Setting the Record Straight on Charter Schools and Achievement:
A Reply to Francesca Lopez

Julian R. Betts and Y. Emily Tang
Department of Economics, University of California, San Diego

In August 2014, we published a working paper, A Meta-Analysis of the Literature on the Effect of Charter Schools on Student Achievement. The paper was commissioned by the Center on Reinventing Public Education (CRPE) at the University of Washington Bothell.

We recently learned of a review of our work by Francesca Lopez, an associate professor in the College of Education at the University of Arizona. The review was published by the "Think Twice Think Tank Review Project" of the National Education Policy Center. We feel compelled to respond to Lopez because her essay, whether by intention or genuine confusion, misrepresents our analysis from start to end.

First, what did we find in our meta-analysis of the charter school effectiveness literature? On average, charter school studies are revealing a positive and statistically significant difference between math achievement at charter schools and traditional public schools. We also find a positive difference for reading achievement, but this difference is not statistically significant. Second, we devote much of our paper to studying not the mean effect, but the variation across studies in the effect of attending a charter school. We find that charter schools' effectiveness compared to nearby traditional public schools varies substantially across locations.

What is the central claim of Lopez? She writes: "The report does a solid job describing the methodological limitations of the studies reviewed, then seemingly forgets those limits in the analysis" (p. 1). She uses words like "exaggeration" and "overstated" (p. 8) to characterize our analysis of the literature, and implies that our conclusions are not "reserved," "responsible," (p. 7) or "honest" (p. 7 and p. 8). Throughout her essay, Lopez falsely projects intentions in our words that simply are not there. We encourage interested readers to review the words that we actually wrote, in their full context, in our abstract, main paper, and our conclusion. We are confident that readers will confirm for themselves that any "overstated" conclusions of which Lopez accuses us are imagined.

There are serious problems with Lopez's arguments. First, she habitually quotes our work in a selective and misleading way. Such rhetorical slights, in which she quotes one of our sentences while ignoring the highly relevant adjacent sentences, or even cutting important words out of our sentences, overlook important parts of our analysis and result in a highly inaccurate presentation of our work. Second, her analysis contains six technical errors. These technical mistakes, some quite serious, invalidate many of Professor Lopez's claims. An appendix to this essay exposes more than two dozen misleading or outright incorrect statements that Lopez makes in a mere 9-page essay. To give readers a sense of the scope and severity of these problems, consider the following examples:

Example 1: A Partial and Misleading Quotation

Lopez insinuates that we exaggerate the positive overall math effect while downplaying the percentage of studies that show negative results. She writes:
"The authors conclude that 'charter schools appear to be serving students well, and better in math than in reading' (p. 47) even though the report finds '…that a substantial portion of studies that combine elementary and middle school students do find significantly negative results in both reading and math—35 percent of reading estimates are significantly negative, and 40 percent of math estimates are significantly negative (p. 47)'"

Here is what we actually wrote on page 47: *"Examining all of these results as separate parts of a whole, we conclude that, overall, charter schools appear to be serving students well, and better in math than in reading. The caveat here is that a substantial portion of studies that combine elementary and middle school students do find significantly negative results in both reading and math—35 percent of reading estimates are significantly negative, and 40 percent of math estimates are significantly negative."*

Lopez uses two rhetorical devices to lead readers to the perception that we overstated findings. First, she separates the two quotations, implying that we are somehow hiding the second result, when in fact we intentionally mention the positive overall mean math effect and the variation in the results across studies side by side. Second, she further misleads the reader by again cutting out part of our sentence. Instead of stating that we have a "caveat" to the positive mean math effect she removes that entire clause.

What makes the approach of Lopez even more misleading is that in the paragraph above, we were bending over backwards to be fair. We cite only one type of study in that quotation: those that combine elementary and middle schools. (These account for about 1/7th of all the studies.) Why did we focus only on those studies in the above quotation? Because these studies were the exception to our conclusion—the ones that produced *the highest percentage of studies with negative and significant estimates*. **Wouldn't one think that if our goal had been to overstate the positive effects of charter schools we would never have chosen to list the result that is the *least* favorable to charter schools in the text above?** For example, we could have stated that for elementary school studies, only 12% showed negative and significant reading results, compared to 71% showing positive and significant results. Or we could have stated that only 11% of elementary school studies showed negative and significant math results, while 61% showed positive and significant results in math.

Lopez fails to list any of the more positive results from the other grade span categories studied *that led us to our overall conclusion*. We noted the exception above precisely because it was an exception. While it is worth noting, it does not refute the other evidence. By citing an exception as a reason to dismiss all of the other results, Lopez misses the main point of a statistical meta-analysis. This is a consistent pattern throughout her essay.

Example 2: Failure to Report Our Findings

Next, consider the implication by Lopez that we should not have included studies by the Center for Research on Education Outcomes (CREDO) due to technical concerns. She writes: *"Nevertheless, CREDO studies, which were 12 of the 20 elementary/middle school studies, remained in the main analyses because 'they include extremely large samples of charter schools" (p. 7)."* Lopez fails to mention to her readers that in Table 4 we re-do *all* of our main analyses after excluding the CREDO studies. In the section entitled *"Our Results Are Not Sensitive to Inclusion of the CREDO Studies,"* which spans pages 29-31, we show that the results are quite similar when we exclude the CREDO studies. Neglecting to inform her readers that we actually devote three pages of our report to doing exactly what she implies we should have done, and, further, neglecting to tell her readers that the results do not change, represent serious and careless omissions.

Example 3: Serious Technical Errors

Lopez also makes some serious technical errors. Here is one example. Lopez completely misunderstands our analysis of the likely sources of bias in the CREDO studies. She claims: "*The authors reported the CREDO studies, however, introduced biases that favored charter schools because of the matching approach used.*" In fact, we did not say that the CREDO estimates introduced favorable bias. On pages 29-30 we explain that we believe that the CREDO charter estimates "if anything, would be biased toward zero". This does *not* mean that the CREDO studies are biased in favor of charter schools at all. While in cases of a true charter effect being negative, a bias towards zero implies an estimate that is less negative than the true effect, in the cases of a true charter school effect being positive, the bias is actually *downwards*—the CREDO estimate may be too small.

Unfortunately, this is far from the only technical mistake Lopez makes. For a complete list of technical errors in the Lopez essay, some quite serious, see Misleading Statements #1, 6, 20, 22, 25 and 26 in the Appendix to this rebuttal. These errors have the compounding effect of making her case seem stronger than it is.

For instance, in Misleading Statement #6 she not only misrepresents the relationship between a lottery-based study and random assignment, but she then makes vague but undocumented allegations against authors of lottery-based studies, claiming that a bias could arise. Not only are we unaware of any lottery-based study that has made the error she alleges, but Lopez does not seem to understand that if the error had occurred, it means that lottery-based studies would probably be *understating* the impact of charter schools on achievement! As another example of a serious technical error by Lopez, in Misleading Statement #26 she accuses the charter school literature of making biased analyses because test score gains are biggest in the lower grades. Lopez is seemingly unaware that the vast majority of charter school studies do not use vertically scaled test scores, making her point irrelevant. And for the few studies for which her point could be relevant, she fails to make a compelling statistical case, and she further fails to point out that many studies divide their analyses into small grade ranges, making her point of marginal importance.

The above examples are just a few of the problems with Lopez's review. She also makes statements presenting subjective judgments as facts. Her concluding statement is this: "A more honest reading of the results would be that they are consistent with the large body of charter school studies: the overall test-score outcomes for charters and public schools are largely indistinguishable."

The statement that outcomes are "largely indistinguishable" has no factual bearing. We present two main results in the paper. First, the mean effect across studies for math is positive and significant for elementary studies, for middle school studies, for a combination of elementary studies, middle school studies, and elementary/middle school studies, and for studies of all grades. Lopez has no justification for her statement. Second, the *variation* in estimated effects across different locations is very big. As our very first sentence in the conclusion emphasizes, some studies produce positive effects, some produce no effects and some produce negative effects. Lopez's claim that test score outcomes are "largely indistinguishable" is a subjective judgment not based in facts.

Conclusion

We stand by our analysis. Lopez completely misreads our paper. In fact, if any good is to come out of Lopez's misconstruction of our work, it is that more people will read our paper so that they can see for themselves what we actually wrote.

APPENDIX: A LIST OF MISLEADING STATEMENTS AND TECHNICAL ERRORS MADE BY LOPEZ

The summary of the review on page 1 includes many incorrect statements. Consider this sentence: "Claims of positive effects when they are not statistically significant, exaggeration of the magnitude of effects, reliance on simple vote-counts from a selected sample of studies, and unwarranted extrapolation of the available evidence to assert the effectiveness of charter schools render the report of little value for informing policy and practice."

We challenge every one of these comments. Let's parse out this laundry list of allegations:

Misleading Statement #1

Lopez writes "This report attempts to examine whether charter schools have a positive effect on student achievement." The very first sentence written by Lopez in our summary is misleading. In statistics we test whether we can maintain the hypothesis of *no effect* of charter schools. We are equally interested in finding positive or negative results.

Misleading Statement #2

"Claims of positive effects when they are not statistically significant" is misleading. For example, our abstract states:
"Overall, for the limited set of charter schools, locations, and years that have been studied to date, charter schools are producing higher achievement gains in math relative to traditional public schools in most grade groupings. No significant differences emerge for reading achievement. However for both math and reading, the bulk of estimates are positive."
Where do we state that a result is statistically significant when it is not? Perhaps she is thinking about our very careful analysis of time trends where we find a greater number of grade spans showing positive math effects in our current study than in our older study from three years ago. We do perform a statistical analysis for time trends. We estimate a positive trend but it is not statistically significant. Our exact summary of that result from page 47: "Although the slope was positive in both cases, suggesting a positive time trend, in neither case was the trend significantly different from zero". We make a similar statement, as plain as day, in the abstract: "There is not a statistically significant link between the years covered by a study and the estimated effect size, but for both math and reading the trend is positive". Again, Lopez ignores our very careful and accurate statements of the findings.

Misleading Statement #3

Next, consider the misleading statement that our report shows a "reliance on simple vote-counts from a selected sample of studies". There are two factual errors here.
First, our report hardly relies on vote-counts. Our main analysis is a careful statistical meta-analysis. But because we are concerned that there are two messages in our study, one about a positive mean effect in math, and a second about variation across studies, we use a weighted histogram approach and a simple count of signs and significance (a "vote-count") as secondary means of showing the same thing that our formal meta-analysis showed. This is hardly a "reliance". It is simply another way of displaying information.
Second, the claim that we do a vote-count "from a selected sample of studies" insinuates incorrectly that we did not include certain studies arbitrarily. We included all studies that met our quality criteria by using one of the methods discussed on pages 5-8.

Misleading Statement #4

Next, her unsupported assertion that we undertake "unwarranted extrapolation of the available evidence to assert the effectiveness of charter schools" is itself completely unwarranted. Do we "extrapolate" to grades outside the grades studied by researchers? No. Do we extrapolate to states without studies? Again, no. Ironically Lopez fails to notice steps we take to ensure that the reader understands that the existing literature is young and that we need to expand the geographical coverage. The very first thing we mention in the section "Challenges for Meta-Analysis of the Literature" on page 3 of our report is this: "These analyses present several challenges. Though improved from the set of studies covered in our 2008 analysis, the available studies offer limited geographic coverage, potentially leading us to overstate the generalizability of results. To provide readers with a sense of how broadly based a given result might be, in section 3 we report not only an overall effect size but also the number of studies and the number of geographic locations underlying a given estimate." Further, we quote from out abstract to show the extraordinary care that we take \to ensure that readers do not extrapolate: "Overall, *for the limited set of charter schools, locations, and years that have been studied to date*, charter schools are producing higher achievement gains in math relative to traditional public schools in most grade groupings." (Emphasis added.) Lopez is completely off track in her charge of extrapolation.

Misleading Statement #5

The report "… finds charters are serving students well, particularly in math. This conclusion is overstated; the actual results are not positive in reading and are not significant in high school math; for elementary and middle school math, effect sizes are very small…"

Although we do conclude that charter schools are serving students well, particularly in math, we make great efforts to show that this is a statement about mean effects.

The first problem in her one-sided summary is that she mentions the two grade spans in which we found no significant math effects, but she neglects to list the four grade spans in which math effects are positive and significant. She cherry picks the results rather than providing a fair and balanced complete account, as we do.

As for the effects of reading, a non-technical reader of her summary statement that "the actual results are not positive" might infer that the actual effects might be negative. On the contrary the reading results are positive but are not statistically significant. Further, we show in Table 8 that for all but one of the six grade spans we analyze, a majority of reading estimates was positive. And in several grade spans, the percentage of studies showing positive and significant effects is far higher than the percentage showing negative and significant results. For example among 17 elementary school studies, 71% showed positive and significant results, compared to only 12% showing negative and significant results. For reading, the *only* case out of six grade spans where the percentage of studies showing negative and significant effects equaled or exceeded the percentage showing positive and significant effects was studies that combined elementary and middle schools, where 35% of studies were in each category.

Lopez also makes a misleading or at the very least, a highly subjective statement that the effect sizes are not big. We return to this in the final section of the rebuttal where we examine evidence that she has crossed the line between a dispassionate scientific analysis and an impassioned opinion piece.

Misleading Statement #6

Lopez writes "For example, the authors include lottery-based studies, considering them akin to random assignment, but lotteries only exist in charter schools that are much more popular than the comparison public schools from which students are drawn. This limits the study's usefulness in broad comparisons of all charters versus public schools."

There are three factual errors in this statement. First, lottery-based studies are not "akin" to random assignment. They *are* random assignment studies. Subject to testing for identical baseline characteristics between the treated students who win the lottery and the control students who lose the lottery, which is something all of the studies we cite do, these are gold-standard studies.

The second error Lopez makes is to state that charter schools with lotteries must be "more popular than the comparison public schools from which students are drawn." This is simply untrue. A charter school with a lottery may be more or less in demand than a given nearby traditional public school. All that is needed is that the number of applications must exceed the *number of seats available.* Third, consider the statement that using oversubscribed charters in lottery analyses makes them of limited use in comparing to all charter schools. While this may be true, Lopez fails to acknowledge that Betts and Tang (2011) were among the first to hypothesize this. Further, Lopez fails to mention that we formally test for differences in effect sizes between lottery studies, propensity score studies, fixed effect studies and other methods. We find no differences for reading across the methods. In the case of math achievement, we find that both propensity score and lottery based studies produce higher effect sizes than the comparison group, which is the CREDO studies and one instrumental variable study. There is hardly a smoking gun here. We have reason to believe that the CREDO studies may sometimes estimate effects that are too small in the sense of being close to zero.

So far, we have discussed only Lopez's one-paragraph summary, in which we have found no less than 6 misleading statements, some of which are outright errors. Sadly for Lopez, the misleading statements continue in extraordinary number in the following pages of her report.

Further Misleading Statements in the Main Body of the Essay:

Misleading Statement #7

Lopez quotes us on page 3 as saying:
"One conclusion that has come into sharper focus since our prior literature review three years ago is that charter schools in most grade spans are outperforming traditional public schools in boosting math achievement." (p. 53)
She then claims that this is an overstatement. In our 2011 study, charter schools' overall estimated math effects were positive and significant for 3 of 6 grade spans for which we conducted an analysis. In our 2014 study, math effects were significant for 4 out of 6 grade spans.[1] Moreover, in comparing the effect sizes between our two studies, three years apart, for math in 5 of 6 cases the effect sizes we estimated for grade-spans were larger in the current study. There is suggestive evidence here that later studies are more likely to find positive effects. Later on in the paper we perform a meta-regression to test whether there is a link between the mid-point year of a study and the estimated effect and we find a positive but insignificant relation. Although Lopez implies otherwise, we make clear the results of that meta-regression: "Although the slope was positive in both cases, suggesting a positive time trend, in neither case was the trend significantly different from zero. " (p. 47)

Misleading Statement #8

Lopez complains that our statement that "…we demonstrated that on average charter schools are serving students well, particularly in math." (p. 36) is overstated.
This is a factual statement (Table 1), with studies on average finding that charter schools are doing as well in reading and in most grade spans significantly better in math.

---

1. The grade spans for which charters significantly outperformed were elementary school, middle schools, a combination of elementary studies, middle school studies, and combined elementary/middle school studies, and studies that combined all grades. The only two grade spans that were not statistically significant were studies of high schools and studies that combined elementary/middle schools.

Misleading Statement #9

Lopez cites a partial sentence from our conclusion in support of her contention that we overstate the case, and yet it is she who overstates. The quotation: "….there is stronger evidence of [charter school] outperformance than underperformance, especially in math" (p. 53 – Lopez cites the wrong page). Sadly, Lopez resorts to chopping this sentence in half, while neglecting to show the balance in our reporting by also failing to quote the preceding sentence. Here is what we actually wrote in the first two sentences of the conclusion:

"The overall tenor of our results is that is that charter schools are in some cases outperforming traditional public schools in terms of students' reading and math achievement, and in other cases performing similarly or worse. But there is stronger evidence of outperformance than underperformance, especially in math".

A reasonable reader would conclude that we are being extraordinarily judicious in our summary of the results. Lopez neglects our attempts to present this balance and, again, cherry picks our statements to support her weak thesis.

The cherry picking continues with her assertions in the first full paragraph on page 3.

Misleading Statements #10, #11, #12, and #13

In the following paragraph on page 3, Lopez makes four misleading statements: "Results of the meta-analysis were presented for various sub-populations (e.g. subsamples of students disaggregated by ethnicity). Despite the fact that there were no differences in most of the comparisons and that the few that found charter schools outperforming traditional public schools had only modest effect sizes, the report nevertheless asserts that charter schools are outperforming traditional public schools." (She then cites several quotations from three **separate** sections of the paper.)

Misleading Statement #10

Lopez seriously distorts our work by comparing results from one set of analyses with our conclusions from another section, creating an apples and oranges problem. Lopez, having apparently read the paper, should know that the conclusions she cites below the above quotation are careful and accurate summaries from models of overall effectiveness that include all charter schools in a given grade range and *all students*, not small subgroups. She misleads the reader by juxtaposing results from small subgroup analyses and then implies that our conclusions stem from them.

Misleading Statement #11

Lopez claims that most of the results are not significant for subgroups. She neglects to report that this is because there are far fewer studies by individual race/ethnicity (for the race/ethnicity models virtually none for studies focused on elementary schools alone, middle schools alone, or high schools) or other subgroups. Further these studies have relatively smaller sample sizes, and therefore the chances that one would be able to detect an effect are lowered. This in **no way** contradicts the findings from the much broader literature that pools all students.

Misleading Statement #12

The claim by Lopez that most of the effects are insignificant in the subgroup analyses is incomplete in a way that misleads. She fails to mention that we conduct several separate analyses in this section, one for race/ethnicity, one for urban school settings, one for special education and one for English Learners.

Students in special education and students eligible for meal assistance both show positive and significant effects in reading in studies that combine all grades. For math *all three grade-spans* show positive and significant effects for students eligible for meal assistance, and students in special education show positive significant effects for studies that combine all grades. Moreover, we do not find negative and significant effects for any group/grade-span. Citing only one of our subgroup analyses, and precisely the one that showed the fewest significant effects, seems like cherry picking by Lopez.

She again cherry picks by failing to report on our other subgroup analysis for urban schools in which we find significant positive effects in reading for 2 of 6 grade spans and positive and significant effects on math for 5 out of 6 grade spans. In none of the grade spans is the effect negative.

For Lopez to claim that "there were no differences in most of the comparisons" is thus highly misleading.

Misleading Statement #13

The claim by Lopez that the subgroup effects that are significant are "modest" in size is misleading. The effect sizes vary considerably. In urban charter schools at the middle school level, we estimate an effect size for math of 0.167. This is enough to raise a student from the 50th percentile to almost the 57th percentile *in a single year.* Lopez, it seems, has an odd view of what a modest effect is. We return to this issue in our rebuttal to Misleading Statement #25 near the end of this document.

Misleading Statement #14

Lopez next conflates statements based on *different analyses* made 9 pages apart, citing them side by side in a way that an unwary reader might take to mean that we are making unwarranted conclusions about mean effects of charter schools on an analysis of the variance. She writes on page 3: "The report displays the varying magnitude of the effects found for each of the 52 studies and used a vote-counting method to determine the number of students that found significant effects in favor of each type of school 'to give a fuller picture of the distribution of effect sizes' (p. 36). The authors conclude that 'charter schools appear to be serving students well, and better in math than in reading' p. 47…."

These two statements should not be put side by side as a way of implying that our results on the *mean* effects are wrong. These are two completely separate analyses. Lopez seems incapable of understanding the distinction. Again, the two main findings of the paper are as follows:
1) Mean Effects: On average the mean effect of attending a charter school is positive and significant in math, and positive but insignificant in reading.
2) Variation in effects: There is considerable variation across locations in the relative effectiveness of charter schools compared to traditional public schools.

Both of these findings are important. (This is why we cite both in the conclusion.) Lopez seems unable to distinguish between mean effects and variance in those effects.

Misleading Statement #15

Continuing with the same badly written paragraph, Lopez insinuates that we exaggerate the positive overall math effect while downplaying the percentage of studies that show negative results. Here is her next misleading statement: "The authors conclude that 'charter schools appear to be serving students well, and better in math than in reading' (p. 47) even though the report finds '…that a substantial portion of studies that combine elementary and middle school students do find significantly negative results in both reading and math – 35 percent of reading estimates are significantly negative, and 40 percent of math estimates are significantly negative (p. 47)'"

Here is what we actually wrote: *"Examining all of these results as separate parts of a whole, we conclude that, overall, charter schools appear to be serving students well, and better in math than*

in reading. The caveat here is that a substantial portion of studies that combine elementary and middle school students do find significantly negative results in both reading and math—35 percent of reading estimates are significantly negative, and 40 percent of math estimates are significantly negative."

Lopez uses two rhetorical devices to lead readers to the perception that we overstated findings. First, she separates the two quotations, implying that we are somehow hiding the second result, when in fact we intentionally mention the positive mean math effect and the variation in the results across studies side by side.

Second, she further misleads the reader by again cutting out part of our sentence. Instead of stating that we have a "caveat" to the positive mean math effect she removes that entire clause.

What makes Lopez's approach even more misleading is that she fails to point out to her readers that in the paragraph above, we were bending over backwards to be fair. We cite only one type of study: those that combine elementary and middle schools. (These account for about 1/7th of all the studies.) Why did we focus only on those studies in the above quotation? Because these sorts of studies were the ones that produced *the highest percentage of studies with negative and significant estimates: 35%.* **Wouldn't one think that if our goal had been to overstate the positive effects of charter schools we would never have chosen to list the result that is the *least* favorable to charter schools in the text above?** We could have stated that for elementary school studies, only 12% showed negative and significant reading results, compared to 71% showing positive and significant results. Lopez completely fails to list any of the more positive results from the other 5 grade spans we study. The reason we made special mention of the negative combined elementary and middle results is because it was the exception to the rule, the positive results from the other 5 gradespans.

Again, the claims by Lopez that we give too positive an outlook on charter schools does not stand up to scrutiny. Meanwhile, she repeatedly quotes only the results that are the least favorable. It is Lopez who gives a misleading accounting of the results.

Misleading Statement #16

Lopez misleadingly includes a quotation from "the summary of the report" that was not written by us. While she has an endnote indicating that the summary is a stand alone document that was not written by us, the inclusion of this quotation without making it clear in the main text that we did not write it is sloppy.

Misleading Statement #17

On page 4 Lopez attacks our analysis of the very small literature on the impact of charter schools in some amusing ways, quibbling about whether our write-up constitutes "an analysis". We will let that odd statement pass, and focus on the following: She states that "… the section does *not* present findings from an analysis. Instead it extrapolates from selected studies to claim favorable outcomes for charter schools." We do not "extrapolate" from the existing studies. On the contrary, we are extremely cautious about our conclusions.

On the small literature on educational attainment, we conclude: *"The papers discussed above also present a smattering of other findings, with varying statistical significance.^{2} Again, the general picture that emerges is one suggestive of large positive impacts of charter schools on high school graduation and eventual college enrollment. It is important to note that this literature is still emerging,*

---

2. Angrist et al. (2013) find that lottery winners were more likely to pass the high school exit examination in Massachusetts, more likely to take an Advanced Placement (AP) exam, and scored higher on the AP Calculus exam. Lottery winners also scored higher on the SAT than lottery losers, but were not more likely to take the SAT. They were more likely to attend college overall (two- or four-year) but this effect is not significant. McClure et al. (2005) find that charter school attendees complete more college preparatory courses in high school.

and currently covers only a limited number of geographic locations.” (p. 52) Where exactly is there "extrapolation"?

The second summary we perform is of two studies that examine attendance and behavior. Our conclusion there reads as follows: *"This literature is obviously very small, but both papers find evidence that charter school attendance is associated with better noncognitive outcomes."* (p. 53)

Again we do not consider our words "extrapolation" when we state in our concluding sentence that there are only two papers.

Misleading Statement #18

Later on, Lopez again takes us to task for our review of the small literature on charter schools and educational attainment. On page 4 of her essay she makes a broad-brush (and unsupported) statement that "Throughout the report, the authors report findings asserting favorable outcomes for charter schools that are not well-supported in the literature." She provides only one example, again selectively quoting the above passage from page 52, ending with "*Again, the general picture that emerges is one suggestive of large positive impacts of charter schools on high school graduation and eventual college enrollment."* But she misleads the unwitting reader by failing to include the very next sentence: "*It is important to note that this literature is still emerging, and currently covers only a limited number of geographic locations."* Again, Lopez selectively quotes in a highly misleading way.

Misleading Statement #19

The top of page 5 quarrels with us about which studies we include and exclude from our main analysis, but yet again, in a highly misleading way. Lopez writes on page 5: *"The authors excluded studies focused exclusively on KIPP schools from the main meta-analysis (although results with KIPP schools are presented for comparison in the report) because the 'schools account for only about 2 percent of all charter schools' (p. 28. Nevertheless, CREDO studies, which were 12 of the 20 elementary/middle school studies, remained in the main analyses because 'they include extremely large samples of charter schools" (p. 7). The authors reported the CREDO studies, however, introduced biases that favored charter schools because of the matching approach used."* This paragraph is seriously off base for several reasons.

First, Lopez seems to fail to understand that the KIPP school studies produce very large positive estimated impacts. We did not include them in the analysis because they would have increased the estimated effect sizes markedly disproportionately to their relatively small share of the charter school universe. She fails to make this basic point. This omission is particularly noteworthy because Lopez accuses us, without merit, of overstating the positive impact of charter schools. Excluding the KIPP studies lowers the overall effects – another observer could argue that these studies should have been included.

Turning to her statement about CREDO studies, she makes two serious errors.

Misleading Statement #20

Lopez completely misunderstands our analysis of the likely sources of bias in the CREDO studies. She claims: "*The authors reported the CREDO studies, however, introduced biases that favored charter schools because of the matching approach used."*   In fact, we did not say that the CREDO estimates introduced favorable bias. On pages 29-30 we explain that we believe that the CREDO charter estimates "if anything, would be biased toward zero". This does *not* mean that the CREDO studies are biased in favor of charter schools at all. While in cases of a true charter effect being negative, a bias towards zero implies an estimate that is less negative than the true effect, in the cases of a true charter school effect being positive, the bias is actually *downwards* – the CREDO estimate may be too small.

Misleading Statement #21

As noted above, Lopez complains that "*Nevertheless, CREDO studies, which were 12 of the 20 elementary/middle school studies, remained in the main analyses because 'they include extremely large samples of charter schools" (p. 7)."* While technically true, Lopez fails to mention to her readers that in Table 4 we re-do *all* of our main analyses after excluding the CREDO studies. The results are quite similar. This is a serious and careless omission by Lopez.

Next, Lopez writes the following sentences in her section on "The Report's Rationale for its Findings and Conclusions" on page 4 of her essay: "The authors only included lottery-based and value-added modeling studies in the primary analyses, contending these 'represent the best methods available' (p. 8). The primary rationale is that a simple tally of conclusions based on positive and negative results accurately and adequately represents the universe of findings without regard to study size, scope, or significance. In the secondary findings, selected narrative reporting is deemed to be valid, although no rationale for inclusion or exclusion is provided."

We find these sentences unclear in their logical structure, but we respond to our best guesses of her intentions here.

Misleading Statement #22

It is simply untrue that we believe that "a simple tally of conclusions based on positive and negative results accurately and adequately represents the universe of findings without regard to study size, scope, or significance." At this point we have serious concerns about whether Professor Lopez understood the statistical analysis in our report. The main analysis is not a "simple tally". Rather it is a meta-analysis, and it does indeed weight studies differently based on the level of statistical precision of each study. We are astounded that Lopez does not seem to understand this basic point. After all, we summarize the method clearly on page 9 and in the appendix to the paper. We do use two other subsidiary methods, each designed to show in a less technical way our major finding that results vary by location considerably. One shows histograms, but these are weighted using the meta-analysis weights, and so they too take into account differences in "significance" across studies. The other is indeed a "simple tally", a counting of how many studies are positive and significant, negative and significant and so on. But Lopez fails to mention that even in this simplest of methods we show the results when we weight the studies by their precision, as calculated in our meta-analysis. (See columns 2 in Tables 8 and 9.) Again, Lopez makes unsupportable statements.

On page 5 Lopez makes strident and off-target criticisms of lottery-based studies, which social science views as the gold standard.

Misleading Statement #23

Lopez frets that charter schools are allowed under federal law to use weighted lotteries to allow minorities a greater chance of attending a charter schools. She fails to document which charter schools actually use weighted lotteries. More to the point, she fails to identify whether any of the lottery-based studies included in our report included schools with weighted lotteries, and at the same time failed to adjust statistically for differential weighting. We would need both of these conditions to apply for her claim of bias to have any merit. We challenge her to bring any such errors to the attention of the authors of the various lottery-based studies.

Further, Lopez does not seem to understand that her concern that some researchers may have ignored the possibility that charters are legally allowed to weight lotteries in favor of minorities likely leads to a downward bias in the estimates. If a researcher failed to adjust for weighting, it would mean that lottery-based studies would probably produce effect sizes that were biased downwards, the opposite bias to what Lopez implies.

Misleading Statement #24

On page 6 Lopez argues that it is "problematic" that we would make conclusions when each of the methods used by researchers have potential statistical disadvantages. This is an odd stance to take as it essentially implies that social scientists should stop all research. But it is the follow-up statement that is more troubling: "This is particularly problematic given that the authors found both lottery-based and propensity score matching studies to be significantly related to the effect size in the meta-analysis for mathematics, which thereby interjects systemic bias in the analysis." This statement is illogical. What we found is a significant difference between the two methods she lists and the comparison group method, which primarily consisted of studies by CREDO. Earlier, Lopez seemed to take our concerns about the CREDO studies seriously, even if she did not understand them. What is to say that gold standard lottery-based studies and propensity score studies are necessarily biased in a positive direction? Could it not be that the comparison group analyses, mainly consisting of CREDO studies, are sometimes biased in the other direction?

Misleading Statement #25

On pages 6-7 of her essay, Lopez criticizes us for re-stating effect sizes in terms of how a median student at the 50th percentile would change his or her ranking within a district. Her insinuations continue. We claim that presenting the effect sizes in terms of percentiles, which any person who has interpreted their performance on a standardized test will understand, is more "transparent" to most people that discussing only "effect sizes". Lopez claims that this explanation is "misleading". Lopez is seemingly unaware that this has become a fairly common practice in the education research literature.

Lopez claims that if a student was far below the 50th percentile, the gains would be smaller. While technically true, the differences are minor. Further, she ignores the fact that as one moves up closer to the median student over time, the gains will become bigger. Consider her example of an effect size of 0.08 (a bit below our estimated effect size of 0.084 for students attending charter middle schools in math). She states that if a student started at the 25th percentile rather than the 50th percentile her gain would be "only" 2.3 percentile points rather than 3.3 points.

To see how much of a problem this is, let's compare predicted math gains for a student who attends a charter school from kindergarten through grade 5, and then a charter middle school from grades 6 through 8, and let's use the average predicted math effects (both significant) of 0.045 and 0.084 respectively (taken from Table 1 of our paper). We will compare the total advantage enjoyed by the student who attends the charter schools relative to an otherwise similar student who attends traditional public schools throughout K-8. Over 9 years, the charter school student is predicted to gain, relative to the non-charter student, $6*0.045 + 3*0.084 = 0.52$ standard deviations. Below is the predicted beginning and end percentile rankings for two charter school students, one who starts at the 25th percentile and another at the 50th percentile. And recall that for both we have comparison students at traditional public schools who remain at the 25th and 50th percentiles between kindergarten and grade 8.

Table 1. Predicted Gain for Two Charter School Student Who Attended a Charter School between Kindergarten and Grade 8, Starting from the 25th and 50th Percentiles

Student	Initial Percentile (kindergarten)	Final Percentile (End of Grade 8)	Gain
A	25	43.9	18.9
B	50	69.8	19.8
Differences	25	25.9	0.9

After 9 years, the gap in predicted gains between students starting at the 25th and 50th percentiles is 0.9 percentile points. One student gains 18.9 percentile points, the other, 19.8 percentile points. Thus, Lopez's accusation that interpreting the effect sizes at the 50th percentile only substantially distorts our results is incorrect.

Lopez slips into the habit of labeling the size of estimated charter school effects as "small" and then using these labels against us. She is perhaps unaware that the U.S. Department of Education will not allow authors to state that an effect size is large or small. We think there is good reason for this policy. Lopez makes questionable claims that effects are small. Worse, she then inexplicably uses this as a criticism of us as researchers. We reported the effect sizes objectively and transparently and for the most part have avoided use of adjectives such as "big" or "small". Any interpretation of the effects as small or large is subjective, and while perhaps worthy of discussion, such interpretation does not change the fact that the effect sizes for math are positive and significant for four of the six grade span categories we studied. We believe that the sort of gains over elementary and middle school calculated above would not be viewed as "small" by a reasonable person.

Misleading Statement #26

On page 7 Lopez cites Shin et al. (2013) who find evidence of decreasing gains in achievement across grades. She argues, without specifics, that if gains are non-linear it would affect our estimates. There are three errors that Lopez makes in this claim. First, and most important, she does not seem to realize that most charter school achievement studies do not model gains in achievement, but rather changes in students' relative standing in the test-score distribution. Her statement is completely irrelevant for such studies. But what about the relatively small number of studies that do use vertically scaled scores? Her point is misguided even for these studies due to two additional errors. First, she fails to make a case as to why a researcher who estimated average achievement gains between, say, grades 3 and 5, would obtain a biased estimate. Second she fails to point out that many papers in the literature avoid comparing large numbers of grades by doing separate analyses for elementary, middle and high schools.

Unfortunately, this first error by Lopez is a particularly serious technical mistake: she is citing a paper that uses a vertically scaled test to infer absolute degrees of learning across grades. Yet the vast majority of charter school studies use the extant state tests, which are *not vertically scaled*. Rather, researchers are using Z scores, which are set to zero with a variance of one within each grade. As such they measure students' relative rank in achievement. Because this format of test scores always produces a mean test score of zero in each grade, the concerns by Lopez that gains in achievement are larger in lower grades is completely incorrect. With Z scores, the average "gain" in achievement from one grade to the next is precisely 0, meaning that on average students do not change their relative position in the test score distribution. It is nonsense to claim that gains are higher in the lower grades. On the contrary, the average gain in all grades are identical: zero. That Lopez would confuse findings from a study with vertically scaled scores with Z scores, and then suggest that gains in Z scores from grade to grade must be decreasing in higher grades is completely incorrect. Again, she makes a serious technical mistake.