Two new reports claim to offer empirical support for the efficacy of voucher programs that allow parents to use taxpayer dollars to send their children to private schools. One report (A Win-Win Solution) is the latest in a series from the Friedman Foundation for Educational Choice. The Friedman report reviews studies purporting to show positive impacts from voucher programs in the US. The other report (The Participant Effects of Private School Vouchers across the Globe) is from the Department of Education Reform at the University of Arkansas. The authors of the Arkansas report conducted a limited meta-analysis of US and international studies of voucher programs. The two reports share a positive view of the impacts of vouchers, and both focus on randomized studies of those effects. Both reports are marred by a number of serious problems and errors, including misrepresentations of the research literature, a failure to acknowledge the limitations of their approaches, not addressing the shortcomings of the theoretical underpinnings of vouchers, and the use of methods that bias the selections of the studies they utilize. The Friedman report is a rudimentary “vote-counting “analysis of an extremely narrow set of 18 studies using a biased counting system. The Arkansas meta-analysis aspires to be “global,” but despite identifying over 9,000 potential studies for the analysis, ultimately uses only 19, almost half of which were conducted by the Arkansas authors or their associates. Moreover, the “global” meta-analysis only encompasses three countries (one of which is consistently misspelled). Together, their manifold serious flaws undercut the trustworthiness and usefulness of these reports.
I. Introduction

The degree to which students benefit from vouchers to attend private schools has been debated for years, with many studies suggesting little to modest benefits, at best, but also no measurable harm. While school choice advocates have insisted that there is a “hidden consensus” in the “highest quality,” randomized studies indicating significant, if inconsistent benefits for students using vouchers, a few recent studies using randomization to examine relative gains for voucher students have found evidence of large negative impacts for those students. This raises the question as to whether there is indeed a change in the “hidden consensus” on the impact of vouchers, or whether it even exists. Using two different approaches, a pair of new reports reviews the evidence, and contends that there is overall empirical support for the efficacy of vouchers.

- The first study, from the pro-voucher advocacy organization, the Friedman Foundation for Educational Choice, offers the latest of its series of reviews on the topic, finding the weight of these studies provides substantial evidence on the efficacy of vouchers in a number of areas. The report, *A Win-Win Solution: The Empirical Evidence on School Choice*, by Greg Forster, is the fourth edition of these summaries, and essentially employs a vote-counting exercise of studies that match criteria set by the author. While the report weighs in on a number of outcomes from voucher programs, including the competitive and fiscal impacts on public schools, the effects on civic values, and on racial segregation, these issues have not been seen as central to questions of voucher efficacy, and are not always illuminated by randomized studies. Instead, the foremost and long-standing focus of the Friedman Foundation has been on the immediate or “first-order” academic effects on students awarded vouchers through a lottery. Since most policy and scholarly interest has been on these first-order impacts (and that is also the exclusive scope of the other report examined here), this review focuses on the Friedman Foundation report’s treatment of the evidence on the achievement effects.

- The other study, *The Participant Effects of Private School Vouchers across the Globe: A Meta-Analytic and Systematic Review*, is from M. Danish Shakeel, Kaitlin Anderson, and Patrick Wolf, scholars at the Department of Education Reform at the University of Arkansas. The third author in particular has long been associated with questions of achievement in voucher programs, having found — sometimes controversially — positive impacts from such programs, although his most recent evaluation found negative impacts that were large and significant.
in Louisiana’s voucher program. This meta-analysis goes beyond the simple vote-counting efforts of previous syntheses, such as those by the Friedman Foundation. The Arkansas report also seeks to move the debate beyond the focus on US programs, and incorporate a global view.

Together, these reports are notable in their efforts to focus on rigorous research, although the ways in which they approach that task raises questions about the degree to which the authors lead the reader to certain conclusions regarding the voucher debate: The Friedman Foundation report demonstrates narrow attention to certain studies in the US that shine a positive light on vouchers. The Arkansas report similarly seeks to elevate a narrower view of empirical evidence on vouchers, while at the same time expanding the geographic basis for that approach to the globe — although in this case that means only examining two other countries. Addressing the questions of what evidence to consider and what to exclude in examining voucher efficacy is a crucial concern in understanding the real and potential impact of vouchers.

II. Findings and Conclusions of the Reports

As noted, the two reports take somewhat different approaches to addressing the achievement impacts for students awarded vouchers, but assert similar conclusions, although the Arkansas meta-analysis offers more sophisticated and specific estimates of those impacts. While both studies focus on randomized controlled trials (RCTs), it is important to note that they present contrasting findings for the US, even though they draw on a similar set of RCT studies. While the Friedman Foundation found that the set of studies from the US “shows that school choice benefits students,” the more nuanced analysis from the University of Arkansas noted that any impacts in the US are relatively small, especially for students using vouchers, and in math, where the impact “was not statistically different from zero.”

The Friedman Report

While vouchers in the modern sense have been a part of the education policy landscape in the US since 1990, the Friedman Foundation report includes only 16 studies that examine academic impacts of vouchers (two additional studies examined voucher impacts on college attendance/attainment). Using a simple vote-counting approach, the report classifies any positive effect for any subgroup as providing evidence of voucher efficacy. The report contends that 12 of the randomized studies show evidence of a positive effect on achievement (in addition to the two examining impacts for college outcomes); while two show no effect, and two show a negative impact. Interestingly, none of the studies had mixed results in the Friedman Foundation accounting; that is, they either showed positive effects for some or all students, no visible effects, or some negative effects. Notably, the author spends considerable effort (a) in arguing that one of the reports, with null findings on the New York program, should be listed as a positive impact (although, to his credit, he ultimately still lists
it in the “no effect” category), and (b) in trying to explain the substantial negative effects shown in both studies of Louisiana’s voucher program.

The Arkansas Report

Rather than simply categorizing whole studies into one of three categories (positive, negative, or null effects), the Arkansas meta-analysis incorporates effect sizes from 19 studies on 11 voucher programs in the US, India, and Colombia, allowing for multiple estimates of effects based on subject area (reading/English or math), location (US or not), funding type (public or private, which the authors also treat as a proxy for full or partial funding, respectively), as well as program longevity. Recognizing the fact that randomized trials can face the challenge of non-compliance, as students randomly assigned to the treatment or control group may make other choices, the authors report both Intention-to-Treat (ITT) and Treatment-on-Treated (TOT) estimates. This essentially makes the perhaps subtle but important distinction between voucher winners and actual voucher users. Thus, the former represents estimates for students assigned to a treatment group, regardless of which school they eventually attended, and the latter provides estimates for students whose subsequent school enrollment aligned with their assignment during the randomization process.

In the Arkansas report, the authors find a substantial range of outcomes. For instance, they note null ITT effects in reading for US programs, although they find substantial impacts for students in non-US programs, driven largely by student gains in Bogota. The report finds evidence of a generally positive trend over time, again in large part due to the program in Bogota. While the report often estimates null or positive impacts, the authors also find negative effects of vouchers. For instance, when they remove the outlier of Bogota to test the robustness of their findings, the authors note an overall negative impact in math for students using vouchers. Overall, they find the impacts are greater for publicly financed programs, for non-US programs, and in reading than math. Although they do not consider the research literature that might shed some light on these patterns (or lack thereof), nor try to explain how these disparate outcomes can be understood in terms of the theoretical underpinnings for vouchers, this last finding may be predicted by the fact that math is a better indication of school effects, and voucher schools have no automatic advantage in effectiveness.

III. The Reports’ Rationales for Their Findings and Conclusions

Both the Friedman Foundation and the Department of Education Reform at the University of Arkansas have long had an interest in the voucher issue — indeed, it is the reason for the Friedman Foundation’s existence. In fact, both are generally seen as advocates for market-based education reforms. As the Arkansas report indicates, school choice is held in these reports as being demand-driven, rather than something pushed by policy elites and privatization advocates. The Friedman Foundation report more explicitly embraces an advocacy
position, befitting its “mission to advance [a] vision of school choice for all children.”
Echoing its founder’s libertarian theories, it decries what it sees as harmful “regulatory burdens” on private schools in choice programs, such as that “participating schools must administer the state test and can be removed from the program if their scores are too low. They are also subject to inspections by public school officials while tests are being administered.” The report concludes that choice is not working better because there are too many regulatory burdens on private schools.

Both reports appear to be motivated by the need for positive evidence on the efficacy of vouchers, and they exclude research that does not align with their criteria. The Arkansas report in particular starts from an understanding of the limitations of extant reviews, which it attempts to document, and expresses the need for a meta-analysis that goes beyond simple vote-counting approaches in order to take into account effect sizes and other concerns not captured in previous approaches. However, it should be noted that the authors and associates have also produced vote-counting analyses of the voucher literature.

Furthermore, both reports hold fast to a strong preference for randomized controlled trial (RCT) studies. Advancing from the position that other methodological approaches do not offer the same high quality insights, both reports ultimately exclude all non-RCT research studies on vouchers. Indeed, there are pronounced benefits from RCTs, particularly in minimizing selection bias. Since researchers want to know how much a voucher impacts achievement outcomes for students, randomization allows for the creation of comparisons where the voucher is thought to be the primary distinguishing factor between groups.

IV. The Reports’ Use of Research Literature

It is a convention in empirical research to situate one’s study in the broader research literature, noting what is already known, where gaps in our collective knowledge may be evident, promising approaches, etc. The Friedman Foundation study generally neglects this important step. Aside from referencing previous editions of its own report, the author only mentions two other works, in an endnote, in establishing the basis for this report. Even then, it erroneously asserts that “most of the studies that rise to a reasonable level of scientific quality have found in favor of private schools,” but then provides only one citation which provides no such claim or evidence.

The Friedman Foundation report in particular conflates different academic outcomes in arguing for a positive impact for vouchers — including both academic achievement and other outcomes such as graduation rates and college attendance. While this present review has focused on achievement, it should be noted that voucher advocates are increasingly turning their attention to non-achievement outcomes (perhaps after realizing that any impacts in achievement are not that compelling). But randomized controlled trials do not account for peer effects in estimating the impact of vouchers, and it could very well be that, while we know that achievement can be impacted by a student’s peers, non-achievement outcomes

http://nepc.colorado.edu/thinktank/review-meta-analysis
might be even more susceptible to the attitudes and dispositions of one's classmates. Moreover, inasmuch as there is an emerging consensus that any impact from vouchers is more evident in non-achievement outcomes than in achievement, this raises the question as to whether policymakers and parents are willing to accept no or negative learning impacts.\(^{18}\)

The Arkansas report offers a more comprehensive treatment of the literature — in fact, three treatments. Unlike the Friedman Foundation, the Arkansas report provides a brief overview of the issue, drawing on a few historical and empirical sources. Then the report provides a systematic critique of the extant reviews of voucher studies (including some previous editions of the Friedman Foundation reports) before describing its own meta-analysis of the literature.

I review here the Arkansas report’s treatment of previous reviews, which serves as the justification for its subsequent meta-analysis. Then in the next section, I review the aspects of the reports that are intended as their primary contribution: the vote-counting analysis in the Friedman Foundation report, and the meta-analysis in the Arkansas report.

Unlike its subsequent meta-analysis, the University of Arkansas’s intended “systematic review of the systematic reviews of voucher effectiveness” does not appear to be as comprehensive, systematic, or careful as it claims. Presumably to establish the need for its comprehensive, “global” meta-analysis, the Arkansas report examines 10 reviews of voucher achievement effects in the US published from 2008-2015, and then engages in some basic analyses of which studies were covered, or — according to the report — should have been covered by these reviews. It is unclear why the report includes only reviews of US voucher programs in justifying a global meta-analysis, especially when other international reviews are already available\(^{19}\) (although the “global” meta-analysis only covers three nations, with the vast majority of the studies coming from the US). While the authors describe in great detail the process for selecting individual voucher studies in their subsequent meta-analysis (see below), the process for selecting reviews of voucher studies is unclear. The “systematic review” neglects to include, for instance, the Friedman Foundation’s 2009 review of voucher studies, and criticizes reviews published over the past three years for their “omission” of recent voucher studies that were only published within the last year.

The Arkansas review of reviews also misrepresents the studies by suggesting that these ten reviews were presented as meta-analyses (only one was). In another instance, a review from Coulson of the Cato Institute is not a review of voucher studies per se, but of public-private school comparisons.\(^{20}\) Studies of school vouchers address a different question than do studies of the relative effects of public and private schools.\(^{21}\) The former examines non-representative subsets of schools from the different sectors, while the latter looks at public and private school effectiveness. Nonetheless, the Arkansas researchers persist in this erroneous conflation of empirical findings.

Similarly, the University of Arkansas review of reviews includes one analysis of the use of voucher research that was explicitly not an analysis of vouchers per se.\(^{22}\) In that study the authors clearly noted that their analysis centered on public-v-private studies, and then an
evolution of policy debates around vouchers, explicitly focusing on studies voucher advocates had highlighted and often misrepresented in addressing the public-private question — not all extant voucher studies. Thus, the University of Arkansas’s assertion that “[e]very study that was released during that period should have been included in the review” makes no sense,\(^2\) since it was clear that not all of these “reviews” were intended as comprehensive treatments of extant voucher studies. Even then, the University of Arkansas report faults authors for not including studies that they in fact clearly cited. This undercuts the integrity of University of Arkansas’s attempts to quantify the comprehensiveness of previous voucher reviews. Likewise, the University of Arkansas report faults reviews such as that from Usher & Kober\(^2\) for not including studies from an arbitrary time period suggested by University of Arkansas authors, even though the authors of reviews clearly focused on a different time period.

Such fundamental errors undermine the credibility of the University of Arkansas analysis. By ascribing a failure of these analyses to do something that they did not claim to do, the University of Arkansas report distracts attention from the actual findings of those studies, which showed limitations of vouchers and research advocating vouchers, as well as the “political motivations of voucher evaluators.”\(^2\)

V. Review of the Reports’ Methods

While the two reports focus on RCTs in voucher research, and assert similar conclusions, they vary considerably in their methods. And they share some of the same limitations and shortcomings.

The Friedman Report

The Friedman Foundation vote-counting analysis includes what is essentially the same set of studies it has drawn upon in previous editions of its report, while adding six additional studies for various reasons. The review includes 18 studies, although only 16 focus on academic achievement. Notably, one prominent school choice advocate, or his students, produced 10 of those studies. While the Friedman Foundation contends that it conducted systematic searches “to help ensure the review was comprehensive,”\(^2\) five studies have been added since the previous edition of the Friedman Foundation report, having come to the author’s attention informally, either through his own ongoing work in the school choice research field or as a result of others in the field bringing these studies to his attention. (It is difficult to work in this field and not be aware of new studies as they come out!)\(^2\)

Yet despite the claim that it is difficult to be unaware of relevant studies, the author had somehow missed a published, peer-reviewed 2006 study (that happened to show no effects
for vouchers) in the three previous versions of the Friedman Foundation report. Moreover, the author’s reliance on “others in the field bringing these studies to his attention” raises concerns about potential bias in the set of selected studies, since voucher advocacy research tends to operate in ideologically defined echo chambers.

The Friedman Foundation uses a simple, and questionable, approach to classifying studies for its vote-counting analysis. Studies are classified into one of three categories, depending on if they show evidence of “no visible effect,” “any negative effect,” or “any positive effect” (with this last category being sub-divided between positive effects for “all” or “some students”).

Of the six studies added to this edition of the Friedman Foundation report, one – showing no effects from vouchers – had been previously missed by the author. In what appears to be an attempt to stuff the ballot box in this vote-counting analysis, another – a 2004 rebuttal – was added to the analysis as an additional vote for positive voucher effects, even though it was not counted as such in previous editions of the Friedman Foundation reports, and, in an apparent case of double-counting, involves the same authors looking at the same program as they had in another study which was also listed as a “positive effect.”

Two recent studies finding large negative effects of vouchers in Louisiana were included in the new Friedman Foundation report. Another new study that was included focused on college attainment (not on achievement effects). The sixth study concluded that

the NYC voucher experiment had little effect across the distribution of student achievement, with the possible exception of small negative effects in math in a small region near the top of the distribution of students who sought vouchers, which fade out over time.... Overall, the distributional findings are most consistent with our ... hypothesis, that vouchers (at least of this magnitude) have no positive or negative effect for the vast majority of students to whom they were offered.

Nevertheless, the Friedman Foundation classifies this report as demonstrating “positive effects” if it has any single positive estimate, even when a “study typically includes multiple analytical models — sometimes many of them, occasionally even more than 100.” (While a single negative estimate could also place a study in the “negative effect” category, there are no such instances of this in the Friedman Foundation report.) The Friedman Foundation claims this approach is a way to avoid accusations of “cherry-picking,” although, as used in the report, the approach gives the appearance of exactly that.

The Friedman Foundation report also uses a questionable approach to classifying studies as showing positive impacts for “some” or “all” students. For instance, the Friedman Foundation classifies the results of the DC voucher evaluation as having positive effects for “all students.” The evaluation indeed found “marginally statistically significant positive overall impact of the program on reading achievement after at least four years. No significant impacts were observed in math.” Yet the evaluators’ analysis of the impact on subgroups found statistically significant impacts in reading only for half the sub-groups studied, and
not for students who left lower-performing schools for the voucher program, started at a lower level, or for male students. Still, the Friedman Foundation approach categorizes such a study as a vote for “positive impacts” for “all students.”

The Friedman Foundation review uses the same vote-counting approach to make the same arguments it has used in previous editions of its report, and expends considerable effort to dismiss findings that do not support the Friedman Foundation’s pro-voucher agenda, without giving equal scrutiny to studies whose findings align with the Foundation’s announced objective. For instance, in the section devoted to discussion of academic outcomes in voucher programs from 18 RCT studies, almost three-quarters of the space is devoted to either (a) explaining why one study “must be regarded as discredited” because it classifies students as African-American if either parent is African American (since this simple issue can change the findings from “no impact” to “positive impact”); (b) justifying why a study of NYC vouchers should be counted as showing positive impacts, even though the authors concluded otherwise (see above); and (c) speculating as to reasons for the “anomalous” but large negative impacts noted in both studies of vouchers in Louisiana — even though the author’s previous speculations as to factors shaping education outcomes have proven to be spectacularly wrong.

The Arkansas Report

As a “global” meta-analysis, the University of Arkansas report offers a much more sophisticated and ambitious approach to estimating the impacts of vouchers on academic achievement than does the Friedman Foundation vote-counting of US studies. Still, the meta-analysis brings its own set of limitations, problems, and errors.

With a few exceptions, the University of Arkansas report is relatively transparent in its methods of identifying and analyzing studies. This is a crucial concern, since methodological decisions can affect the outcomes, and the strength of any meta-analysis is based on the selection process used to include or exclude studies, and the quality of those studies. Here, the meta-analysis is used to draw data from the different studies to generate more precise (and potentially more statistically powerful) estimates of the average impact of voucher programs.

Despite the transparency, and the laudable goal of moving the discussion beyond US programs, there are a number of questions, concerns, and potentially problematic methodological decisions that may bias the findings of the meta-analysis. For instance, the report acknowledges that “the conclusion one draws about the efficacy of vouchers is heavily influenced by the body of studies one reviews.” Yet, although the report discusses at length the process for identifying studies, the review then “utilized subject matter experts in the field and snowballing techniques to find additional relevant studies.” Yet we don’t know who these experts are. Since it might be expected that they may be drawn from ideologically defined networks (indeed, only colleagues affiliated with the Department of Education Reform at the University of Arkansas are listed in the acknowledgements), it would have been
useful to note how many, and which, of the studies ultimately chosen for inclusion in the meta-analysis came from such sources. In fact, the authors write that their search led to four non-US studies being “uncovered,” but this included one study led by one of the co-authors of the meta-analysis.\textsuperscript{39} In fact, nine of the 19 studies ultimately used in the meta-analysis, from an initial set of over 9,000 considered, were conducted either by one of the co-authors of the meta-analysis, their co-authors from another voucher study, or colleagues at the Department of Education Reform at the University of Arkansas.

In assembling the larger set from which to identify studies for the meta-analysis, the authors conducted searches of online databases — a sensible and relatively transparent approach explained in the report. But the authors included only studies published in English, and searched for terms like “voucher” and “opportunity scholarship.” Such approaches could be problematic, as the word “voucher” is often used in some other countries more in the sense of a “coupon.” While the authors included “education” or “school” to make sure their search would return primarily those studies focused on Freidman-style vouchers, there is still a prior but untested assumption that such programs and researchers in other countries use the term “voucher” to describe the sort of programs of interest. Similarly, the alternative search term — “opportunity scholarship” — was a phrase suggested by pollster and word-smith Frank Luntz in his advice to Republican members of the US Congress because it polled much better (66%) with American parents than did the term “voucher” (23%).\textsuperscript{40} Thus, it is far from clear whether the report’s search strategy really returned a globally representative set of studies.

In addition to excluding any study not available in English, the report also excluded any unpublished studies available in theses or dissertation databases, under the logic that they “expect that any experimental evaluation of a school voucher program that is the subject of an original thesis or dissertation will be sufficiently important that it also will be released as a study report or journal publication.”\textsuperscript{41} Yet a well-known challenge for meta-analyses is to avoid or account for publication bias, and it is very possible that even a rigorous, high-quality treatment of vouchers will be less likely to be published if it produces null results. This unfortunate decision might be expected to bias the University of Arkansas meta-analysis to make estimated effects appear more pronounced by excluding studies finding null results.

As with the Friedman Foundation report, the University of Arkansas meta-analysis focuses only on RCT studies. While a defensible decision — albeit not the only choice — for a meta-analysis, it is important to remember that such a narrow approach excludes a rich array of quasi-experimental and other studies that can also shed light on the voucher question. Even if other studies did not meet University of Arkansas’s criteria, they should have been considered at least in the preliminary discussion in order to inform the analysis in terms of theoretical, policy and contextual considerations, especially since many of these extant studies focus on larger and more developed voucher programs.\textsuperscript{42} And because they are often larger in scale and can offer insights into school and home-background factors not accounted for in most RCTs, such studies sometimes offer a broader and more illuminating light on school choice issues.\textsuperscript{43} Indeed, the exclusive focus on RCTs means not only eliminating studies employing different approaches, but excluding the learned lessons of whole coun-
tries like Sweden and Chile that have a longer history with more comprehensive voucher programs than, say, the small-scale, targeted programs in Charlotte or Dayton included in the Arkansas report.

The fact that the University of Arkansas report imposed criteria that narrowed the pool of over 9,000 studies to just 19 for the meta-analysis, 15 of which were in the US, along with two pairs of studies on India and Bogota, Colombia (repeatedly mislabeled as “Columbia” in the report) suggests a shrunken vision of the globe. Thus, this is a “global” meta-analysis in the same sense that the championship for American baseball is the “World Series.”

Even then, there is concern that drawing on international data in this regard involves equating some rather disparate programs and contexts. While the report notes that all the programs share some basic factors, and that all the studies are RCTs, there are still important distinctions neglected by such an exercise. Just as “vouchers” can mean different things in different countries, even the basic idea of “public” and “private” schools can be very different across contexts. The US distinction between public and private school sectors is hardly universal. The US private school sector enjoys substantial autonomy relative to public schools, even though the public sector is relatively decentralized, and funding of the private sector is almost always from private sources (despite the efforts of voucher advocates). But “private” and “independent” schools in other nations are often more regulated than US public schools, and many nations provide substantial funding to the private sector, including religious schools. The University of Arkansas report does not appear to acknowledge such considerations, and instead appears to be based on a rather US-centric set of assumptions.

Similarly, the programs included in the analysis are hard to compare to each other. Programs (and schools) are likely more similar within countries, and more different across national boundaries. As the authors note, the differences between public and private schools in Colombia might be much greater than what is seen in the US, and explain the fact that one city — Bogota — skews the overall results for the meta-analysis. Moreover, many low-fee private schools in India are often simple store-front, mom-and-pop operations that are very difficult to equate to, say, a private religious school in New York; so vouchers within such disparate contexts might be expected to have very different impacts and introduce quite distinct dynamics. Likewise, specific policies differ: programs may be open to students based on family income; others have residency requirements; some, such as Colombia, have minimum academic standards — thus the programs are created with different objectives in mind, including equity, achievement, competition, and institutional support. Indeed, the cases bring very different historical, demographic, policy and institutional contexts that might be expected to shape voucher programs, their uptake, use, and effects, but these considerations are brushed aside. In fact, without consideration of such policy differences, the report makes the unsubstantiated claim that “most publicly-funded vouchers must be accepted as the full cost of educating the child.” That is simply not true, for example, with the large and long-standing Chilean voucher program.

Despite the dramatically varied contextual and program issues, the meta-analysis treats the programs examined in the 19 studies as “functionally equivalent,” and combines and an-
alyzes the data from these studies, presenting estimates for math compared to reading/English public to private funding, longevity, and geography (US v India/Bogota). The meta-analysis does not consider other important issues, such as religious v. non-religious schools, urbanicity, the relative amount of the voucher in different countries (other than a vague proxy of public v private funding) or relative to funding for public schools in cities with vouchers, etc. Perhaps most importantly, after presenting us with distinctly different outcomes by subject, geography, and funding, the University of Arkansas meta-analysis is ill-equipped to offer insights into which factors might explain more or less effective voucher programs. Instead, it presents us with inconsistent and haphazard outcomes — for some unexplained reason, vouchers “work” in some programs, for some students, in some subjects, but hurt similar students in others — that call into question the validity and usefulness of the theoretical foundation for vouchers.

VI. Review of the Validity of the Findings and Conclusions

A primary strength of RCTs is their internal validity — that is, they exemplify a strong design that allows for useful insights into the data being studied. However, they have relatively weak external validity, presenting serious challenges for generalizing to other populations. The RCTs highlighted by both reports may do a relatively good job of providing estimates of the impacts of vouchers on the specific populations offered and using them in given contexts (although with some limitations, discussed below). But this also means that findings from these particular program evaluations cannot be readily generalized to the broader population, as with calls — such as from the Friedman Foundation — to “scale up” voucher programs. Essentially, RCTs offer insights only on the question of impacts for the students who tend to apply for these specific programs, and not the population in general, and for the schools (or programs) that are oversubscribed, and not for all private schools that could potentially accept vouchers.

The meta-analysis does not fundamentally address these limitations. Even though the approach draws data from evaluations of different programs, the studies still depend on students who self-select into the randomization process by choosing to apply — a distinctive and observable act that suggests that they may differ qualitatively from others who did not so choose. Moreover, RCTs are limited in that they only attempt to tell us if, not how, a voucher may have had an impact. They shed little light on questions of school effects, demographic/family background factors, or peer effects, even though these latter issues are known to be outsized factors, according to the research literature, and offer no real insights into promising practices that might be replicated.

Indeed, both reports appear to place all their faith in the randomization process, since it eliminates selection bias — “at least in theory,” the Arkansas report notes — so that they pay little attention to such limitations for RCTs for addressing the questions of expanding vouchers to a broader scale, even though the Friedman Foundation report calls explicitly for
that. Indeed, a number of researchers, including the University of Arkansas authors’ own colleagues, have recently acknowledged the limitations of randomization, pointing out that it is inappropriate to generalize to broader populations of students and schools from these studies, and that scaling up the programs might diminish any returns. RCTs, as they have been leveraged in voucher research, have been very narrow in what they can tell us.

Moreover, one of the primary concerns with meta-analyses in particular is that the results are dependent on the specific sets of studies chosen for review. Therefore, a transparent selection process is helpful to minimize such concerns, but disclosure of potential biases is also important. The University of Arkansas report makes note of the possible biases of some of the reports it reviews, for instance, noting that the Center on Education Policy is “generally viewed as opposing market-based reforms such as school vouchers.” But the Arkansas report fails to disclose its own funding source, or that of the Department of Education Reform, nor note how it is perceived.

VII. Usefulness of the Reports for Guidance of Policy and Practice

The reports reviewed here have garnered attention in the broader policy discussion, although often primarily with pro-voucher advocates, reflecting the echo-chamber nature of the discussion reflected in these reports. Neither report has been independently peer reviewed, and they suffer from the problems noted above, which undercut their credibility.

Both reports make a number of methodological choices that shape their results. They focus on randomized controlled trials for defensible reasons, but without acknowledging the limitations of RCTs, even though those limitations are widely known. The sets of RCT studies selected for use are also problematic, with the Friedman Foundation report using a process that is not as systematic as it indicates, and in fact is shown to have been erroneous, while the University of Arkansas report’s selection criteria left it with a highly skewed set of studies, and no insights from the longest-running and more comprehensive national voucher programs. At the same time, the University of Arkansas report makes no effort to account for publication bias in its selection process, which may exaggerate its findings.

The reports also take very different approaches to other voucher “impacts.” The University of Arkansas meta-analysis focuses only on academic achievement, and not on issues such as the impact on segregation, which is known to be a detrimental factor in countries that have embraced vouchers. The Friedman Foundation report does attempt to address some of those other factors, but in its treatment of academic achievement (the focus of this review) it conflates learning gains with non-cognitive outcomes such as college attendance, without apparently recognizing that these are distinct issues that are possibly susceptible to different influences besides the offer or use of a voucher.

Overall, the reports present findings that are not particularly helpful for advancing our understanding on the impacts of vouchers on student achievement. Together they tend to pres-
ent a mixed set of findings, with benefits appearing for one group in one context, but not for the same group in another city, or even in another subject. The results from the global meta-analysis are shaped largely by one city in one country, and the report is not designed to tell us why vouchers would have a greater impact there. Indeed, we don’t know why voucher impacts — positive or negative — appear in one case, but not in another. Thus, the theoretical underpinnings motivating vouchers, as set out by Milton Friedman and subsequent theorists, do not appear to be very strong when applied to the real world, even when tested by voucher advocates.

The Arkansas report concludes with recommendations for more RCTs, without offering any guidance about how researchers might overcome the limitations of randomization to take into account peer effects, or improve upon generalizability. But the report also encourages consideration of “the cost-benefit tradeoffs associated with voucher programs” — by which the authors mean to suggest that vouchers are “cost effective, since they tend to generate achievement outcomes that are as good or better than traditional public schools, but at a fraction of the cost.” Instead of focusing only on saving money when educating disadvantaged students, a better approach would be to consider vouchers in terms of their potential for enhancing student learning relative to the effects of other interventions’ ability to do so. For example, instead of looking at school vouchers alone, we should be considering the effect sizes of school vouchers compared to, say, housing vouchers, integrated classrooms, or smaller class sizes. In fact, perhaps the largest meta-analysis on education interventions has done just that, looking at the effect sizes of various influences on student learning, and finding little support for school choice programs. Drawing on over 800 meta-analyses encompassing over 50,000 studies of different influences in academic achievement, that meta-study ranked the school choice option 107th out of 138 factors, beneath class size, professional development for teachers and peer tutoring...but above summer vacation and television.
Notes and References


http://nepc.colorado.edu/thinktank/review-meta-analysis


46 The Arkansas report suggests that it “is a study of achievement effects of low-cost private school vouchers on low-income inner-city children,” (p. 25) but includes studies of voucher programs that are not relegated to urban areas.


http://nepc.colorado.edu/thinktank/review-meta-analysis 19 of 22


http://nepc.colorado.edu/thinktank/review-meta-analysis
Arkansas Department of Education Reform, 41.


59 The meta-analysis looked at charter schools, not vouchers, perhaps pointing to the limited embrace of the term “voucher.”

Authors: Greg Forster; and M. Danish Shakeel, Kaitlin P. Anderson, & Patrick J. Wolf

Publishers/Think Tanks: Friedman Foundation for Educational Choice; and Department of Education Reform at the University of Arkansas

Document Release Dates: May 18, 2016; and May 10, 2016

Review Date: June 30, 2016

Reviewer: Christopher Lubienski, University of Illinois

E-Mail Address: club@illinois.edu

Phone Number: (217) 333-4382

Suggested Citation: