ideological and political wars, as exemplified by the math wars (Klein 2007) and reading wars (Pearson 2004).

The disputes cannot be settled with more rigorous research methods such as RCT. Supporters of direct instruction have presented abundant convincing evidence scientifically gathered over 40 years that supports its effectiveness in raising student academic achievement[4] (Hempenstall 2012, 2013). Nonetheless, they have failed to convince the critics. It is not because of their lack of data or rigorous research method or design they used to collect the evidence. Most of the opposition does not come from doubting the evidence-supported effectiveness of DI, but stems from a different set of concerns such as the rigidity and prescriptiveness of the approach,

[2] Direct instruction (di) here refers to the general pedagogical approach characterized by explicit instruction. It includes both the lower and upper case *di*s (Rosenshine 2008).

[3] The National Institute for Direct Instruction published a 102 page long bibliography of writings on Direct Instruction in 2015, each page containing about 12 entries (National Institute for Direct Instruction 2015).

[4] Although *What Works Clearing House* found the effects to be small or indiscernible based on its reviews of two programs using direct instruction (What Works Clearinghouse 2006, 2007).

inconsistency with developmental theories, inappropriateness for certain groups of children and contexts, sustainability of the effects over time, suppression of learner autonomy and development of creativity, and other potential damaging side effects (Adams and Engelmann 1996; Hempenstall 2013; Tarver 1998).

These concerns are essentially about the adverse effects of DI. But supporters of DI tend to refuse to acknowledge them as such. Instead they discount these concerns as "myths,"(Adams and Engelmann 1996; Hempenstall 2013; Tarver 1998) blaming ideological bias and lack of understanding of direct instruction. Or they try to produce more evidence, while their critics continue to show evidence of its adverse effects.

A more rational and productive approach would be for both sides to acknowledge that DI, like all medical products has effects and side effects. The evidence supporting DI's effectiveness seems overwhelming, but there is also evidence showing that it can cause damage. For example, a review article by Penelope L. Peterson in 1979 found that DI could suppress creativity and problem solving while boosting achievement test scores. After reviewing over 200 studies, she concluded:

> …with direct or traditional teaching, students tend to do slightly better on achievement tests, but they do slightly worse on tests of abstract thinking, such as creativity and problem solving. Conversely, with open teaching, students do somewhat worse on achievement tests, but they do somewhat better on creativity and problem solving. Furthermore, open approaches excel direct or traditional approaches in increasing students' attitudes toward school and toward the teacher and in improving students' independence and curiosity. In all these cases, the effects were small. (Peterson 1979, p. 47)

Peterson's observations seem theoretically reasonable. Direct instruction can be effective "in promoting rapid and efficient learning of target material," (Bonawitza et al. 2011, p. 322), but it can negatively impact creativity because "instruction necessarily limits the range of hypotheses children consider" (Bonawitza et al. 2011, p. 322) and their attempt to explore novel situations. This hypothesis was confirmed by two independent studies conducted in two separate labs. Both studies were published in the journal *Cognition* in 2011 (Bonawitza et al. 2011; Buchsbauma et al. 2011).

In one study, Bonawitza and colleagues conducted two experiments among preschoolers. In the first experiment, the children were randomly assigned into four conditions: pedagogical, interrupted, naïve, and baseline. The task was to play with a novel toy. In the *pedagogical* condition, the experimenter acted like a teacher using direct instruction. She told the children "Look at my toy! This is my toy. I'm going to show you how my toy works. Watch this!'' She then proceeded to demonstrate one of the multiple ways of playing with the toy. The *interrupted* condition had exactly the same treatment except that the experimenter interrupted herself and moved away from the scene immediately after the demonstration. In the *naïve* condition, the experimenter told the children she had just found the toy and as if by accident discovered the way to play with the toy by saying "Huh! Did you see that? Let me try to do that!" She then performed the same action. In the *baseline* condition, the experimenter did not demonstrate the use of the toy. She simply

called children's attention to the toy by saying: "Wow, see this toy? Look at this!" In all conditions, after the initial introduction, the experimenter encouraged the children to figure out how the toy worked and then left them to play with it (Bonawitza et al. 2011, p. 322).

The researchers video recorded all sessions and compared children's total time playing, the number of unique actions performed, the proportion of time spent on the demonstrated function, and the total number of functions discovered. Their data suggest:

> …that teaching constrains children's exploration and discovery. Children who were taught a function of a toy performed fewer kinds of actions of the toy and discovered fewer of its other functions, than children who did not receive a pedagogical demonstration, even though all children were explicitly encouraged to explore the toy (Bonawitza et al. 2011, p. 325).

The results of the second experiment further confirmed findings of the first. It found that children could infer pedagogical intentions. In other words, even if children are not directly instructed but are given the opportunity to overhear instructions to their peers, they believe teaching is happening and so they should follow the instructor.

The other study provides more evidence that direct instruction is efficient and effective in teaching targeted knowledge, but inhibits curiosity and creativity. The study was also conducted with a group of preschoolers using toys. The results show that children in situations where the experimenter adopted the role of an instructor and directly gave instructions and demonstrations were more likely to imitate the instructor than in other conditions. But they were found to be less likely to explore and come up with novel solutions (Buchsbauma et al. 2011).

Similar findings are reported in math learning (Kapur 2014). Students generated more solutions to the problems before instruction than after. Students who received instruction first tended to produce only the correct solutions they were told. Kapur (2016) suggests that instruction seems to constrain students' search for novel solutions, which is necessary for creativity and inventiveness.

"Unproductive success" is a concept applied to direct instruction by Manu Kapur in a recent article in *Educational Psychologist.* According to Kapur, "unproductive success" is intervention that "may maximize performance in the shorter term without maximizing learning in the longer term…it is possible for students to show high performance on memory tasks or carrying out problem-solving procedures without a commensurable understanding of what it is that they are doing" (Kapur 2016, p. 290). In a quasi-experimental study where students are placed in two different conditions: Direct Instruction where students received lectures from the teacher and Productive Failure where students were first asked to solve complex problems and then the teacher explains about "canonical" solutions. The study found that students in the direct instruction condition were initially more successful in solving well-structured problems (Kapur and Bielaczyc 2012). However, in the end, their performance on tasks that required deeper conceptual understandings was inferior to students under the Productive Failure condition.

Obviously we should be very cautious with drawing any general conclusions based on just a few studies. However, the philosophy underlying the studies is what we need in educational research. The approach that simultaneously studies effects and side effects of educational interventions holds great promise to resolve long fought battles with empirical evidence. These studies essentially confirm that DI can be effective in knowledge transmission AND suppress creativity and curiosity. In education, we need both effective ways to transmit knowledge and foster creativity. Thus DI has its place in education. However its side effects need to be minimized. To advance education it is thus more important to direct efforts to explore when DI should be used for what purpose and population as well as strategies to mitigate its adverse effects on creativity.

More important, studies like these give consumers information to make choices. Knowing the effects and side effects of educational interventions, consumers (policy makers, educators, parents, and students) can decide what interventions they would adopt and the potential risks and benefits of their decisions. In the case of DI, if further investigations confirm that the effectiveness in promoting rapid and efficient mastery of knowledge and skills of direct instruction comes at the expense of creativity and curiosity, education consumers can choose if, when, or how much they wish to adopt direct instruction or discovery learning. What is of particular importance is whether the damage to creativity and curiosity is long term and extends beyond the immediate situation. A one-time treatment of direct instruction is unlikely to inhibit children's curiosity and creativity for life. But what if children are exposed to only direct instruction for 12 years or longer? Would it cause them to become less creative?

The best or the worst: The conflicting evidence of performance

In recent years, as interest in learning from top performing educations systems grows, international assessment programs have become the *de facto* arbiter of the quality of education in the world, using students' test scores as evidence. As a result, East Asian education systems have been branded as the best education systems in the world because their students have consistently topped the world in the two most influential international testing programs: Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA). Since the 1960s when the First International Mathematics Study (FIMS) was conducted, East Asian education systems have ranked the best in test scores (Baker 2007; OECD 2011, 2013a, 2014; Zhao 2012a, 2015).

It is thus no surprise that East Asian education systems have become the object of idolization and source of ideas for improving education. Eager admirers and learners from around the globe have come to them for lessons in policy, curriculum, school management, teachers and teaching. Academic scholars, policy wonks, journalists, and casual observers have produced numerous articles, books, documentary films, and blog posts praising these systems and making a wide range of recommendations for other systems around the world (Bieber and Martens 2011; Darling-Hammond and Lieberman 2012; Figazzolo 2009; Jensen 2012; Lamb

and Fullarton 2002; Meyer and Benavot 2013; National Research Council 1999; Nelson 2002; OECD 2011; Schleicher 2013; Schmidt 1999; Tucker 2011, 2014). Some of these recommendations have been taken seriously and implemented. For example, much of the argument for international benchmarking in the U.S. Common Core State Standards was fueled by Eastern Asian education systems' superior performance (Common Core State Standards Initiative 2011). The U.K. has decided to have half of its primary schools, about 8000, adopt the math teaching method practiced in Hong Kong, Singapore, and Shanghai because of their performance (Telegraph Reporters 2016). It is safe to say that Eastern Asian education systems have had a significant impact on educational policies and practices around the world in the past two decades.

There is no shortage of critics who question East Asia's superior performance over the years. Some question the validity and reliability of the methods employed by PISA and TIMSS (Feniger and Lefstein 2014; Kreiner and Christensen 2014; Morrison 2013; Stewart 2013), others doubt the sampling of participants (Loveless 2014). There certainly are issues with the tests that identified Eastern Asian education as the best systems, but the real big problem is the adverse effects these systems may have on their children.

East Asia's students have ranked among the lowest in terms of their reported confidence in mathematics, despite their top performance in test scores (Zhao 2016c, d). In the 1995 TIMSS results, for example, the percentages of students reporting themselves as 'very confident' in math and science in 1995 in East Asian systems were much lower than in Australia, the UK, and the US (Figs. 1, 2). Compared to U.S. students, who have always scored much lower in TIMSS mathematics, East Asian students have consistently reported lower confidence in math (Zhao 2016d).

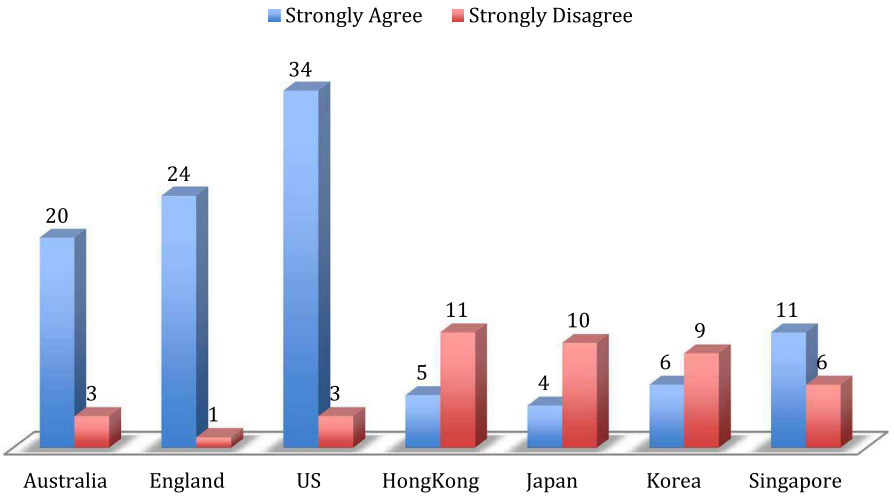

Fig. 1 Percentage confident in maths ('Usually do well in Maths'), eighth grade, TIMMS 1995

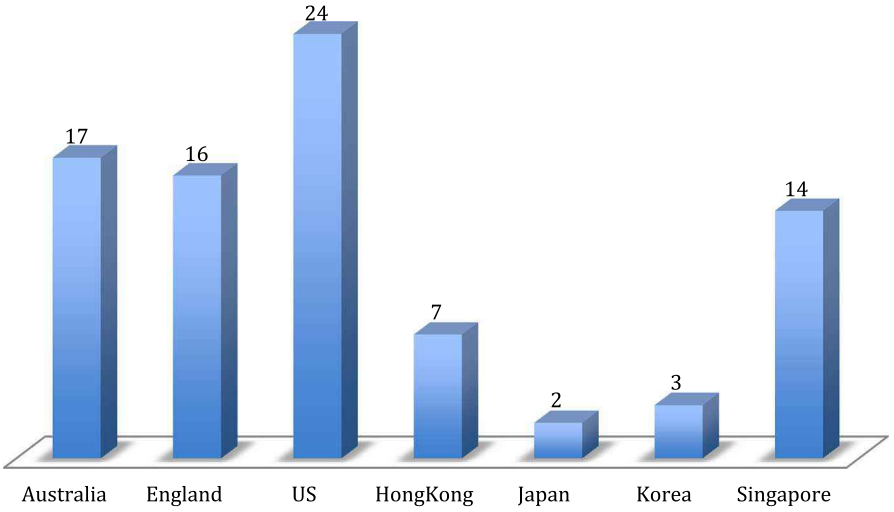

Fig. 2 Percentage confident in science, eighth grade, TIMMS 1995

PISA has shown similar results. In the 2012 PISA study, students in East Asian systems reported the lowest self-confidence in the world (OECD 2013b). A high proportion of students in these educational systems worried they "will get poor grades in mathematics." More than 70% of students in Korea, Chinese Taipei, Singapore, Viet Nam, Shanghai-China, and Hong Kong-China—in contrast to less than 50% in Austria, United States, Germany, Denmark, Sweden, and the Netherlands—"agreed" or "strongly agreed" that they worry about getting poor grades in math (OECD 2013b).

East Asian students have also ranked low in their attitudes toward the tested subjects: math, reading, and science. They tend to like the subjects less and value them less than students in other countries (Mullis et al. 2008, 2012; OECD 2013b; Zhao 2014, 2016c, d). Moreover, while their performance in tests continued to improve over the years, their confidence and attitudes have not. Instead, they seem to decline (Zhao 2016c, d).

This seems counterintuitive because it is more reasonable to expect high performance should make the students more confident and more interested in the subjects. This is in fact the case within education systems. For example, TIMSS has reported positive correlations between test scores and confidence as well as attitude within education systems (Loveless 2006). In other words, at the individual level, students who scored higher tend to be more confident within education systems.

Thus the fact that East Asian students have low confidence is a system effect. That is, these systems have a small percentage of students who have confidence in their math. It does not mean that high-performing students within these systems necessarily have low confidence or vice versa. Instead, it means that these systems have somehow made a large number of students lose confidence and interest in math, science, and reading, while helping them achieve excellence in testing.

Therefore it is not unreasonable to hypothesize that these educational systems may be effective in preparing students to achieve excellent scores AND effective in lowering their confidence and interest. They help with improving test scores but hinder the development of confidence and interest. This hypothesis has not been confirmed with large scale and longitudinal studies, but there is preliminary evidence.

One piece of evidence is the negative correlation between students' confidence and test scores across educational systems. In the 2003 TIMSS, math scores and confidence in the subject were significantly negatively correlated, in both 4th and 8th grade ($r = -0.59$ and $r = -0.64$ respectively) (Loveless 2006). The correlation between nations' average scores and enjoyment is also negative ($r = -0.67$ for 4th grade and $r = -0.76$ for 8th grade). Similar negative relationships exist between PISA scores and students' non-cognitive qualities. "It seems that pupils in high-scoring countries also develop the most negative attitudes to the subject" (Sjøberg 2012, p. 1). There is a significant negative correlation between students' self-efficacy in science and their scores in the subject across education systems in the 2015 PISA results, for example (OECD 2016). Additionally, PISA scores have been found to have a significant negative correlation with entrepreneurial confidence and intentions (Zhao 2012b).

Another piece of evidence is the peculiar trend in TIMSS outcomes over the past two decades. While East Asian education systems continued the upward trend in test scores, there has been a slight decline in confidence and attitude. The same trend is observed for the United States (Zhao 2016d). This appears to suggest that strategies that raise test scores may indeed negatively affect confidence and attitude.

Learning from others is always a good idea, but we must do so with great caution. The East Asian systems may be the best at producing outstanding test scores, they are also the worst in cultivating confidence in and a positive attitude toward the subjects. If indeed the policies and practices that raise test scores also hurt confidence and attitude, we must carefully weigh the risks against the benefits. Do we care more about test scores or confidence and attitude?

When risks outweigh benefits: Test-based accountability

America could have avoided the significant damages caused by test-based accountability if side effects had been taken seriously in educational research and policy making. Although the No Child Left Behind Act (NCLB) ("No Child Left Behind Act of 2001," 2002), the law that made standardized testing a common fixture of American education, has been widely condemned for the damages it has caused, most of the criticism came too late. The damages are treated as unintended consequences, only to be uncovered afterwards. The damages could have been avoided if studying and reporting side effects had been a requirement for all proponents of policies and practices. If NCLB proponents had carefully studied and publicly disclosed its potential risks, would it have received such overwhelming support from both sides of the aisle in congress and the general public?

The adverse effects of testing-based accountability imposed by NCLB are extensive (Ginsberg and Kingston 2014; Hess 2011; Nichols and Berliner 2007; Ravitch 2013; Tienken and Zhao 2013). Nichols and Berliner (2007) report a wide range of negative effects of high-stakes testing on American public education. The prevalence of various forms of cheating on standardized testing is one of them. Numerous instances of cheating by school administrators and teachers have been reported in recent years (Toppo et al. 2011), including the infamous case in Atlanta Public Schools (Vogell 2011; Zhao 2014). The exact magnitude of the problem is difficult to quantify, but some surveys found that about 10% of teachers and administrators admit to some form of cheating or assisting their students to obtain better results (Nichols and Berliner 2007). Without any intention to defend the unethical behaviors of cheating, Nichols and Berliner believe that high stakes testing has placed teachers "in untenable positions, environments where pressures encourage questionable behavior" (2007, p. 34).

Another side effect, or collateral damage, of high stakes testing is the exclusion of certain types of students from education. Nichols and Berliner found "considerable evidence that some educators have shaped the test-taking pool in their schools or districts through such extraordinary practices as withdrawing students from attendance rolls. Others have purposefully demoralized students, causing them to give up…" (Nichols and Berliner 2007, p. 57). For example, Lorenzo Garcia, former superintendent of El Paso, kept almost half of the students eligible for 10th grade from taking the 10th grade exam by not allowing them to enroll in the school, retaining them at 9th grade, or rushing them into 11th grade (Sanchez 2013). As a result, "[h]igh-stakes testing creates conditions in which a great number of our most vulnerable and less advantaged students are denied a chance at a productive life" (Nichols and Berliner 2007, p. 57).

Curriculum narrowing is one more form of side effect causing damages by denying vulnerable and less advantaged students of a true education (Ginsberg and Kingston 2014; Hess 2011; Tienken and Zhao 2013). NCLB has led to a significant narrowing of the education experiences of American children (McMurrer 2007). With increased emphasis on improving test results in literacy and numeracy, schools had to reduce the amount of time and resources devoted to other subjects and school activities. The reduction is much more significant for schools serving disadvantaged students than for those serving more advantaged children (Tienken and Zhao 2013).

Additionally, high stakes testing has been associated with the distortion of instruction, turning teaching into test preparation (Ginsberg and Kingston 2014). Curriculum narrowing and test preparation are likely to undermine "the meaningfulness (or validity) of the resultant test scores" (Nichols and Berliner 2007, p. 141). Furthermore, it has led to a decline of ethical behaviors in state education departments. Multiple instances of states and districts manipulating drop out rates and misrepresented test results have been documented (Nichols and Berliner 2007). Finally, Nichols and Berliner (2007) suggest that high stakes testing has the effect of undermining American education and demoralizing both teachers and students.

Yet, all these damages came without much benefit. NCLB did not deliver its promise to close the achievement gap and raise academic achievement (Hout and Elliott 2011). In other words, NCLB caused all these damages for nothing. The

achievement gaps remain (Zhao 2016b) and students' academic achievement has not significantly improved (Brown 2016).

A call to study side effects

These examples suggest that educational programs, approaches, and policies are not unlike medical products: when they cure, they can hurt. They also suggest that considering both main and side effects can help resolve artificially divisive issues in education and help advance the field. They further suggest that the negative effects of educational policies, programs, and products have not gone entirely unnoticed. There indeed exists a body of literature that reports the negative effects of certain education approaches, programs, and policies. However, the negative effects of educational products have not been treated the same way as side effects of medical products in a number of crucial aspects.

First, there is no regulation that asks developers of education interventions to study and disclose potential side effects when providing evidence for their effectiveness. As a result, the majority of educational product developers and proponents have focused exclusively on marshaling evidence to show benefits and effects. Even review and synthesis studies such as the numerous meta-analysis studies have been conducted to prove or disprove the effectiveness of certain approaches or policies (e.g. Hattie 2009; Shakeel et al. 2016; What Works Clearinghouse 2014), without much attention to the potential negative effects. Thus consumers (teachers, parents, education leaders, students, and the public) only have information of what works, without knowledge of the potential costs associated with it. In cases where potential damages may be greater than benefits, it is perhaps better not to adopt the product even if it is effective in some way.

Second, the negative effects of educational products, when occasionally discovered, are not considered an inherent quality of the product or policy. Rather, they are often treated as unintended or unanticipated consequences or results of poor implementation. While not all negative effects can be anticipated and it is reasonable to believe that policy makers or product developers in education intend to do no harm, some negative effects can be predicted based on past experiences and sound reasoning in advance of their actualization. For example, the side effects or collateral damages of NCLB reported by Nichols and Berliner could have been anticipated based on Campbell's Law, which states: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell 1976, p. 49). The corruption resulted from NCLB thus could have been avoided or at least mitigated had the policy makers heeded the warning of Donald Campbell.

Third, when side effects in education are occasionally reported, they often come from opponents and critics of certain products. But the opponents and critics often do not consider impartially the effects of the product or policy, nor do they have access to or resources to conduct original studies concerning the product or policy. As a result, the reported side effects are often brushed aside as lacking objectivity or

scientific rigor, or motivated by ideology. This is one of the reasons behind the long lasting "wars" in education—with two bodies of opposing literature co-existing in parallel places without much genuine interactions. As a result, much energy is wasted in research that does not lead to improvement.

Education as a field has been slow to improve because it has largely failed to build on past experiences (Bryk 2015). One way to build on past experiences is to consider side effects as an integral aspect of research. Research on side effects can force product and policy developers to improve their products and policies so as to minimize side effects or to develop alternative products and policies that cause less damage. It can also better guide consumers in their decision to choose the products and policies that best suite them, considering both effects and side effects. Moreover, seeking to understand side effects can prevent damaging products and policies, despite their effectiveness, to be adopted in advance. If the risks of a product are higher than its benefits, it should not be allowed to enter schools.

Recommendations

The field of education is stuck in wars over many interventions, some old and some new. Besides direct instruction, there are fights over class size, charter schools, school vouchers, the common core, teacher merit pay, teacher certification, testing-based accountability, to name just a few. These are all expensive and consuming battles that can have significant and long lasting impacts on education.

Education can benefit much from making side effects an essential and important aspect of any study that attempts to prove the effectiveness of educational products, just like medical research. We should learn from medical research and begin to build a culture that makes side effects an integral part of our research. We can start by taking a few actions:

1. Research organizations such as the American Educational Research Association (AERA) and academic journals can require research articles to include both main effects and side effects. Such a requirement would force or encourage educational researchers to pay attention to side effects in their studies. It is not likely to expect a government organization like the FDA to issue such a requirement, but it is quite possible for leading organizations and journals to start the movement.
2. Federal clearing houses such as what works should consider and include information about the negative effects of educational approaches, methods, products, or policies. This action would serve two purposes. First, it serves as an encouragement for educational researchers and product developers to seriously consider side effects. Second, it provides consumers—educators, parents, and policy makers—to make informed decisions.
3. Education researchers, policy makers, and product developers should voluntarily study side effects and disclose such information. Eventually I hope all education products and policies will carry a warning label that discloses confirmed and potential negative effects.

4. Consumers of educational research, policy, and products should ask for information about both effects and side effects. For example, when a publisher comes to promote products, school leaders, teachers, and parents should also always demand information about side effects of the products. The public should expect to know side effects of proposed policies and practices.
5. Program evaluation should include investigating both effects and side effects. In other words, we cannot only look at how effective a program has been, but also what damages it might have done. For example, when looking at school privatization and vouchers, we cannot claim success because participants' test scores are better (Shakeel et al. 2016). We must also look at the potential negative impact on the development of participants' other abilities, teachers, and even non-participating students affected by the program.
6. Reports of side effects after the implementation of interventions should be considered seriously. Instead of discarding them as unintended consequences or improper implementation, or simply complaints by unhappy parents, students, or teachers, it is the responsibility of the policy and product developers' to investigate and respond to such reports. The FDA monitors side effects and recalls products all the time when a product's risks outweigh benefits.

No doubt these recommended actions also have side effects that need to be addressed. When we begin to address the side effects of these recommendations, education would have advanced. This is how we improve.

Acknowledgements Ken Frank of Michigan State University, James Basham and Jason Travers of University of Kansas, and Yechen Zhao of Stanford University read drafts of the manuscript and provided invaluable suggestions.

References

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

Baker, K. (2007). Are international tests worth anything? *Phi Delta Kappan, 89*(2), 101–104.

Barker, B. (2010). *The pendulum swings: Transforming school reform*. Sterling, VA: Trentham Books.

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.

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18–20.

Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education, 46*(1), 101–116. doi:10.1111/j.1465-3435.2010.01462.x.

Bonawitza, E., Shaftob, P., Gweonc, H., Goodmand, N. D., Spelkee, E., & Schulzc, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322–330.

Brent, G., & DiObilda, N. (1993). Effects of curriculum alignment versus direct instruction on urban children. *The Journal of Educational Research, 86*(6), 333–338.

Brown, E. (2016, April 27). U.S. high school seniors slip in math and show no improvement in reading. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/us-high-school-seniors-slip-in-math-and-show-no-improvement-in-reading/2016/04/26/9b8f033e-0bc8-11e6-8ab8-9ad050f76d7d_story.html?utm_term=.7a9d458243cd.

Bryk, A. S. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Buchsbauma, D., Gopnika, A., Griffithsa, T. L., & Shaftob, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120*(3), 331–340.

Campbell, D. T. (1976). *Assessing the impact of planned social change*. Hanover New Hampshire, USA. Retrieved from https://www.globalhivmeinfo.org/CapacityBuilding/OccasionalPapers/08AssessingtheImpactofPlannedSocialChange.pdf.

Common Core State Standards Initiative. (2011). *Common core state standards initiative*. Retrieved from http://www.corestandards.org/.

Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3–13.

Darling-Hammond, L., & Lieberman, A. (Eds.). (2012). *Teacher education around the world*. New York: Routledge.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384–397.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251.

Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Education Policy*. doi:10.1080/02680939.2014.892156.

Figazzolo, L. (2009). *Impact of PISA 2006 on the education policy debate*. Retrieved from http://download.ei-ie.org/docs/IRISDocuments/ResearchWebsiteDocuments/2009-00036-01-E.pdf.

Ginsberg, R., & Kingston, N. (2014). Caught in a vise: The challenges facing teacher preparation in an era of accountability. *Teachers College Record, 116*(1), n1.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 34*(2), 90–103.

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

Hempenstall, K. (2012, November 11). *Reviews supporting direct instruction program effectiveness*. Retrieved from https://www.nifdi.org/news/hempenstall-blog/403-reviews-supporting-direct-instruction-program-effectiveness.

Hempenstall, K. (2013, October 10). *Why does direct instruction evoke such rancour?* Retrieved from https://www.nifdi.org/news-latest-2/blog-hempenstall/389-why-does-direct-instruction-evoke-such-rancour.

Hess, F. M. (2011). Our achievement-gap mania. *National Affairs, Fall 2011* (Number 9), 113–129.

Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. Washington, DC: National Academies Press.

Jensen, B. (2012). *Catching up: Learning from the best school systems in East Asia*. Retrieved from Melbourne: http://www.grattan.edu.au/publications/129_report_learning_from_the_best_main.pdf.

Kaestle, C. (1985). Education reform and the swinging pendulum. *Phi Delta Kappan, 66*(6), 422–423.

Kaplan, K. (2016, December 9). Cooling cap helps cancer patients preserve their hair during chemotherapy, clinical trial shows. *The Los Angeles Times*. Retrieved from http://www.latimes.com/science/sciencenow/la-sci-sn-cooling-scalp-chemotherapy-20161209-story.html.

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83.

Klein, D. (2007). A quarter century of US 'math wars' and political partisanship. *BSHM Bulletin: Journal of the British Society for the History of Mathematics, 22*(1), 22–33.

Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*. doi:10.1007/s11336-013-9347-z.

Lamb, S., & Fullarton, S. (2002). Classroom and school factors affecting mathematics achievement: A comparative study of Australia and the United States Using TIMSS. *Australian Journal of Education, 46*(2), 154–171.

Loveless, T. (2006). *How well are American students learning?* Retrieved from Washington, DC: http://www.brookings.edu/~/media/Files/rc/reports/2006/10education_loveless/10education_loveless.pdf.

Loveless, T. (2014). *PISA's China problem continues: A response to Schleicher, Zhang, and Tucker*. Retrieved from http://www.brookings.edu/research/papers/2014/01/08-shanghai-pisa-loveless.

McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Retrieved from Washington, DC: http://www.cep-dc.org/displayDocument.cfm?DocumentID=312.

Meyer, L. A. (1984). Long-term academic effects of the direct instruction project follow through. *The Elementary School Journal, 84*(4), 380–394.

Meyer, H.-D., & Benavot, A. (2013). *PISA, power, and policy: The emergence of global educational governance*. Oxford: Oxford University Press.

Morrison, H. (2013, December 1). *Pisa 2012 major flaw exposed*. Retrieved from https://paceni.wordpress.com/2013/12/01/pisa-2012-major-flaw-exposed/.

Mullis, I. V. S., Martin, M. O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). *TIMSS 2007 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Retrieved from Chestnut Hill, MA.

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Retrieved from Chestnut Hill, MA.

National Institute for Direct Instruction. (2015). *Writings on direct instruction: A bibliography*. Retrieved from Eugene, OR.

National Research Council. (1999). *Global perspectives for local action: Using TIMSS to improve U.S. mathematics and science education*. Washington, D.C.: National Academy Press.

Nelson, D. I. (2002). *Using TIMSS to inform policy and practice at the local level. CPRE policy briefs. Access ERIC: FullText* (CPRE-36). Retrieved from Pennsylvania: http://www.gse.upenn.edu/cpre/. For full text: http://www.cpre.org/Publications/rb36.pdf. Provider: OCLC.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

No Child Left Behind Act of 2001, 107–110, Congress (2002).

OECD. (2011). *Strong performers and successful reformers in education: Lessons from PISA for the United States*. Retrieved from Paris: http://dx.doi.org/10.1787/9789264096660-en.

OECD. (2013a). *Programme for international student assessment (PISA)*. Retrieved from http://www.oecd.org/pisa/aboutpisa/.

OECD. (2013b). *Ready to learn: Students engagement, drive, and self-beliefs*. Retrieved from Paris: http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-III.pdf.

OECD. (2014). *PISA 2012 results: What students know and can do: Student performance in mathematics, reading and science (Volume I) [Revised edition February 2014]*. Retrieved from Paris: http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-i.htm.

OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education*. Retrieved from Paris: http://dx.doi.org/10.1787/9789264266490-en.

Pearson, P. D. (2004). The Reading wars. *Educational Policy, 18*(1), 216–252.

Peterson, P. L. (1979). Direct instruction: Effective for what and for whom. *Educational Leadership, 37*(1), 46–48.

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York: Knopf.

Roehler, L. R., & Duffy, G. G. (1982). Matching direct instruction to reading outcomes. *Language Arts, 59*(5), 476–480.

Rosenshine, B. (2008). *Five meanings of direct instruction*. Retrieved from Lincoln, IL: http://www.centerii.org/search/Resources%5CFiveDirectInstruct.pdf.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?*. New York: Teachers College Press.

Sanchez, C. (Writer). (2013). El Paso schools cheating scandal: Who's accountable? Washington DC: NPR.

Schleicher, A. (2013, December 3). *What we learn from the PISA 2012 results*. Retrieved from http://oecdeducationtoday.blogspot.com/2013/12/what-we-learn-from-pisa-2012-results.html.

Schmidt, W. H. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Dordrecht; Boston: Kluwer Academic Publishers.

Schwerdt, G., & Wuppermann, A. C. (2011). Sage on the stage: Is lecturing really all that bad? *Education Next, 11*(3), 63–67.

Shakeel, M. D., Anderson, K. P., & Wolf, P. J. (2016). *The participant effects of private school vouchers across the globe: A meta analytic and systematic review*. Retrieved from Little Rock: http://www.uaedreform.org/downloads/2016/05/the-participant-effects-of-private-school-vouchers-across-the-globe-a-meta-analytic-and-systematic-review-2.pdf.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington DC: National Academy Press.

Sjøberg, S. (2012). PISA: Politics, fundamental problems and intriguing results *Recherches en Education, 14*. Retrieved from http://www.recherches-en-education.net/spip.php?article140.

Slavin, R. E. (1989). PET and the pendulum: Faddism in education and how to stop it. *Phi Delta Kappan, 70*(10), 752–758.

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher, 31*(7), 15–21.

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5–14.

Sneader, W. (2005). *Drug discovery: A history*. Hoboken, NJ: Wiley.

Stewart, W. (2013, December 3rd). *Is Pisa fundamentally flawed?* Retrieved from http://www.tes.co.uk/article.aspx?storycode=6344672.

Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities, 33*(2), 114–136.

Tarver, S. G. (1998). Myths and truths about direct instruction. *Effective School Practices, 17*(1), 18–22.

Telegraph Reporters. (2016, July 23). *Half of primary school pupils to follow Chinese style of learning maths with focus on whole-class teaching*. Retrieved from http://www.telegraph.co.uk/education/2016/07/12/half-of-primary-school-pupils-to-follow-chinese-style-of-learnin/.

Tienken, C. H., & Zhao, Y. (2013). How common standards and standardized testing widen the opportunity gap. In P. L. Carter & K. G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 113–122). New York: Oxford University Press.

Toppo, G., Amos, D., Gillum, J., & Upton, J. (2011, March 17). When test scores seem too good to believe. *USA TODAY*. Retrieved from http://www.usatoday.com/news/education/2011-03-06-school-testing_N.htm.

Tucker, M. (Ed.). (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Boston: Harvard Education Press.

Tucker, M. (2014). *Chinese lessons: Shanghai's rise to the top of the PISA league tables*. Retrieved from Washington, DC: http://www.ncee.org/wp-content/uploads/2013/10/ChineseLessonsWeb.pdf.

U.S. Department of Health and Human Services (Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER)). (2012). *Guidance for industry and investigators: Safety reporting requirements for INDs and BA/BE studies*. Retrieved from Silver Spring, MD: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM227351.pdf.

Vogell, H. (2011, July 6, 2011). Investigation into APS cheating finds unethical behavior across every level. *The Atlanta Journal-Constitution*. Retrieved from http://www.ajc.com/news/investigation-into-aps-cheating-1001375.html.

What Works Clearinghouse. (2006). *Reading mastery*. Retrieved from http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=417.

What Works Clearinghouse. (2007). *Direct instruction*. Retrieved from http://ies.ed.gov/ncee/wwc/interventionreport.aspx?sid=139.

What Works Clearinghouse. (2014). *What works clearinghouse: Procedures and standards handbook version 3.0* Retrieved from Washington, DC.

What Works Clearinghouse. (2015). *Frequently asked questions*. Retrieved from http://ies.ed.gov/ncee/wwc/Document.aspx?sid=15-wwc.

Zhao, Y. (2012a, December 11). *Numbers can lie: What TIMSS and PISA truly tell us, if anything?* Retrieved from http://zhaolearning.com/2012/12/11/numbers-can-lie-what-timss-and-pisa-truly-tell-us-if-anything/.

Zhao, Y. (2012b). *World class learners: Educating creative and entrepreneurial students*. Thousand Oaks, CA: Corwin.

Zhao, Y. (2014). *Who's afraid of the big bad dragon: Why China has the best (and worst) education system in the world*. San Francisco: Jossey-Bass.

Zhao, Y. (2015). *Lessons that matter: What we should learn from Asian school systems*. Retrieved from Melbourne: http://www.mitchellinstitute.org.au/reports/lessons-that-matter-what-should-we-learn-from-asias-school-systems/.

Zhao, Y. (2016a). *Counting what counts: Reframing education outcomes*. Bloomington, IN: Solution Tree Press.

Zhao, Y. (2016b). From deficiency to strength: Shifting the mindset about education inequality. *Journal of Social Issues, 72*(4), 716–735.

Zhao, Y. (2016c, November 29). *It must be the chopsticks: The less reported findings of 2015 TIMSS and explaining the East Asian outstanding performance*. Retrieved from http://zhaolearning.com/2016/11/29/it-must-be-chopsticks-the-less-reported-findings-of-2015-timss-and-explaining-the-east-asian-outstanding-performance/.

Zhao, Y. (2016d, November 29). *Stop copying others: TIMSS lessons for America*. Retrieved from http://zhaolearning.com/2016/11/30/stop-copying-others-timss-lessons-for-america/.