



School of Education, University of Colorado at Boulder
Boulder, CO 80309-0249
Telephone: 802-383-0058

NEPC@colorado.edu
<http://nepc.colorado.edu>

DUE DILIGENCE AND THE EVALUATION OF TEACHERS

FACT SHEET CONCERNING L.A. TIMES' POSTS OF FEBRUARY 14, 2011

This fact sheet was prepared in response to February 14 statements by the *Los Angeles Times* concerning the National Education Policy Center (NEPC) report called *Due Diligence and the Evaluation of Teachers*.

The *Times'* responses to the NEPC's original report and to our efforts to correct the newspaper's subsequent reporting have repeatedly misrepresented and distorted the facts. Specifically, the newspaper's so-called "readers' representative" has made demonstrably false claims about what was included and not included in the NEPC report. We discuss those false claims, as well as other misrepresentations and distortions, below.

What is at stake here is not a battle over semantics or arcane statistical details. The *Times* contends that the teacher effectiveness ratings it published online were built on sound research, offering a fair and reliable assessment of the relative quality of individual teachers in the Los Angeles Unified School District. Parents are encouraged to rely on its searchable database in order to make choices about who will teach their children.

The NEPC report explains that the model used to construct the *Times'* teacher database is not adequate to that task. Using a stronger, alternative model, 53.6% of the teachers in the database would fall into a different effectiveness category for reading than the one assigned by the *Times*. While the NEPC researchers explain why they think a stronger model is preferable, that's not really the point. Instead, the point is this: when two reasonable models reach such different results, the *Times'* decision to publish ratings based on their preferred model is reckless.

Background

In an apparent pre-emptive response to the imminent (February 8, 2011) release of the NEPC *Due Diligence* report the *Times* published on February 7 a story under the headline: “Separate study confirms many Los Angeles Times findings on teacher effectiveness,” and with the subtitle, “A University of Colorado review of Los Angeles Unified teacher effectiveness also raises some questions about the precision of ratings as reported in The Times.”

The *Times* headline and its reporting were so misleading that in response NEPC posted a “Fact Sheet” on its website (<http://nepc.colorado.edu/publication/due-diligence>). That first Fact Sheet walks readers through the most misleading and false statements in the *Times’* February 7 article.

However, the *Times* has essentially “doubled down” on its reporting, defending the accuracy of its teacher ratings as well as its coverage of the NEPC report. The controversy over the veracity of the *Times’* reporting and its justification for publishing the names and “effectiveness” ratings of 6,000 LAUSD teachers has continued, and the *Times* on February 14 published a post by its “readers’ representative” and a statement from the management defending its reporting (<http://latimesblogs.latimes.com/readers/2011/02/times-responds-to-criticism-of-teacher-analysis.html> and <http://latimesblogs.latimes.com/readers/2011/02/los-angeles-times-stands-by-its-teacher-ratings.html>)

This, our second Fact Sheet, is in response to these two posts from the *Times*. The *Times’* posts misrepresent the NEPC research and its implications for the *Times* teacher rating database.

First Contention Made by the *Times*

In her post, the *Times* readers’ representative, Deirdre Edgar, notes that *Times* readers were concerned about two issues: “Did the article accurately reflect the findings of the study? Does the study invalidate the ‘Grading the Teachers’ series?’

On the question of whether or not the *Times* assertion that the NEPC analysis “confirms the broad conclusions” of the *Times* teacher rating research, Ms. Edgar quotes *Times* assistant managing editor, David Lauter, “For parents and others concerned about this issue, that’s the most significant finding: the quality of teachers matters. So although they disagree with us about how to measure the teacher effect, it was entirely accurate to say that their study confirmed some parts of our work and criticized others.”

Fact: The NEPC report unambiguously states that the analysis used by the *Times* cannot produce accurate teacher ratings. The only confirmation found in the NEPC study is of so-called variation. That is, the *Times* model does in fact produce different numbers for different teachers. But this finding does not in any way address whether or not the numbers have any validity or usefulness.

There has been ample opportunity for Dr. Buddin, the author of the White Paper on which the *Times* relies, to contest the NEPC report's conclusions regarding validity. To our knowledge, he has not. Nor has the *Times* produced any expert on value-added assessment to dispute this point.

To be clear, the NEPC report found that the *Times'* teacher ratings are too inaccurate to be useful for readers wanting to find out a given teacher's actual effectiveness.

Comment: Mr. Lauter appears to frame the issue as a technical argument between experts. He contends that because one value-added model produces one set of rankings and another such model produces a different set, who is to say which expert is right? Mr. Lauter confuses the issue. It is precisely **because** the ratings are so unstable and so sensitive to different model choices that the *Times* should never have published individual rankings of teachers.

The core issue is that the *Times* cannot honestly warrant to anyone who uses its database that any given teacher rating is valid. In fact, there is a very good chance that a teacher's rating *would change* if a different, reasonable model were used. Since the *Times* cannot make this warrant, its ratings have a high potential to mislead parents or anyone who wishes to use them to determine a teacher's effectiveness. This is important as a matter of sound policy, but it is also an issue of journalistic integrity.

Second Contention Made by the *Times*

Ms. Edgar comments, "It's important to note that the Colorado study was not based on entirely the same data Buddin used."

Fact: The NEPC obtained its dataset through a California Public Records Act request of the LAUSD. The PRA request specified that the district was to provide the exact same dataset provided to the *Times* and used by Dr. Buddin. We have no reason to believe that the LAUSD failed to comply. The NEPC researchers then used the Buddin White Paper to guide their attempt to replicate his research. Whenever they ran into a question or were unable to replicate something, they attempted to reach Dr. Buddin for clarification and received only limited responses.

Ultimately, the numbers never matched, which is certainly a reason for concern. Replicability of research is a core element of scientific practice; when research cannot be replicated, it calls into doubt the original study and findings. To be absolutely clear, this is not to accuse Dr. Buddin of any wrongdoing. Rather, the non-replicability is a red flag and signals the need for further investigation to better understand the data and the process used to reach the results he reports.

In any case, it is reasonable to begin with the assumption that the same dataset was used by both sets of researchers. The next step was to determine which teachers and which students could be included in the analyses. For instance, if a teacher appears only once in the dataset

and teaches only 10 students, she cannot reasonably be included. The NEPC researchers followed all guidance available to them in the Buddin White Paper, attempting to use exactly the same data as did he.

This is rhetorical jujitsu: the inability to replicate the *Times* analysis is now used by the *Times* as an argument to defend its analysis. In any case, the two sets of analyses did appear to use very similar data, and neither the *Times* nor Dr. Buddin have offered, to our knowledge, any evidence that key findings of the NEPC study would be different had the exact same data been used in both analyses.

Third Contention Made by the *Times*

The statements by the *Times* readers' representative and *Times* management both make note of the fact that NEPC has received funding from teachers and their unions.

Here is Ms. Edgar: "It also is worth noting that the policy center is partly funded by the Great Lakes Center for Education Research and Practice, which is run by the top officials of several Midwestern teachers unions and supported by the National Education Assn., the largest teachers union in the country and a vociferous critic of value-added analysis."

The *Times* management puts it this way in its post: "Finally, a major source of funds for the policy center is the Great Lakes Center for Education Research and Practice, a foundation set up by the National Education Association and six major Midwestern teacher unions affiliates. The NEA was one of the teacher union groups that backed an unsuccessful call for a boycott of The Times when 'Grading the Teachers' was first published."

Fact: This work was in fact funded in part by the Great Lakes Center; a funding credit to that effect is clearly stated on the published report. It is also a fact that no NEPC funder has ever had an editorial say in our research. The content of this report as well as the choice of expert authors are far beyond the reach of our funders. Again, no funder has any role in the NEPC editorial process, no funder sees NEPC documents until they are publication-ready, and no funder has any say on whether a document will be published.

Comment: It is peculiar that the *Times* fails to see that commissioning a study and then vehemently defending it when it is found to be flawed represents a conflict of interest of a far greater magnitude than that suggested by the line of attack directed at us. Is it appropriate for a newspaper to create news, defend that work in a "news" story that misrepresents a critique, subsequently attack those offering the critique, and then represent its own account as the work of unbiased reporters?

Fourth Contention Made by the *Times*

Ms. Edgar offers the following criticisms of the NEPC analysis and of NEPC: “The Colorado researchers did not contact [the] *Times* before preparing their report in order to compare data sets, and they have not explained the reasons for the dropped data. They also have not disclosed basic information about some of their statistical techniques that would allow outside researchers to assess their work.”

Fact #1: Some of this is addressed above (see the discussion of the second contention). But there’s a bit more here. The focus of the NEPC reanalysis was Dr. Buddin’s White Paper, the research report on which the *Times* relies as the foundation of its reporting. And Professor Briggs, the NEPC lead author, did attempt to address inconsistencies and questions with him.

Fact #2: The claim that basic information about the statistical techniques used by NEPC researchers is missing from our report is flatly false.

The report is detailed, the appendix provides even greater details, and the researchers even offered the following: “In an effort to be as transparent as possible about what we have done and any mistakes we may have made, the script files that were used to generate our results are all available upon request.” (See endnote 27 of page 30 of the report. The “script files” essentially contain the computer programming language.)

Comment: Such public disclosure and openness about decisions (and scripts), exposing important work to public scrutiny and to expert replication, is a foundation of scientific inquiry. We call upon the *Times* and Dr. Buddin to further this inquiry and allow this process to continue by releasing complete information about their data choices as well as their script files. In addition, the *Times* has represented in correspondence with Professor Derek Briggs, lead author of the NEPC report, that it conducted sensitivity analyses similar to those set forth in the NEPC report. If this is the case, disclosure of details of that work would greatly add to the public’s understanding of any due diligence conducted by the paper.

Fifth Contention Made by the *Times*

Ms. Edgar writes, “The researchers based their work on a pool of roughly 11,000 teachers. But The Times published scores for only 6,000 teachers because the project excluded scores from any teacher who had not taught at least 60 students. When informed of that discrepancy over the weekend, Derek Briggs, the lead author of the Colorado study, said in an e-mail to The Times that the use of the 60-student limit “**serves to mitigate**” [emphasis added] some of the shortcomings his study had alleged. Briggs and his colleagues, however, have not made that information public.”

Ms. Edgar’s statement contains both a misrepresentation and a factual error.

Fact #1: Here is the full context, from a February 5 email to Jason Felch from Professor Briggs: “I agree that use of N>60 criterion by the L.A. Times **serves to mitigate** [emphasis added] the issue of false positive and false negatives. But I stand by our statement in the executive summary of the report that ‘it is likely that there are a significant number of false positives (teachers rated as effective who are really average), and false negatives (teachers rated as ineffective who are really average) in the L.A. Times’ rating system.’ The fundamental point we were making is that the application of a 95% confidence interval to group teachers into three categories will be more conservative than a quintile approach that groups teachers into five. This remains true with or without the N>60 restriction.”

It appears that Ms Edgar has pulled these three words out of context to misrepresent Professor Briggs’ meaning to the *Times*’ readers.

Fact #2: Professor Briggs’ comment that the *Times*’ 60-student limit mitigates the issue of false positives and false negatives *is* provided to the public – in the NEPC report itself. Find it on page 31, endnote 38. In addition, the report’s Executive Summary clearly states, “Using the *Times*’ approach of including only teachers with 60 or more students, there was likely a misclassification of approximately 22% (for reading) and 14% (for math).”

This is a crucial point. The NEPC report and Professor Briggs were responsive to the concerns raised by Mr. Felch and, prior to the report’s release, we addressed those concerns in a prominent (executive summary) and detailed (endnote 38) manner. The full email from Professor Briggs to Mr. Felch was also linked to in the first Fact Sheet and is available on the NEPC website here: <http://nepc.colorado.edu/files/BriggsFelchemail.pdf>

A week later, Ms. Edgar and the *Times* editors falsely accused Briggs and the NEPC of failing to “[make] that information public,” and in the process incorrectly represented the report’s contents.

Fact #3: The claim itself is largely a distraction.

The NEPC authors were replicating the Buddin White Paper’s analysis, which is why the 60-student limit was not originally imposed on the data. But this issue affects only a small part of the NEPC report’s analysis – the so-called “precision” analysis. It has no effect whatsoever on the accuracy (validity) of the analysis that lies at the heart of the NEPC report. This has been explained to representatives of the *Times*, yet we still see repeated attempts by *Times* to muddy the waters with this claim.

Accuracy (validity) and precision (reliability) are very different things. If a doctor uses an eye chart to diagnose a broken leg, the patient may provide precise answers, reading the eye chart perfectly again and again. But these results cannot be validly used to determine whether the patient’s leg is broken. The *Times*’ complaint is that the embargoed report failed to note that the paper had taken a step to mitigate concerns about the precision (reliability) of their analysis. The NEPC report notes and responds to that complaint. But, the important point is that even if Dr. Buddin’s value-added model produced precisely the same ranking score for a given teacher

over and over again, that does not mean that it is capable of producing valid teacher effectiveness ratings.

Conclusion

The *Times* has not been simply reporting on teacher evaluations or ratings. It has been *creating* them and publicizing them. This unusual position confers upon the *Times* a profound obligation to ensure that any ratings it publishes are both valid and reliable. It is incumbent on the paper's reporters and editors to cautiously report on the effort's weaknesses.

Moreover, this ethical obligation is amplified when the *Times* is presented with a critique of the social science work that the paper had commissioned and used. Yet inexplicably the story about the critique was assigned to the same reporter who wrote and has repeatedly defended the original story, and this assignment was apparently made by the same editor who worked on the original story. The result, not surprisingly, was an attempt to mislead readers and whitewash the critique.

As the two NEPC fact sheets document, this is not a matter of experts merely disagreeing. This is a matter of mistakes in judgment and in fact. We call upon the *Times* to stop trying to defend the indefensible, pull down its invalid teacher ratings, and set about the difficult business of getting its story right.

The first NEPC Fact Sheet is available at: <http://nepc.colorado.edu/files/FactSheet.pdf>