

IMPACT AND YOUR DEATH BED: PLAYING THE LONG GAME

Dan Goldhaber

Center for Education Data
and Research
University of Washington
Seattle, WA 98103
dgoldhab@uw.edu

WHAT DO WE MEAN BY IMPACT?

I attended my first Association for Education Finance and Policy (AEFP) conference (then American Education Finance Association) in 1995 in Savannah, Georgia. The world of education research has changed a great deal since then, but I suspect what has not changed for most people in the “education research community” (a term meant to include education researchers, policy makers, practitioners, and other wonks) is the underlying reason we are all here. In short, we want to engage in, or with, research that matters, research that helps education institutions improve the academic and life outcomes of students.

At its heart, this essay is about whether education research matters for students.¹ Thus, it’s useful to begin with a definition: for this essay, I will define *death-bed impact* as the direct line between education research and student outcomes. Yes, it’s a bit macabre, but the distinction between the impact that you really care about versus the various ways that research impact is measured (for instance, many academics and researchers conceive of impact as the number of citations or the impact factor of the journal in which they publish) is important for what follows.

So, what are the steps necessary for education research to have a death-bed impact?² To begin, there should be some societal agreement as to the outcomes we value. This is an obvious, but not trivial, matter. We never know for sure what students get out of their schooling, and many of the outcomes that schools affect aren’t observed until much later in a student’s life. The weight we place on different measures of student outcomes, or how schools contribute to the development of students (such as test scores, educational attainment, or assessments of social-emotional learning), differs from person to person. Indeed, some might argue that other schooling contributions, like preparing

1. For a longer version of this essay, which touches on important topics such as Wilt Chamberlain’s free throws, the youth of today, and more on what AEFP members think about research impact, see CEDR Working Paper No. 07242017-1 at <http://cedr.us/publications.html>.
2. For a more comprehensive discussion on research and policy connection, see Tseng (2012).

doi:10.1162/EDFP_a_00246

© 2017 Association for Education Finance and Policy

students to participate in a democratic society or teaching tolerance of others, are equally or more important than test score gains or years of schooling. The varying roles that schools play, combined with different values we place on these roles, make it difficult to agree on how to measure school progress and interpret the import of research findings (Cody 2011; Whitehurst 2016).

Clearly, there must also be lines of communication between researchers and policy makers and practitioners (henceforth P&P). Sometimes this communication will come through intermediary organizations (e.g., AEI, CAP, Brookings) and sometimes directly from researchers.³ And there is little doubt that relationships between policy makers and researchers make a difference. The research itself is more likely to be actionable if it is informed by the needs of P&P, and P&P are likely to turn to trusted sources when obtaining information about the state of the literature on a topic. It is through these relationships that we might expect to see findings influence policy most immediately.

But we also might hope for research to have bigger, longer-run impacts by establishing facts and knowledge with which policy makers must wrestle in the course of the policy-making process. For this to occur, there needs to be some general agreement among researchers evaluating empirical evidence about both what research shows and, relatedly, what it portends for policy or practice (henceforth “policy”). Researchers will of course disagree about the findings from particular studies.⁴ And it’s not crazy (or necessarily uncommon) to look at research, agree on the basic empirical findings, but disagree on what these mean for policy (e.g., Corcoran and Goldhaber 2013). Some level of consensus, however, is likely necessary in order for research to influence policy. In the absence of some degree of basic agreement among researchers on both the substance and implications of findings, it is hard to see how P&P could think there is a research consensus suggesting a particular course of action.

The final step toward death-bed impact is that P&P should have the incentive to make decisions that are aligned with improving student outcomes.⁵ There are, of course, many individual examples where important decisions at both the individual and institutional levels don’t appear to be driven by the weight of abundantly clear empirical evidence, but the basic assumption is that, overall, the political process will lead to good societal outcomes. I used to take that as a given but the politics and policy making of spring 2017 does seem to call this into question (I miss the median voter!). Indeed, if the past year has not led you to question the degree to which empirical evidence (or “facts”) matter for important decisions, well, I question your powers of observation. For

3. There are certainly many researchers who make a good faith attempt to get their work out into the public sphere. But I’ve also been to enough presentations that are ostensibly geared to a broad audience where I find myself staring at slides full to the brim of regression coefficients printed in 8-point font to know that “locked away in the ivory tower” is a phrase that exists for a reason.
4. For a great example that gets into technical weeds, see discussion (and disagreement) about the degree to which falsification tests undermine the validity of value added (Goldhaber and Chaplin 2015; Chetty, Friedman, and Rockoff 2016, 2017; Rothstein 2017).
5. One might (and probably should) also argue for a fifth step associated with the importance of good implementation for positive results. See Durlak (2011) and O’Donnell (2008).

the sake of sanity, I'm going to pretend that the political process, judged over the long haul, is not broken.⁶

In what follows I describe the results of a survey of AEFPP members assessing the degree of their consensus on particular empirical findings, their assessment of how politics and research influence policy, and how the research and dissemination process might change so as to increase the likelihood that empirical evidence will influence decision making. Finally, I'll offer a few thoughts, which are often unburdened by empirical evidence, about how education research and dissemination has changed recently, the role social media has played in these changes, and what I think we face for research to matter in the way we might want it to.

Two things before you begin. First, as you read, please keep in mind the distinction between *impact* and *death-bed impact*, as that distinction is central to the essay. Second, I had a (brilliant) friend read an earlier draft of this and he/she (you can guess from the acknowledgments, but I'll never tell) suggested that I warn readers that I'm "going to say things that some people may find offensive." I'm not precisely sure what those things are, but you've been warned!

WHAT'S THE PROBLEM?

I was once an idealist who believed that research could and would have *impact* if only we did it carefully enough and reported the findings honestly enough, but I've grown increasingly pessimistic that good research is regularly used to make good policy.

The availability of new data, analytic techniques, demand for rigor by the U.S. Department of Education's Institute for Education Sciences, and the general "hotness" of education as a research and policy area, have all contributed to drawing talented scholars into the field. I don't think it's an exaggeration to say there has been a revolution in the quality of educational research over the last ten to fifteen years. But while we might be in a golden age of education research, a nontrivial share of it is of questionable quality. This, combined with seeing how decisions were made when I served on a local school board, and not seeing much evidence that high-quality research generally translates into improved practices and student achievement, contributes to a rising tide in my level of cynicism about the research-policy connection.

Let's begin by looking back for a moment. About a decade ago, I co-authored a paper with Dominic Brewer in which we explored the incentives that drive education research. We argued that there are various failures in the market for education research, which often lead to the distribution and amplification of bad work. I won't fully reiterate what is already spelled out in an excellent article,⁷ but a few issues merit mentioning.

6. This is also of practical import as one could write a tome on any part of the process connecting research and policy: Even with the vast powers of the AEFPP presidency, I am limited by a higher power—a higher power I know as Lisa Jelks (Jelks, personal communication, January 2017)—to "something like a 10–15 page double-spaced manuscript." In the unlikely event that this article is read by people who have not interacted with her, Lisa is the superb managing editor of *Education Finance and Policy*, and she and Amy Schwartz, EFP's editor, were nice enough to give me a bit more space, but not enough to give a full treatment to this topic.

7. I had intended for this section to be largely evidence, and certainly citation free, but come on, I have a chance to up my impact right here—Goldhaber and Brewer (2008)—and I can't resist the irony given what follows in the next sentence.

The academy provides strong clear incentives for focusing on academic-types of measures of impact (i.e., *not* death-bed impact). Thus, a common strategy once the first stage of research is complete is to present it at conferences, submit it to a journal, and hope it gets published—somewhere along the line picked up by more accessible media outlets. There’s nothing wrong with this strategy and the process itself likely provides researchers with valuable feedback, though it is important to note there are many tiers of journals, so the extent to which journal publications signify that research is really vetted varies.

Also, even back in the day (let’s call the early 1990s “the day”) researchers with enough pull could get compelling work published in the *New York Times* or *Wall Street Journal* without it having first received a full journal vetting. Obviously, academia still values journal publications so research today continues to receive peer-reviewed vetting, but social media has changed the publicity game quite a bit. It’s no great insight to note that today we live in a world with a more rapid news cycle, greater ideological fragmentation, and the ability for researchers to *push* research out the door directly by posting working papers on publicly accessible Web sites and/or calling attention to new research findings using Twitter.⁸

There have always been significant publicity benefits for being first with research—reporters are generally going to be more receptive to writing about work that is novel than replication work, especially if the replication shows the same result as prior research. So how is social media related to the pressure to be first? Well, back in the day reporters might scan new journal publications to get story ideas. Thus, a fair amount of peer-reviewed critique would already have occurred. Today, reporters are much more likely to get ideas directly from researchers themselves. Indeed, I have been told by several reporters that tweets from researchers they follow represent the primary means by which they get new education stories.

Much of what gets tweeted (or otherwise pushed out to the media) has been vetted (again, I do not necessarily think publication in a journal constitutes “thorough” vetting, but it’s something), but certainly not all. And one consequence of the incentive to be first is that research is less likely to get the kind of professional critique it might have received prior to seeing the bright lights of the press.⁹ Although there are clearly upsides to being able to more broadly and quickly disseminate work, the fact that research is more likely to get out the door without a thorough critique increases the likelihood that mistaken results lead to misinformation that may inform policy debates.

It has also become common to use social media to gain prominence by connecting with an audience that has a specific ideological perspective. Diane Ravitch’s blog (<https://dianeravitch.net>) is a terrific example of this. I’m not suggesting that the posts

8. Though, as Danial Nexton (@dhnexon) points out in a 24 June 2017 tweet, “The disposition to aggressively self-promote is, shockingly enough, unevenly distributed among political scientists.” I appreciate the irony with which this tweet was pointed out to me!

9. As but one example of what this means, I have seen some egregious cases where underpowered research findings receive coverage in popular news outlets. The problem here is that it is difficult for many in the media to distinguish research that finds a particular program does not work (based on an assumption of what effect size would constitute “working”) from the situation where a researcher did not have a sample size sufficiently large to find an effect that would signify that a program is working. I’m not sure he is the originator, but credit to Cory Koedel for introducing me to the term “uninformed zero” to describe underpowered, yet highly touted, results.

on Diane's blog are wrong, or that there's anything wrong with ideology per se, but I hope we can all agree that it is more difficult to adjust one's views based on new empirical evidence knowing that it may cut against a readership's expectations on an issue. I also hope we can agree that although blog posts and tweets are effective avenues for reaching people, what is said in 140 characters can occasionally lack nuance or the multitude of caveats that appropriately go along with research findings.¹⁰

In case this seems too abstract, I should mention that social media impact clearly matters in ways that are quite tangible to researchers. It is, for instance, now commonplace to measure scholarly contributions in part based on relatively new social media-based metrics. A good example of this is Rick Hess's *Edu-Scholar Public Influence Rankings*. This ranking is based on traditional academic measures of impact (such as Google Scholar citations), but also broader measures of dissemination (such as Web mentions and scholars' Twitter-based Klout scores). The standing of scholars on Hess's rating system are touted by both universities and individual researchers through press releases and on CVs. And these rankings can be influential in determining salaries and promotions.

Lest you think I am adopting a holier than thou attitude, I readily admit to being a participant in the media game—from mentioning media impact in proposals, to actively weighing the benefits of being first with novel work versus waiting longer to get more precise estimates or more feedback, to checking Rick Hess's ratings for my personal ranking. Hence, I am not criticizing researchers or institutions for using social media or their related measures of impact—we are all simply responding to the incentives we face.

Nevertheless, we should recognize that there is probably less room today for the mealy-mouthed (but sometimes more empirically sound) researcher in an ideologically fragmented world that values quick and digestible results over nuanced findings. In particular, most research does not yield definitive yes-or-no answers about what ought to be done in terms of policy. Research offering nuanced findings that could shed light on important avenues for making incremental progress for school system improvement is not likely to get much attention.

Another problem in the market for education research is that consumers of education research—the P&P community being an important constituency—often do not have enough technical knowledge to judge whether education research studies are good or bad.¹¹ The flow of education research has been turned up to 11 on the dial that goes to 10, but it's a mix of good and bad. And regrettably, despite efforts to establish research quality gatekeepers, such as the What Works Clearinghouse, I think what Dom and I wrote back in 2008 stands today: “it appears that the growth in research media outlets is exceeding the capacity of gatekeeping institutions to separate good research from bad. This is deeply problematic, because most consumers of the work will not have the time or capacity to judge its quality” (p. 217).

10. More generally, it is probably hard for researchers to walk back from findings when presented with new evidence that conflicts with these findings, especially if the findings were presented with nuance and caveats.

11. I've certainly met a number of members of the P&P community who do have an excellent grasp of what constitutes good research (many of whom are involved with AEFPI!), but I'd classify these folks as the golden unicorns (the rarest of the unicorn breed).

In sum, I'd argue that the central problem we face is not too little education research—rather, there is too much research of dubious quality making it into the public domain. As a result, there is a good deal of confusion about what research actually does suggest for improving schools, which limits its positive impact. And if there is an inverse relationship between the clarity of the message about what education research suggests and the power of adult interests and institutional inertia to maintain what might be an ineffective status quo, then we should expect school productivity gains to be slow.

I'd go so far as to characterize what I've described above as an existential crisis, which is why I turned to the AEFPP membership to either tell me I am wrong—good research is affecting policy in positive ways—or I'm right, but there's a solution to this big problem.

SO WHAT DO YOU (AEFP SURVEY RESPONDENTS) THINK?

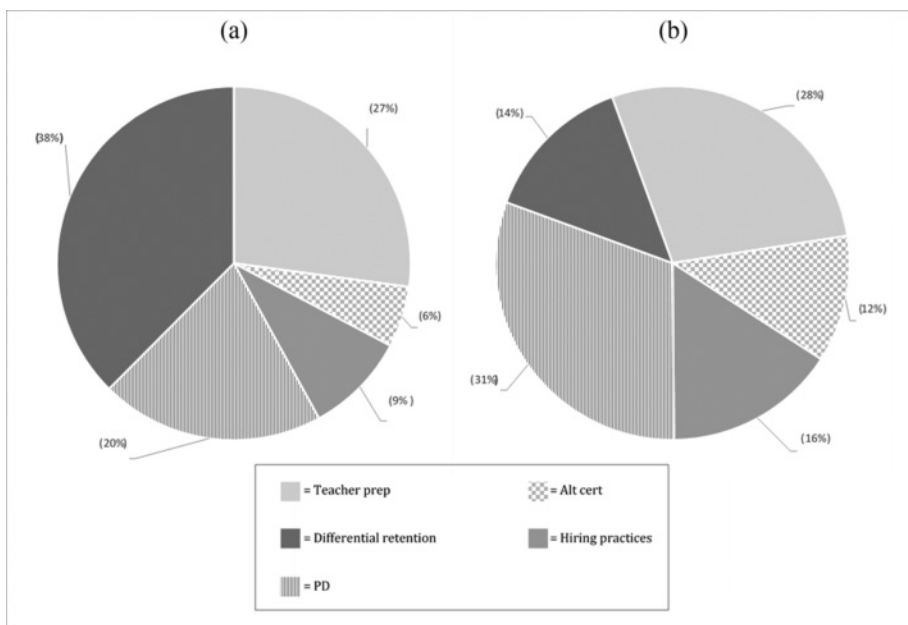
I expressed my concerns about whether research affects policy at the 2017 general session of the AEFPP conference, and asked those in attendance to respond to a simple four-question survey. This survey was designed with three specific goals in mind: (1) to better understand how much agreement there is about research findings; (2) to assess the degree that AEFPP members think we are limited in our ability to improve student outcomes by lack of knowledge or by political hurdles associated with implementing particular policies; and (3) to elicit suggestions about how to make good research matter more for education decision making.¹²

Before getting to the survey results, I suppose I should give you some statistics enabling you to judge how seriously to take the findings presented here. The survey was sent out to AEFPP's 1,045 active members (as of March 2017), and there was a total of 270 respondents to the survey (thank you again to those who took the time to respond!). This suggests a response rate of about 26 percent. Things are a bit fuzzier than this, however, because the survey was clearly directed to those who were at my presidential address during the 2017 conference. There were roughly 800 people at the conference and I would guess (based on seating capacity in the general session room) there were about 500 people in attendance (thanks, Raj!). Thus, I think it's reasonable to say the response rate of those who were in attendance at the address is about 54 percent: not great, but not terrible either.

To gauge the extent to which there was agreement on specific empirical findings, I asked respondents to first prioritize five different strategies for improving “the overall quality of the teacher workforce” based “solely on your understanding of the *empirical evidence* in support of an action.”¹³ Following that I asked for a ranking of the same five strategies based on “your understanding of *all relevant factors* (i.e., not just empirical evidence but also considering what is politically feasible).”

12. You can view the survey instrument in a separate online appendix that can be accessed on *Education Finance and Policy's* Web site at www.mitpressjournals.org/doi/suppl/10.1162/EDFP_a_00246.

13. I chose to focus on the quality of the teacher workforce both because it dovetails with what I believe to be a research consensus that teacher quality is the key schooling variable influencing student outcomes, and because it looks like we are facing an end—premature to my mind—to the national focus on teacher quality (at least at the federal level). I thank Jim Wyckoff for mocking up a similar set of categories for a discussion that occurred at the 2017 CALDER conference.



Notes: Teacher prep = improving the preparation of teacher candidates; alt cert = giving school systems greater discretion over teacher hiring; hiring practices = improving the practices for hiring teachers; PD = improving the performance of in-service teachers; differential retention = differential retention of existing teachers based on effectiveness.

Figure 1. Top Priority for Improving the Overall Quality of the Teacher Workforce: (a) Based Solely on Empirical Evidence and (b) Based on All Relevant Factors.

The five strategies that I asked survey recipients to prioritize were: improving the preparation of teacher candidates (“teacher prep”); giving school systems greater discretion over teacher hiring—for example, relaxing of teacher licensure requirements (“alt cert”); improving the practices for hiring teachers (“hiring practices”); improving the performance of in-service teachers—for example, through professional development and mentorship programs (“PD”); and differential retention of existing teachers based on effectiveness (“differential retention”).

So what did you, AEFP members, think? Figure 1 shows the proportion of all survey respondents who picked each of the five strategies as their top choice *based on the empirical evidence alone* (figure 1a) and *based on all relevant factors* (figure 1b).

When asked to focus on empirical evidence, the plurality of respondents (nearly 40 percent), chose differential retention as their top choice, followed by teacher prep (27 percent), and PD (20 percent). The final two categories, alt cert and hiring practices, together garnered only 15 percent.¹⁴ I would not say the distribution represents an overwhelming consensus about what research suggests for improving the quality of the teacher workforce, but it was more consensus than I expected. It is also interesting that differential retention was most often selected as the top strategy, whereas PD was one of the least favored (26 percent reported it was their lowest priority; see Appendix table A.1). PD is a ubiquitous strategy used by school systems (Miles et al. 2004), although

14. Reported percentages do not necessarily add to 100 due to rounding. See the online appendix for more details on prioritization based on the *empirical evidence* and *all relevant factors*.

Table 1. Transition Matrix (Question 1 to Question 2) for Top Priority for Improving the Overall Quality of the Teacher Workforce

		Q2: Based on All Relevant Factors (%)					Total (row)
		Teacher prep	Alt cert	Hiring practices	PD	Differential retention	
Q1: Based Solely on Empirical Evidence	Teacher prep	16.1	1.9	2.2	4.9	1.9	27
	Alt cert	0.4	2.2	0.0	0.4	2.2	5
	Hiring practices	1.1	1.1	4.9	2.2	0.0	9
	PD	5.2	2.2	0.7	12.4	0.0	20
	Differential retention	5.2	4.1	7.9	10.9	9.7	38
(%)	Total (column)	28	12	16	31	14	

Notes: To be considered in this table a respondent must have made priority rankings for both questions 1 and 2. Percentages in cells are based on $N = 267$. Small differences between figure 1 and table 1 values for alt cert due to rounding. Teacher prep = improving the preparation of teacher candidates; alt cert = giving school systems greater discretion over teacher hiring; hiring practices = improving the practices for hiring teachers; PD = improving the performance of in-service teachers; differential retention = differential retention of existing teachers based on effectiveness.

there is little evidence that school systems are actively dismissing large shares of the teacher workforce based on measures of effectiveness.¹⁵

There were dramatic shifts from the distribution when respondents were asked to rank the same five strategies based on *all relevant factors* (figure 1b). For example, when considering all factors PD (31 percent) and teacher prep (28 percent) are the top two choices, and differential retention (14 percent) ranks fourth behind these and hiring practices (16 percent), and just in front of alt cert (12 percent).¹⁶

Table 1, a transition matrix, shows the first-choice category selected considering all relevant factors (in the columns) based on respondents first choice category for the empirical evidence alone (in the rows), and vice versa. For instance, 16.1 percent of the total sample chose teacher prep based both on the empirical evidence and considering all relevant factors, whereas 1.9 percent of the sample chose *teacher prep* based on the empirical evidence but shifted to *alt cert* when considering all relevant factors.

Respondents tend to stick with their first choice (i.e., the largest share of the distribution tends to be on the diagonal, where respondents' first and second choices are the same).¹⁷ But there are also cases with quite large shifts in responses (from empirical evidence alone to considering all factors); differential retention (–24 percentage points) and PD (+11 percentage points) are good examples. There are many ways to interpret the differences in the responses to questions 1 and 2. For example, some respondents may feel that a strategy offers great promise despite the fact that the evidence on which

15. This may not be the general perception as the “war against teachers” rhetoric suggests otherwise (see, e.g., Dayen 2015; Gamson 2015). But data from the most recent (2011–12) *Schools and Staffing Survey* show the average percentage of teachers dismissed or nonrenewed for any reason was about 2 percent, and the figure dismissed for poor performance was about a half of a percent (see https://nces.ed.gov/surveys/sass/tables/sass112_2013311_dis_008.asp). The District of Columbia Public Schools has one of the most “aggressive” *differential retention* initiatives (and well-known, too, given the 2008 cover of *Time Magazine* with Chancellor Michelle Rhee holding a broom), but even here fewer than 4 percent of teachers each year have been dismissed for poor performance in recent years (Dee and Wyckoff 2015).

16. Here, too, there were no statistically significant differences in responses by responder type.

17. If respondents never deviated in their first choice based on whether they considered only the empirical evidence or all relevant factors, then all the off-diagonal cells would have a value of zero.

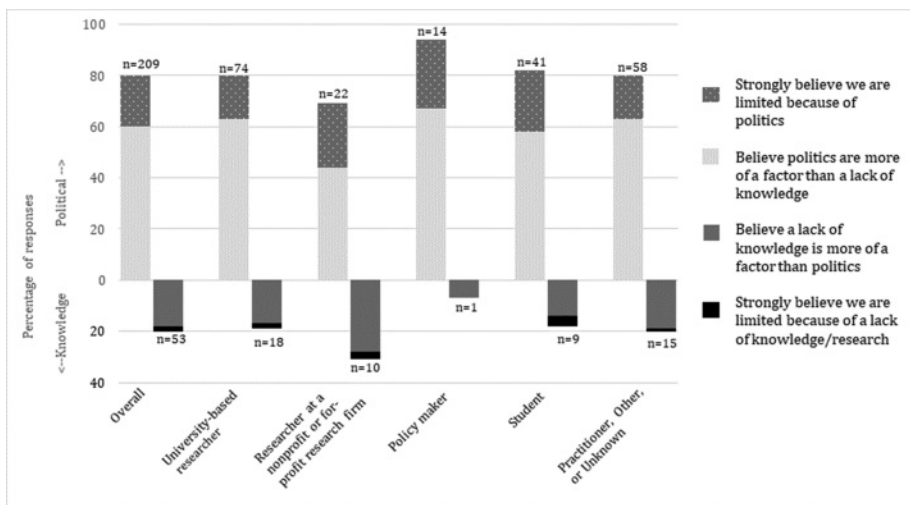


Figure 2. Percentages Rating Politics or Lack of Knowledge as More Important Factor Limiting Our Ability to Improve Student Outcomes.

the strategy is based is rather thin. It is also likely to reflect the obvious, however—politics plays an important role in the ability to implement schooling reforms, often with what is good for adult interests taking precedence over what is best for children.¹⁸

To dig deeper into this issue, I next asked respondents to rate the degree to which “we are limited in our ability to improve student outcomes more because there is a lack of sufficient evidence about what to do . . . OR because policy makers fail to act on existing knowledge due to political realities.” The responses to this question are shown in figure 2. Large majorities of each responder type believe that politics are the more limiting factor in making progress in improving student outcomes: Over 80 percent of the sample stated that politics is the more important factor (and there were no statistically significant differences by responder type).¹⁹

I find the above results depressing. The views of the respondents are largely in line with my own view that the politics of education reform and policy making is a significant hurdle to making schools better. Luckily, I do not need to end the paper on this somber note as I asked the AEFPP survey recipients for suggestions for how to make progress.

THOUGHTFUL SUGGESTIONS FOR HOW TO MAKE PROGRESS

The final question survey recipients received depended on whether they previously picked that we are limited based on a “lack of sufficient evidence about what to do . . . OR because policy makers fail to act on existing knowledge due to political realities.” For those who believed (or strongly believed) that lack of knowledge was more

18. So obvious, in fact, that I choose not to offer particular author citations, rather I cite Reality (Any Year).

19. This leads me to ask: What’s up with young people today? I would have expected students going into education research would be more inclined to think that the lack of knowledge about how to improve is the limiting factor. And, somewhat surprisingly, the group with the highest percentage choosing “lack of knowledge” (either “strongly” or “more of a factor”) are researchers at nonprofit or for-profit research firms (though the responses are also not significantly different from the other respondent types). Make of this what you will.

of a factor I asked for “some concrete thoughts on the types of research questions that need to be answered.” For those who believed (or strongly believed) that politics were more of a factor than lack of knowledge in limiting our ability to improve student outcomes, I asked for “some concrete thoughts about how to make research matter more.” As I would expect from members of AEFPP, there was a large number of very thoughtful comments.^{20,21}

Ideas for New Research

I’m going to begin with a brief recitation of some of the ideas for research from those who fell on the lack of knowledge side of the spectrum—brief because it constituted only about 20 percent of respondents.²²

Nearly all the suggestions for research fell into three broad areas: (1) teacher preparation and development, (2) curriculum, and (3) out-of-school factors. For instance, on teachers, one respondent noted:

we have made a lot of progress in identifying which teachers are high quality but much less progress in understanding what factors and/or interventions help teachers improve their teaching quality. Without being able to improve low-quality teachers or get more individuals who would be high quality teachers to enter the profession, any policies that target hiring or firing teachers will do more to reshuffle the high-quality teachers across schools than improve teacher quality overall.

I happen to agree with this comment and think it nicely illustrates a dilemma we face. Research hasn’t made much progress with identifying the interventions that help teachers improve their teaching quality, but it’s not for lack of trying. There are nearly as many studies on the effects (or non-effects) of professional development than all other K–12 research topics combined. And more research funding goes to study this topic than any other. Okay, truth be told, I can’t empirically justify the claims in either of the prior two sentences, but it wouldn’t surprise me if they were true. So why do we continue to invest in research in this area? I think it’s because improving the skill set of the teachers already in the labor market is the most politically expedient means of trying to move the needle on teacher quality. It is also true that finding ways to change the quality of incumbent teachers is key if we want to move the needle quickly. It would, for instance, likely take twenty to thirty years to see large changes in the workforce via differential retention or initiatives designed to improve the teacher candidates who are newly hired.

It is clearly difficult to implement policies that have significant consequences for a well-organized group of adults. The truth about the “war against teachers” does not comport with the rhetoric, at least in terms of teachers losing their jobs for performance reasons (look back at footnote 15), but that does not mean teachers do not feel like

20. Sadly, I do not have the space to include many of them in this essay.

21. Quotes from the survey responses are italicized to distinguish them from other quotes.

22. Note, respondents who are not identified by name requested anonymity, and each quote not attributed to a particular person comes from a separate person. Also, in some cases I made minor grammatical or spelling corrections to the passages in quotes.

they are under assault. And this feeling clearly affects teacher-based policies. Similarly, redistributing teacher talent so that disadvantaged students have better access to high-quality teachers is also politically difficult. Having served on a school board, I know full well the difficulty of moving an effective teacher from a school with politically active (often more affluent) parents into a disadvantaged school with parents who are less likely to turn out for the next school board election.²³

Although we grapple with the politically delicate policies around teacher effectiveness, there are other, less controversial areas of research that may be overlooked. Mark Steinmeyer, for instance, notes that:

[t]here is good evidence that curricula choice could matter a good deal for student achievement . . . since choosing one curriculum over another is (relatively!) easier politically than altering school governance, choice options, teacher hiring and retention rules or increasing financing, it might be curricula is the low-hanging fruit of school reform

This recognizes an area where research might be expected to have a larger impact, since curriculum choice does not threaten adult interests to the same degree that high-stakes teacher policies do. Indeed, it is quite surprising that we don't know more on this front: School systems make textbook purchasing decisions all the time; even small differences between textbooks A and B could be quite important given that a textbook will be chosen (Kane 2016). While we are figuring out how to tackle political problems (see below), we should be looking for this sort of low-hanging fruit.

Addressing the Political Problem

There were varying degrees of cynicism among those who felt that politics represented the more significant hurdle (again, this is about 80 percent of the respondents). For instance, one respondent suggests

that the vast majority of politicians have no interest in investigating what research is currently out there. While some local school administrators and practitioners may be interested in the latest research and what tools can be effective in improving learning, in general I believe that our elected officials do not share this interest.

But there were also more optimistic assessments. Marty West, for instance emphasizes the long run, noting:

Politics will always play a larger role than empirical evidence in shaping policy outcomes at any given point in time. So, in the short run, politics dominate. In the long run, however, new ideas generated through research and the empirical evidence amassed to support them do play a role in shaping public opinion (or the views of key stakeholders in an issue area) and therefore politics. The question then becomes whether there are strategies to speed the process through which ideas and evidence gain

23. Nevertheless, I do agree with Rebecca Wolf (in her survey response) that it would be good to know more about how “much [do] teachers have to be paid to change the overall distribution of teacher quality?”

acceptance by the public. I'm actually not sure how much we can do here, but part of the strategy has to be paying attention to and investing resources in the dissemination of research through outlets other than academic journals and conferences.

The “other than academic journals” theme was reflected in a large number of comments. Some of the comments get to the specifics of the timing of research or how it is framed. Cara Jackson, for example, wrote, “Research will matter more if presented within a relevant policy window.” Another respondent noted: “In terms of how to share/present research: Doing nothing is often costly. Make that more evident in findings.”

Several respondents noted that policy makers are more likely to pay attention to findings that are contextually relevant (which, in some cases, might mean based on the schools and students over which they are making policy) to them. Seth Hunter’s comment reflects this:

Before I became an academic I worked in a state education agency. My experience tells me this is a problem of presentation and that effective presentation of research is highly context-specific. The presentation of research should ultimately aim to personally convince the policy maker that research implications align with their political values/beliefs without compromising the integrity of the research. . . In a nutshell—all policy making is local.

This is consistent with Rachel Feldman’s view that data and empirical evidence are sometimes less compelling than a good story:

this means telling stories. Politicians may say they want the data, but when decisions are made, they end up falling back on their own experiences—unless we can replace it with a more compelling story. Rather than delivering the data, we need to deliver substantive narratives for the data.

Tracy Weinstein notes the importance of just showing up:

Based on my experience working with legislators in states across the country I think it is imperative that researchers show up more. There are so many folks who are not researchers communicating the work of the research community to legislatures and doing so in a way that often over sells what the literature says or fails to represent the full body of knowledge. Researchers need to be more present, more visible, and more committed to getting their research into the hands of staffers and legislators. If you aren't part of the conversation, someone else is filling that void. I realize this is a complicated issue and there is often fear that engaging at all is somehow going to get you tagged with one side of a debate or the other, but it's absolutely possible to be visibly discussing your work and remain non-ideological and true to the science.

Most comments suggest a need for more research briefs (or shorter pieces). Andrew Biggs, for instance, writes, “Accessibility of research to non-academics has to be a

priority . . . Academics shouldn't sacrifice rigor, but make an effort to generate versions of their research that laymen can understand." I agree with Andrew, but it is also easy to agree with the "more accessible—no sacrifice" position. Sometimes that can be done—some research designs and findings are clear enough that a policy brief, or even a one-pager, can omit without much lost the multitude of caveats that often accompany an academic journal article. Unfortunately, that's not always the case, as findings are often messy and context-specific, meaning there are tradeoffs when it comes to condensing work into shorter, more accessible products.

Some argue that we err too much on the side of caution in terms of advocacy and taking public stances. For example, one respondent pointedly urges greater courage: "I think that highly respected, high-profile researchers need to be willing to risk some of their 'academic credibility' to take stronger political stands on issues where the findings from the empirical literature diverge strongly from what is done in practice." I disagree to some extent with this comment; there are lots of examples of high-profile researchers (though arguably not enough or always the right ones) who have waded into the policy arena.²⁴ More importantly, I'm not sure this solves the problem because there is significant disagreement among these researchers on any number of key issues. Lori Taylor's beliefs reflect what I said above about the publicity benefits of being first with research and the difficulty of replication studies garnering attention:

Researchers get published for disagreeing with one another. Confirmatory work is not publishable in good journals. It gives naïve researchers trolling Google Scholar the impression that we lack consensus, even when we mostly agree. Why should the politicians listen when most of what they hear is noise?

The issue that Lori highlights helps create a situation where, as she also notes:

[research] can be cherry-picked to suit almost any political purpose, making it seem like evidence-based policy making is occurring when in reality, the importance of high-quality research is being diluted.

All of this is consistent with the idea that part of the problem we face is too much research, much of it bad and overly ideological. If this is a large part of the problem, then more policy briefs won't necessarily lead us down the road of making more empirically oriented decisions. Nevertheless, an intriguing suggestion that I think would lend credibility to research is made by Leanna Stiefel, who suggests that "foundations give grants jointly to conservative and liberal think tanks or researchers, insisting that all papers be joint authored."²⁵

A number of respondents argued for new or expanded roles for the P&P community in research questions and design. The basic idea is that it's important to "[f]ind a way to give teachers and administrators ownership over these policy initiatives and how they are implemented." Researcher–practitioner partnerships (RPPs) were called out explicitly as "a direct way to make research more relevant and useful to policy makers by highlighting

24. See, for instance, the numerous adequacy court cases or the *Vergara v. California* (2014) trial.

25. But hey Leanna, what about those of us who are middle-of-the-roaders, we need grants too!

questions already under consideration as opposed to ones with no political support or interest in the moment.”

RPPs are indeed very much in vogue these days. The Institute for Education Sciences (IES) (https://ies.ed.gov/funding/ncer_progs.asp), as well as private foundations such as the Laura and John Arnold Foundation (Laura and John Arnold Foundation, *Policy Lab Background*, 28 April 2017, e-mail correspondence), the Spencer Foundation (<http://www.spencer.org/research-practice-partnership-program>), and the William T Grant Foundation (<http://rpp.wtgrantfoundation.org/>) have all invested in creating or facilitating RPPs. There are also a number of examples of longstanding partnerships between various researchers and school systems that have produced a large amount of policy research. I can say firsthand that the work I’ve done with Spokane Public Schools is some of the most rewarding work in which I’ve engaged, and not just because of the research it has produced. Indeed, more important to me than the published research has been the interactions I (and some top-notch colleagues) have had with practitioners who have the ability to affect the lives of children directly.²⁶ My guess is that the little things that come out of discussions with folks from Spokane (and will never receive any academic attention) are far more important to improving the school system than any of the published work. I’ve learned a great deal from these interactions, and it helps me believe I’m having a death-bed impact.

But, although I’m a fan of RPPs, I’m not sure about either their scalability or sustainability. Researchers and districts may need each other, but there are various hurdles that make the formation of partnerships challenging (Turley and Stevens 2015)—and it remains to be seen if some of the partnerships, which now receive external funding, are sustainable in the face of possible shifts in the priorities of funders.

Another issue is that partnerships privilege established researchers who have not only had time to develop relationships with school systems, but have established scholarly records that increase the likelihood they can secure funding for the partnership. A closely related issue pertains to *where* these partnerships happen. They tend to happen in large urban districts that are in geographic proximity to research universities, given that these are the districts likely to be in the public eye and to have the institutional capacity to establish partnerships. This leaves the vast majority of districts and schools without benefits that come from a partnership, and raises concerns that the research findings arising from such districts may not generalize to, for instance, smaller rural or suburban districts. Thus, one thought on the scalability front is that younger scholars eager to work on policy problems and data access seek out school systems that need research help. This might not be the more glamorous, big-city districts, and it might not come with funding. But it could still be worthwhile, not just to do good research, but also (yes, it sounds trite) to do good.

A final recommendation that showed up several times in the survey responses is to “[g]et more trained researchers (MPP, EdD, PhD) into government and politics.” A disposition toward empirical evidence plus the ability to distinguish high-quality research from bad is a great mix in a policy maker. I’ve been there myself in a different lifetime

26. I’m sneaking another cite in: If you want to know more about this work with Spokane, see Goldhaber, Grout, and Huntington-Klein (2017); you’ll laugh, you’ll cry, you will experience the panorama of emotions and emerge out the other side better for it.

(my school board days) so I know it can be a frustrating experience. That said, I do hope that more people with a strong inclination toward empirical evidence throw their hats in the political ring.

I'll close this section on a positive note cast by Eric Parsons. Eric urges us to have faith: *“Overall, I think the biggest hope is to play the extremely long game, where the bulk of good evidence slowly, slowly, slowly comes to be considered the common knowledge to the extent that no one even considers it reasonable to push back against it.”* I hope Eric is right!

CONCLUSION: (MOST OF) YOU TOO SHOULD HAVE AN EXISTENTIAL CRISIS

What might you make of what's written above? Well, on the one hand it could largely reflect the (early onset) “does what I do even matter?” form of a midlife crisis. But, I think it's difficult to ignore the fact that most of you, AEFPP respondents, also feel that it is politics, not a lack of knowledge, that is limiting our ability to improve education outcomes. It would be naïve to think that politics would not factor into decisions. We live in a democracy, so politics obviously does and should matter in policy making. But I'm worried about the longer run. Is the research community effectively establishing and communicating the knowledge and facts with which the P&P community should wrestle when making decisions? Does the way that research is translated into the public domain mean that adult interests too often trump what is in the best interests of children?

What can be done if the answer to the above questions is “no”? A good first step is always to call for more research! In particular, I'd say we need research to gain an understanding of the conditions under which research is most likely to affect policy.²⁷ We know some (e.g., Tseng 2012), but far from enough, about the links between educational research and decision making. Fortunately, there are new efforts to better understand this process (see, for instance, the National Center for Research in Policy and Practice and the Center for Research Use in Education, the two IES-funded knowledge utilization centers).

Unfortunately, the problem of making research matter more is complicated by the fact that there is hardly a consensus among researchers about what the evidence means for policy making (at least on the question of increasing the quality of the teacher workforce). Thus, I'd say we also need more investment in consensus building. Maybe that comes in the form of research quality gatekeepers, like the What Works Clearinghouse, but I'm also intrigued with the idea of encouraging (e.g., look back at Leanna's funding idea) those *seen* as being on different sides of ideological debates (e.g., on the merits of class size reduction, school accountability policies, charter schools) to design and engage in joint work on hot-button issues.

There is also clearly interest in bringing policy makers and researchers closer together. As I described in the last section, there are efforts aimed at this, such as funding for the building of policy labs or the funding of researcher-practitioner partnerships.

27. There are some school systems that have made striking progress in building evidence-based cultures (I would, for instance, put the District of Columbia Public Schools in this category) that have translated into student achievement, but it's not clear why this occurs in some places and not others.

(AEFP is also trying to do its part to foster connections between researchers and the P&P community through conference sessions designed around conversations between the two communities that hopefully lead to future collaborations.) Ultimately, I think the decision makers should be the constituency demanding research. So maybe the best strategy is to try to encourage the states and districts to increase their capacity to engage in or digest research—my anecdotal impression is that there is currently tremendous variation in capacity across states. Maybe this happens by getting researchers elected to policy-making positions, as was suggested above, but I also think we need to look more to programs, like Harvard’s Strategic Data Project (<https://sdp.cepr.harvard.edu/fellowship>), that are specifically designed to develop talent around data analytics in school systems and state education departments.

I’m going to close with a bit of advice (which hopefully does not come off as patronizing or preachy). First, to the funders out there, think hard about how you measure impact. You have a role to play when it comes to determining whether it is true for the research community that “all press is good press.” For funders who value good empirical work, I would encourage looking beyond media-based measures of impact, emphasizing more the degree to which researchers are devoted to empiricism, including asking reviewers to judge researchers on this criterion when weighing funding decisions. Second, the incentives need to change in academia as well. It is no surprise that scholars pay attention to journal publications and citations as those are the currency of the academic realm. I’m not sure how to judge the level of public engagement by academic scholars, but, if we want more of it, public engagement must be more valued in determining tenure, promotion, and compensation.

And, finally, a bit of advice to young scholars: play the long game. What I’m suggesting is that it is worthwhile to sacrifice some short-term media impact in order to maintain credibility as a researcher who bows down primarily to empirical evidence. I wish I could suggest this is the best thing for one’s career but, unfortunately, I’m not sure that is the case. I do, however, believe that it is the best strategy for those wanting to develop a reputation as a straight shooter and, I hope, to have a larger death-bed impact.

ACKNOWLEDGMENTS

I would like to thank Dominic Brewer, Nate Brown, Jordan Chamberlain, Carrie Conaway, James Cowan, Elizabeth Farley-Ripple, Cory Goldhaber, Cyrus Grout, Katharine Strunk, and Roddy Theobald for their feedback and suggestions on the survey instrument or earlier drafts of this essay. I would also like to thank all Association for Education Finance and Policy survey respondents for their thoughtful input and time spent completing a survey. The views expressed here are those of Dan Goldhaber and do not necessarily reflect those of the University of Washington or AEFP. Any and all errors are solely Dan’s fault.

REFERENCES

- Chetty, Raj, John Friedman, and Jonah Rockoff. 2016. Using lagged outcomes to evaluate bias in value-added models. *American Economic Review* 106(5):393–399. doi:10.1257/aer.p20161081.
- Chetty, Raj, John Friedman, and Jonah Rockoff. 2017. Measuring the impacts of teachers: Reply to Rothstein. *American Economic Review* 107(6):1685–1717. doi:10.1257/aer.20170108.

Cody, Anthony. 2011. *Teachers: How do we propose to measure student outcomes?* Available http://blogs.edweek.org/teachers/living-in-dialogue/2011/01/teachers_how_do_we_propose_to.html. Accessed 21 August 2017.

Corcoran, Sean, and Dan Goldhaber. 2013. Value added and its uses: Where you stand depends on where you sit. *Education Finance and Policy* 8(3):418–434. doi:10.1162/EDFP_a_00104.

Dayen, David. 2015. *We must despise our kids: Our ugly war on teachers must end now*. Available http://www.salon.com/2015/10/06/we_must_despise_our_kids_our_ugly_war_on_teachers_must_end_now/. Accessed 18 September 2017.

Dee, Thomas S., and James Wyckoff. 2015. Incentive, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management* 34(2):267–297. doi:10.1002/pam.21818.

Durlak, Joseph A. 2011. *The importance of implementation for research, practice, and policy*. Washington, DC: Child Trends Research Brief No. 2011-34.

Gamson, David A. 2015. The dismal toll of the war on teachers. *Newsweek*, 5 October.

Goldhaber, Dan D., and Dominic J. Brewer. 2008. What gets studied and why: Examining the incentives that drive education research. In *When research matters: How scholarship influences education policy*, edited by Frederick M. Hess, pp. 197–217. Cambridge, MA: Harvard Education Press.

Goldhaber, Dan, and Duncan Dunbar Chaplin. 2015. Assessing the “Rothstein Falsification Test”: Does It really show teacher value-added models are biased? *Journal of Research on Educational Effectiveness* 8(1):8–34. doi:10.1080/19345747.2014.978059.

Goldhaber, Dan, Cyrus Grout, and Nick Huntington-Klein. 2017. Screen twice, cut once: Assessing the predictive validity of applicant selection tools. *Education Finance and Policy* 12(2):197–223. doi:10.1162/EDFP_a_00200.

Kane, Thomas J. 2016. Never judge a book by its cover—Use student achievement instead. Available www.brookings.edu/research/never-judge-a-book-by-its-cover-use-student-achievement-instead/. Accessed 12 July 2017.

Miles, Karen Hawley, Allan Odden, Mark Fermanich, and Sarah Archibald. 2004. Inside the black box of school district spending on professional development: Lessons from five urban districts. *Journal of Education Finance* 30(1):1–26.

O’Donnell, Carol L. 2008. Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research* 78(1):33–84. doi:10.3102/0034654307313793.

Reality. Any Year.

Rothstein, Jesse. 2017. Revisiting the impacts of teachers [Comment]. *American Economic Review* 107(6):1656–1684. doi:10.1257/aer.20141440.

Tseng, Vivien. 2012. The uses of research in policy and practice. *Social Policy Report* 26(2):3–16.

Turley, Ruth N. López, and Carla Stevens. 2015. Lessons from a school district-university research partnership: The Houston Education Research Consortium. *Educational Evaluation and Policy Analysis* 37(1S):6S–15S. doi:10.3102/0162373715576074.

Whitehurst, Grover J. “Russ.” 2016. Hard thinking on soft skills. *Brookings Evidence Speaks Reports* 1(14):1–10.

APPENDIX A: PRIORITY RANKING DATA

Table A.1. Priority Rankings for Improving the Overall Quality of the Teacher Workforce

	Panel A: Priority rating based on <i>empirical evidence (%)</i>					Panel B: Priority rating based on <i>all relevant factors (%)</i>				
	HIGHEST		LOWEST			HIGHEST		LOWEST		
	1	2	3	4	5	1	2	3	4	5
Teacher prep	27	23	16	22	12	28	25	18	13	15
Alt cert	6	18	20	23	33	12	15	18	27	28
Hiring practices	9	19	34	24	14	16	18	34	26	6
PD	20	24	15	14	26	31	28	12	11	18
Differential retention	38	17	15	17	14	14	13	18	23	32

Notes: Teacher prep = improving the preparation of teacher candidates; alt cert = giving school systems greater discretion over teacher hiring; hiring practices = improving the practices for hiring teachers; PD = improving the performance of in-service teachers; differential retention = differential retention of existing teachers based on effectiveness.