

U.S. Department of Education
May 2016

Going public: Writing about effectiveness studies and their results

Mark Dynarski
Pemberton Research



REL 2016-173

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased, large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

May 2016

This report was prepared for the Institute of Education Sciences (IES) by Decision Information Resources, Inc. under Contract ED-IES-12-C-0057, Analytic Technical Assistance and Development. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Dynarski, M. (2016). *Going public: Writing about effectiveness studies and their results* (REL 2016-173). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

Contents

Writing about effectiveness studies and their results	1
Tip 1: Make the contrast clear	2
Tip 2: Make causal statements only when they result from causal research designs	3
Tip 3: Present numbers simply and concretely	8
Tip 4: Describe effects in meaningful units	10
Tip 5: Present findings as information for policy	13
Putting the parts together	15
References	16
Figures	
Figure 1. Differences between experiments and quasi-experiments	6
Figure 2. A simple graphic displaying numbers	10
Boxes	
Box A. Examples of unclear and clear contrasts in a study	3
Box B. Types of experiments	4
Box C. Accurate and inaccurate causal statements	7
Box D. Numbers presented to facilitate readers' understanding	9
Box E. Four ways to put effect sizes into context	12
Box F. Examples of presenting findings as information for policy	14

Writing about effectiveness studies and their results

Studies of effectiveness in education are now numerous, and a professional society and its journal are devoted to publishing findings from them. Yet, a common complaint of policymakers is that effectiveness studies are not easily understood. A recent report on interviews with policymakers about their use of research concluded that “communication of findings could be enhanced by reporting research evidence in more succinct, nontechnical, and readable formats” (Nelson et al. 2009). An effectiveness study that is unclear or too difficult to read is a lost opportunity to convey findings and to stimulate more research and innovation.

Effectiveness studies in education measure whether a program, policy, or approach improves outcomes. Five aspects of an effectiveness study, if written clearly, will enable readers to grasp findings more readily:

First, the study contrasts one or more approaches or interventions. Describing the contrast accurately is a starting point. Typically, the contrast is defined by the kinds of services received by study participants.

Second, the study uses a research design, such as an experiment or a quasi-experiment, to select students and schools with which the contrast will be implemented. Some designs yield effects that are causal and others do not. The choice of design creates limitations on the findings; the study needs to acknowledge those limitations so that its readers can understand what has been learned.

Third, the study needs to relate numbers, possibly many numbers, which could be characteristics of the sample, measured effects, or manipulations of measured effects. Conveying many numbers clearly is a challenge.

Fourth, as its name implies, an effectiveness study yields measures of one or more effects. Putting those effects into context—*are they small or large relative to some benchmark or to a gap that policy wants to close?*—helps readers to gauge their importance.

Fifth, the study’s findings have implications for policy, and explaining implications can be the point at which the writer conveys the study’s main takeaway. It is a point where writers could claim too much or too little. Writing clearly about implications helps readers know what key points they should take away from the findings.

The 5 “tips” or guidelines presented in this paper are intended to help writers convey the five key aspects of a study to non-expert readers. Reports have other components that writers may need to present briefly or at length: what previous research has found, the rationale for the current study, key research questions, properties of data collection instruments, and so on. Whether to discuss these other components and at what length will depend on the particular context of a study. But when writing about effectiveness studies, it is important to include a description of the five key aspects presented here, and emphasizing these aspects in a study’s executive summary will be useful.

The audience for this guide is researchers writing for policy readers. Researchers could be in academic settings, research organizations, or research departments of government organizations, such as federal or state agencies

or school districts. Policy readers may be in government, nonprofit organizations, or the media, or they may be members of the informed public. Though aimed at writers, the guide may also help policy readers understand what they should be looking for when they read effectiveness studies. A forthcoming companion document will go into depth on how to write descriptive analyses, and an earlier brief (<http://ies.ed.gov/ncee/pubs/REL2014051/>) is focused on writing research concepts in everyday language.

Tip 1: Make the contrast clear

Ask yourself:

Does the study clearly describe the difference between what is happening with the treatment group—the intervention, policy, or approach being tested—and what is happening with the control group?

An effectiveness study centers on a *contrast*. It can take many forms: the difference between two approaches for teaching reading, between using a software package to support math instruction or using a conventional textbook, between students attending charter schools and students attending traditional public schools, between schools that use different turnaround models, and so on. A study can have more than one contrast—perhaps it contrasts three approaches for teaching reading—but it always has to have at least one.

Studies of innovative or new approaches often focus on describing that approach and measuring its effects relative to current practice. The term “business as usual” is sometimes used to describe current practice: “the reading intervention was contrasted with business as usual.”

What this “business as usual” is may be evident to researchers and to the schools or classrooms in which the study is being conducted. When researchers writing about an effectiveness study use it as a shorthand phrase, it has two problems. First, a reader unfamiliar with the context may have no idea what “business as usual” is. Second, if a study is being conducted in more than one location, business as usual in one location may differ from business as usual in another. Two neighboring school districts may teach reading differently, and two school districts in different states almost certainly do.

Box A presents an example of a clear and unclear contrast. The first passage gives readers half of a whole. They learn that the treatment was the “Algebra Works” curriculum. However, unless readers are math teachers in the control-group schools, they do not learn what the treatment is being contrasted with. Ultimately, no matter whether the study reported that students in treatment schools performed better, worse, or the same on an algebra test, readers will have to wonder what has been learned, because they do not know what the intervention was contrasted with. The second passage gives readers the other half. The “Algebra Works” curriculum is being contrasted with the “Algebra Now” curriculum.

The example may seem contrived—the only difference between the two passages is that the second one names the curriculum used by the control group. But the difference highlights the point: the key to presenting a clear contrast is *viewing it as two arms of equal importance, as if two interventions were being*

compared. In social science research, business as usual is not the lack of services. It consists of something. If a study finds no differences between outcomes of the treatment and control groups—that is, “null effects”—the conclusion is not that the treatment is “ineffective.” The two interventions are *equally* effective. Writers should ensure that readers understand the contrast so that they come away with the right conclusion from the study.

Box A. Examples of unclear and clear contrasts in a study

An unclear contrast:

Schools randomized to the treatment group implemented the “Algebra Works!” curriculum, which combined lessons from its textbook with lessons in a computer lab. Schools assigned to the control group continued to teach Algebra as they previously had.

A clear contrast:

Schools randomized to the treatment group implemented the “Algebra Works!” curriculum, which combined lessons from its textbook with lessons in a computer lab. Schools assigned to the control group continued to use their existing Algebra 1 curriculum based on the “Algebra Now” textbook.

Tip 2: Make causal statements only when they result from causal research designs

Ask yourself:

- Is the study design causal?
- Are statements causal when they should not be?
- For studies that are not causal, are limitations of noncausality discussed?

Is the study design causal?

Causal relationships are at the heart of social-science research on whether policies, practices, or approaches are effective. An intervention that improves outcomes is “causing” the improvement. But the expression “correlation does not mean causation” is a reminder that outcomes might improve for other reasons. The challenge is to eliminate those other reasons.

Experiments are the tools that researchers use to eliminate those other reasons. Experiments can take different forms, and Box B shows some examples. One form uses a “randomization device” to split a population into two groups. (A random number generator in spreadsheet software works well.) The offer of the intervention—the policy, practice, or approach—is the only variable that differs between the two groups. By construction, the offer of the intervention is then the only explanation for why outcomes differ, after

allowing for chance differences that arise from the randomization process itself. The offer of the intervention causes outcomes to differ (the “effect”).¹

Box B. Types of experiments

Experiments are important for measuring effects because they yield *causal* estimates of those effects. The following examples summarize some experimental approaches:

Random assignment: A randomization method by which study participants are assigned to treatment or control groups.

Regression discontinuity: A method by which study participants are assigned to the treatment group on the basis of a cutoff value of a variable, such as a test score. For example, students scoring below the twentieth percentile on a reading test may be assigned to participate in a supplemental reading program.

Single subject (also called single case): A range of designs in which a treatment is given to study participants and then taken away from them their outcomes with and without the treatment are compared.

Natural experiments: A type of experiment in which a law, policy, or regulation creates a treatment group and control group. For example, a voucher or charter-school lottery creates treatment and control groups (those that are selected and those that are not).

Other forms of experiments can be used, depending on circumstances. For example, an intervention that is given to an individual, removed, given again, and removed again (a type of “single subject design”) can yield an estimate of the intervention’s effect that is causal (assuming that intervention effects “stop” after the intervention is removed). Also, an intervention that is offered only to those with a “score” above (or below) a threshold value of a characteristic, such as a test score (known as a “regression discontinuity design”), can yield an estimate of an effect that is causal.

A common and similar-sounding approach for studying effectiveness is a “quasi-experiment.” An example of a quasi-experiment is a study that compares the reading skills of students in a group of schools that volunteer to use a new reading curriculum to the reading skills of students in another group of schools that use the existing reading curriculum. Because schools volunteer, intervening variables can affect outcomes, even if schools are outwardly similar on characteristics that the study observes (“observables”). But the two groups of schools may not be similar on characteristics that the study does not observe (“unobservables”). The groups

¹ The focus here is on the *offer* of the intervention. Random assignment creates two equivalent groups, and an experiment’s ability to measure causal effects derives from the equivalence of those groups. Formally these are ‘intent to treat’ effects, which should be distinguished from effects of *using* the intervention.

may be managed differently, have different levels of teacher experience, or have different kinds of after-school programs that students attend. Without an experiment, a researcher cannot rule out the possibility that these unobservable characteristics are affecting outcomes.

Are statements causal when they should not be?

The “quasi” qualifier is important for understanding the different abilities of experiments and quasi-experiments to measure causal effects. One way to think about how they differ is to imagine all characteristics that can affect an outcome inside a rectangle (see Figure 1). Many characteristics might be there. The rectangle can be divided into characteristics that a study observes and characteristics that it does not observe. For example, if the outcome is reading skill, a study might measure students’ reading in their homes, the training and experience of their classroom teachers, and reading skills of their classroom peers.² The same study might not measure student motivation to read. What is measured and not measured depends on how ambitious the study is about collecting data and what the extent of resources is, but it is safe to say that researchers cannot know whether they have measured *all* relevant characteristics.

In an experiment, the rectangle is the same for the treatment and control groups. Researchers refer to groups being “balanced” on observed and unobservable characteristics. In a quasi-experiment, however, only the rectangle that represents observed characteristics is the same for treatment and control groups. Whether the groups are balanced on unobservable characteristics is unknown and, being unobservable, cannot be known. When a quasi-experimental study observes differences between outcomes in the treatment group and outcomes in the control group, results inevitably mingle the effect of the intervention on outcomes and the effect of unobservable characteristics on outcomes.

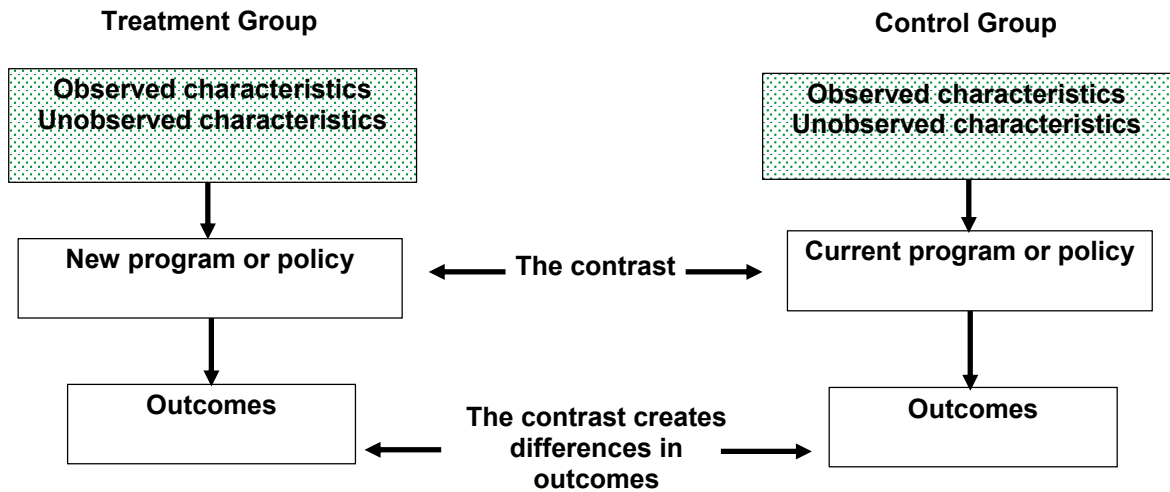
These differences in how experiments and quasi-experiments adjust for observed and unobservable characteristics should be kept in mind when writing about findings. Box C shows examples of findings being accurately and inaccurately described.

² To depict how interventions are intended to affect outcomes, some studies use schematic “logic models” that relate inputs and structure to outcomes. For more information about logic models, see <http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=404> and <http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=401>. A typical logic model focuses on the intervention and not its contrast with services received by the control group.

Figure 1. Differences between experiments and quasi-experiments

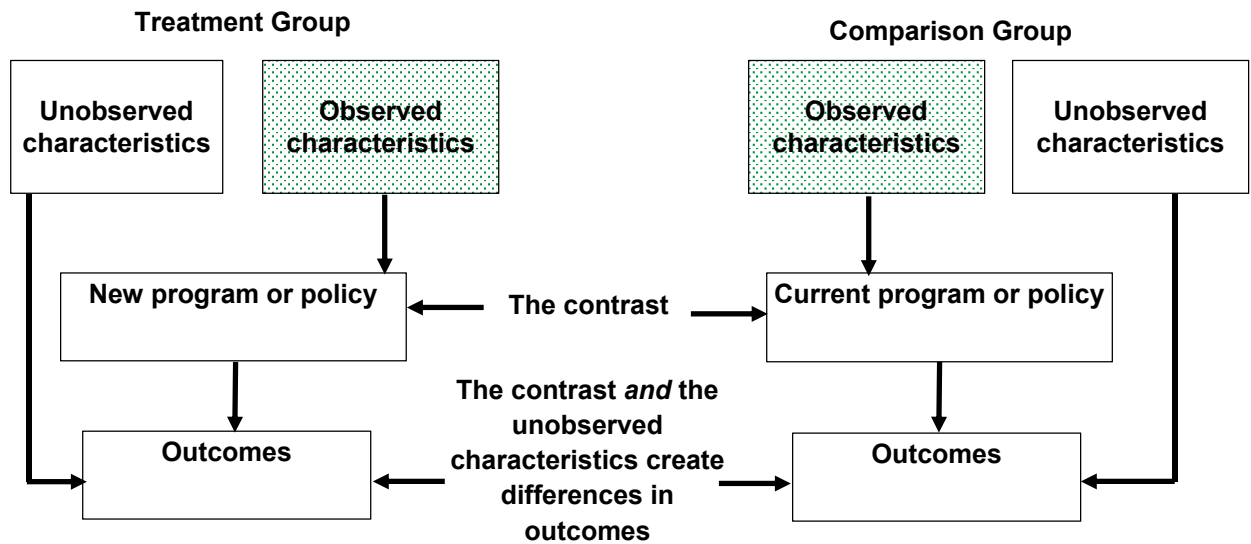
An experiment


Treatment and control groups are balanced on both observed and unobserved characteristics



A quasi-experiment

Treatment and control groups are balanced only on observed characteristics



 Balanced characteristics

Box C. Accurate and inaccurate causal statements

An **accurate causal** conclusion, from an experiment in which college students were randomly assigned to have coaches:

Coaches increased student persistence in college by 5 percentage points.

An **accurate noncausal** conclusion, from a quasi-experiment in which students volunteered to have coaches and were matched to other students who did not volunteer:

Students with coaches had higher rates of college persistence than similar students that did not.

An **inaccurate noncausal** conclusion stated as a causal conclusion, from a quasi-experiment in which students volunteered to have coaches and were matched to other students who did not volunteer:

Coaches increased college persistence.

The first passage is from an experiment that studied the effect of coaching on keeping college students in school. The statement that “coaches increased student persistence in college by 5 percentage points” is causal in nature. Because of the study’s design, the findings showed that coaching causes persistence. The second example is the same intervention but was studied by using a quasi-experiment. The accurate statement is that students with coaches had higher rates of persistence. The third example is the same quasi-experiment, with its conclusion stated inaccurately as a causal one: coaches increased persistence. In a quasi-experiment, unobservable characteristics could affect why some students persisted and others did not. Differences in persistence could arise because of those unobservable characteristics. A quasi-experiment is not able to make causal claims simply by stating its findings as if they were from an experiment.

For studies that are not causal, are limitations of noncausality discussed?

Readers might easily overlook differences between experiments and quasi-experiments. The difference between an intervention *causing* an effect and an intervention being *associated with* an increase in outcomes may seem like an overly fine distinction that only researchers need worry about. But this is a distinction with a difference. Causal evidence provides a strong basis for policy. A noncausal statement possibly could be a basis for policy, but policymakers should be cautious with findings from these designs. On the basis of the third example, a policymaker might think that coaches benefit students, and, in turn, provide more funding for coaches. In fact, the quasi-experiment found coaches *may* have benefited students. Because of the study’s design, it cannot rule out other plausible explanations. Students that volunteered for coaches may have had unobserved characteristics, such as being more motivated to succeed in college.

Findings from experiments have stronger causal foundations than findings from quasi-experiments, but the stronger foundations do not guarantee that implementing what was studied will yield the same effects. As the first “Going Public” brief explains, even if results indicate that an intervention has effects, researchers writing

about the results of a study should not oversell their finding that “this intervention works.” The finding should be tempered to say that it worked for the specific study population and implementation experience. Researchers should discuss these limitations plainly when writing reports. Reports generally also should discuss other limitations that might arise from the sample, the data-collection process, analysis issues, or other aspects that help readers understand the findings.

Tip 3: Present numbers simply and concretely

Ask yourself:

- Are numbers compounded or turned into fractions?
 - Are number concepts used differently in the same sentence?
 - Can graphs or figures make the numbers more visual?
-

When writers include compounded numbers (meaning numbers that result from arithmetic operations) and fractions in their reports, readers are required to “do the math.” Readers also have to use more “working memory” when three or four numbers in a row are presented. Psychologists have long studied “working memory” and the role of working memory in math cognition. Working memory is not large—currently it’s thought to be about three to four items. Numbers use working memory too, so three or four numbers in a row will slow readers down. Working with numbers can even induce physical stress.³

³ Daniel Kahneman (2011) describes psychological experiments in which researchers asked participants to add the value of one to each place in a three-digit number—that is, to respond to a number like “436” with the answer of “547.” As they did, their heartbeat increased, their respiration quickened, and their pupils dilated—all consistent with physical stress. Asking participants to add three to each place rather than one heightened the stress. It was harder to do.

Writers have close familiarity with their data and analyses, so it's easy to see how they end up writing the results of their efforts with number-packed text. Writers already have worked through their numbers. The numbers make sense and tell a story. But putting numbers close together and manipulating them as fractions, ratios, or percentages creates dense text that quickly becomes hard for readers to understand. Box D presents an example of many numbers and concepts presented in a way that is hard for readers to understand. When writing about research findings, present numbers in a way that minimizes the load on working memory.

Box D. Numbers presented to facilitate readers' understanding

An example of a sentence that taxes working memory:

The treatment group mean was 68 percent, and the control group mean was 36 percent, which implies an effect of 32 percentage points and a relative effect of 89 percent.

The sentence includes four numbers, two different kinds of fractions, and a difference between numbers:

- The first kind of fraction is the treatment group mean of 68 percent and the control group mean of 36 percent; they may be proportions of students graduating from high school or achieving reading proficiency.
- The third number is the effect, the difference of 32 percentage points between those two proportions.
- The second kind of fraction is the effect divided by the control group mean (32 percentage points divided by 36 percentage points).

Using more text will help readers absorb the numbers more easily:

The data indicate that 68 percent of the treatment group and 36 percent of the control group achieved proficiency. The difference of 32 percentage points suggests a large effect. The treatment group mean was 89 percent larger than the control group mean.

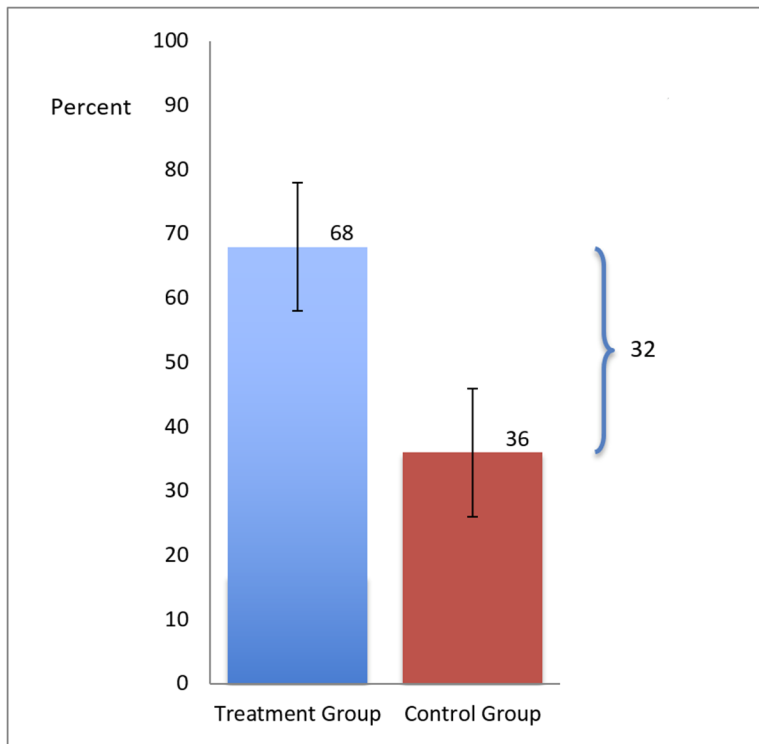
Another approach for presenting numbers is to use graphics or visuals. Tukey (1990) notes three properties that authors should strive for in presenting findings as visuals: the finding should hit the reader fast, hit the reader between the eyes, and be unavoidable.⁴ Figure 2 shows the example numbers from Box D as a graphic. A reader will quickly see that the average value for the treatment group is about twice the value for the control group. The bracket on the right provides an explicit value of 32 for the difference. Writers may have different views about whether to present the confidence interval for the mean values (the lines in the columns). Adding confidence intervals is useful—if intervals cross, statistical tests will likely find that the difference between the means is not significant—but the tradeoff is that every line and bar adds information that readers need to

⁴ There is an enormous literature on presenting numbers visually, anchored by Tufte's (2001) classic volume.

process. Finding the balance at which a graphic presents just the “right” amount of information depends on the audience.

Graphics should be used with care. Presenting many numbers in one graphic may add too much complexity, essentially turning text with too many numbers into a graph with too many numbers. And presenting one graph after another can be numbing. Linking a study’s main findings to graphics is a way to determine which graphics to add and which are extraneous. Most studies have only a few main findings, and only a few graphics may be needed to present them.

Figure 2. A simple graphic displaying numbers



Tip 4: Describe effects in meaningful units

Ask yourself:

- Are effect sizes described in meaningful units?
- Are effect sizes compared to expected rates of growth, closing of gaps, findings from other studies, or cost-effective amounts?

Ultimately, effectiveness studies yield numbers: changes in test scores, reductions in behavior incidents, or improvements in attendance rates, for example. Researchers commonly use the “effect size” to convey a sense

of how large effects are. Effect size is commonly defined as the difference between the treatment and control group average outcomes (the effect) divided by the standard deviation of those outcomes in the population. Effect sizes are useful because they are pure quantities. In other words, they are not associated with a unit of measurement (for example, dividing 4 pounds by 2 pounds produces the number 2, which is a pure quantity), so effect sizes for two or more outcomes can be compared, regardless of how different the outcomes are. Effect sizes can be compared for outcomes that measure behavior, reading test scores, number of times students participated in a discussion, or any other outcome. When writing about the results of effectiveness studies, writers can use effect sizes to indicate that an intervention's effect on behavior, for example, was larger than its effect on test scores.

Although effect sizes are useful tools for researchers, they are not as useful for policymakers. For policymakers, effect sizes are unit-less measures that do not convey whether the finding is meaningful, any more than saying "two" conveys meaning. Two what? To avoid this problem when writing for policy readers, consider using one or more of the approaches described by Hill et al. (2008), Bloom et al. (2008) and Harris (2008) (see Box E).

Box E. Four ways to put effect sizes into context

1. Compare them to normative expectations for growth over time in student achievement.

“Normative expectations” place a study’s effect against the backdrop of what is normal. Hill et al. (2008) and Bloom et al. (2008) present tables of average achievement test scores in different grades and show the effect of a year of learning at different grade levels. The tables show that a year of school has an increasingly smaller effect as children age. Writers can use these tables to calibrate how large their effect size is compared to a year of learning for students at the same grade level. For example, Hill et al. (2008) report estimates that fifth grade, on average, had an effect size of 0.40 for reading. A study showing that a reading intervention for fifth graders had an effect size of 0.20 can report that its finding is equal to about half a year of schooling based on data from the Hill et al. study.

2. Compare them to policy-relevant gaps in achievement.

To compare effects to policy-relevant gaps, scores for black and white students on the National Assessment of Educational Progress are often used as benchmarks. For fourth graders in 2013, the average reading score was 232 for white students and 206 for black students. The standard deviation was 37 for all students. Measured as an effect size, the 26-point gap between black and white students becomes an effect size of 0.70. A study reporting that a reading intervention had an effect size of 0.35 can report that its finding is equivalent to closing the black-white achievement gap by half.

3. Compare them to findings from previous research.

This approach uses previous research to create a distribution of measured effects for similar interventions. A study of a reading intervention could report that its measured effect is in, say, the top 10 percent of measured effects for reading interventions. The [What Works Clearinghouse](#) allows users to create their own tables of findings from previous research, which provides a tool for using this approach.

4. Compare them to costs and benefits for the intervention being studied.

This approach asks whether an intervention has benefits that exceed its costs. The approach implicitly uses effect sizes but presents them as money quantities. For example, improved reading ability correlates with higher future income, so an intervention that improves reading ability generates higher future income. That gain in income can be compared with the cost of an intervention. Similarly, high school graduation is known to lead to higher income, so the cost of an intervention to reduce dropping out can be compared with the benefit of higher incomes. This approach can be difficult to implement. Many effectiveness studies do not report intervention costs, and studies may be estimating effects on outcomes such as incidents of inappropriate behavior or measures of social learning that have indistinct relationships with future earnings.

Tip 5: Present findings as information for policy

Ask yourself:

- Are findings presented as information for policymakers?
 - Are findings based on evidence and not on editorializing?
-

Policymakers may want researchers to condense findings of long and complex studies into a few pages or even a bottom line, which creates the challenge to researchers of presenting their findings as actionable information for policy. Tip 4 discussed ways to put findings into quantities that are useful for policy. Tip 5 is about the challenge of presenting overall findings as information for policy.

Harris (2008) notes that two kinds of policy decisions are common. One is a decision among alternatives that are related and “narrow.” For example, to improve reading comprehension, a policymaker might be choosing between interventions to improve reading comprehension. The second is a decision among alternatives that are far apart, or “wide.” A decision might be about whether to spend more on programs to improve reading comprehension or more on advanced-placement courses in high school. At high levels of government, decisions might be as wide as whether to spend more on education or more on health care.

Researchers should, as much as possible, align the study’s contrast with the policy being considered. Most effectiveness studies support “narrow” decisions. A study contrasting two approaches for teaching reading will have findings that help choose between them. That same study does not support a decision between spending on more effective teaching of reading versus spending on health care. A study comparing education and health outcomes for countries that spend different amounts on the two might inform a decision in that wide frame.

Box F shows examples of presenting findings as information for policy. The first study examined how students responded to failing a high-school exit exam. A policymaker can use these findings to consider whether the positives of the high school exit exam outweigh the negatives.

The second example puts the study’s results into an unclear policy context. In the example’s first sentence, results suggest that public schools “appear to be” performing well compared to private and charter schools. The hedge might signal readers that the author is unclear about the study’s conclusions. The example goes on to say that the study’s findings “suggest significant reasons” readers should be “suspicious” about claims that public schools are failing, and it “raises substantial questions” about the premise of school reform. The reader is left to infer what the significant reasons and substantial questions are. The example tells readers that the study could be interpreted as refuting claims about the effectiveness of approaches for reforming public school: what some claim is true might not be true.

Box F. Examples of presenting findings as information for policy

1. A clear example from a single study:

Whatever the underlying mechanism, failing a high school exit examination has long-lasting effects on students. Students who failed were less likely to go on to college than similar students who passed.

2. An unclear example from a single study:

The study's results suggest that public schools appear to be performing well compared to similar private and charter schools. The findings suggest reasons to be suspicious of claims of failure in the public schools and raise substantial questions about a premise of school reform.

3. A clear example from a review of studies:

The findings of available studies indicate that accountability contributed to achievement growth. The positive effects were clearer for mathematics than for reading, and were particularly clear when the outcome measure was based on a national test, such as NAEP.

4. An unclear example from a review of studies:

The analysis revealed substantial discrepancies among studies. Policymakers should become aware of the variability of evidence in the literature, and should be cautioned against relying only on research that is consistent with their own views about accountability.

It would be unsurprising if the second example were met by confusion. The writing places results into a wide frame when a narrow one was appropriate. Had it stayed within its narrow frame, it would have reported something like “the study found that students in public schools, private schools, and charter schools had similar achievement levels after adjusting for characteristics of students attending those schools.”

Syntheses of related studies provide a stronger basis for policy than a single study. This stronger basis underlies efforts by organizations that synthesize research, such as the [Cochrane Collaboration](#), [the Campbell Collaboration](#), and the [What Works Clearinghouse](#). Again, the key to putting a synthesis of findings into context is to consider the contrasts being synthesized. A set of studies implies a set of contrasts. The third example in Box F is from a synthesis of studies looking at relationships between accountability and test scores. Its orientation is entirely empirical: the quantitative evidence points to accountability being related to achievement growth, more so in math, and more so when a national test was used to measure achievement. Presumably a reader can verify each of these claims in the study's tables.

The fourth example is from a synthesis of studies that also examined accountability and test scores, and included many of the same studies included in the third example. The first sentence states that the review found that results varied between studies. If the purpose of the review is to illuminate that results vary, stating this finding makes sense. But the example then moves from an empirical stance (findings vary) to a judgmental stance—policymakers should not use that variability as a way to support their own views.

Findings of effectiveness studies are best presented as evidence rather than as urgings. A study's conclusions or recommendations need to be based on evidence it presents, possibly integrated with findings from other research. When conclusions or recommendations are opinions or views of the writer, they are no longer based on evidence. The example about high school exit exams presents evidence. It does not editorialize about whether exit exams are right or wrong or whether policymakers should support them or not. The second example comparing school types expresses a point of view about what policymakers *should* think. Readers might wonder whether a researcher expressing a point of view also is presenting evidence selectively to support that point of view. Researchers may feel tempted: they know more about the findings than anybody else and perhaps conclude they are entitled to speculate on implications of the findings. This temptation is best avoided. There are settings in which researchers can editorialize or are asked to, but a study's findings will be clearer if they are presented as disinterested evidence and not mixed with editorializing

Putting the parts together

Effectiveness studies are important opportunities to measure effects, and responsibility falls on researchers to convey what was measured and what was found. A study that is clear on the five features of research studies discussed here is more likely to resonate with readers and help policymakers grasp findings accurately. Readers need to know what the question was, understand the contrast and the approach used to answer the question (causal design or not), be able to follow the numbers, be able to gauge effects against some context or benchmarks, and understand how the effects answer fits into a policy context. Crafting the executive summary around these five features will be useful. A study report that provides an unclear contrast, makes statements outside its design, piles up its numbers, and does not put its findings into policy perspective is limiting itself from helping to shape policy. Clear writing is important for conveying study findings in a way that is understandable and useful to policymakers and other readers.

References

- Bloom, H., Hill, C., Black, A., and Lipsey, M. (2008). *Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Education Interventions*. New York, NY: MDRC. Retrieved June 2015, from http://www.mdrc.org/sites/default/files/full_473.pdf.
- Harris, D. (2009). "Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs." *Educational Evaluation and Policy Analysis* 31(1), 3-29.
- Hill, C., Bloom, H., Black, A., and Lipsey, M. (2008). "Empirical benchmarks for interpreting effect sizes in research." *Child Development Perspectives* 2(3), 172-175.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Strauss, Giroux.
- Nelson, S. R., Leffler, J. C., & Hansen, B. A. (2009). "Toward a research agenda for understanding and improving the use of research evidence." Portland, OR: Northwest Regional Educational Laboratory. Retrieved February 12, 2015, from <http://educationnorthwest.org/sites/default/files/toward-a-research-agenda.pdf>.
- Tufte, E. (2001). *The Visual Display of Quantitative Information (2nd edition)*. Cheshire, CT: Graphics Press.
- Tukey, J. (1990). "Data-based graphics: Visual display in the decades to come." *Statistical Science* 8(3), 327-339.

