Focusing on Fundamentals: A Reply to Koski and Horng

Sarah F. Anzia and Terry M. Moe

March 2014

Koski and Horng argue that our study of seniority-based transfer rules is narrow and that its findings only apply under limited circumstances. They say that their own study, by contrast, takes a broad frame—so broad that, as detailed in their 127-page technical report, they estimated 40 different models. They argue that the bigger picture that emerges from their more extensive analysis is that there is little evidence for our finding that these rules have a negative impact on the percentage of experienced teachers in disadvantaged schools relative to advantaged schools.

As a starting point, we should note that Koski and Horng's published article does not mention this technical report, its 40 models, or its broad frame. The article simply presents an empirical analysis of seniority-based transfer rules and concludes that there is no impact. In our view, the article is properly assessed based on what it actually says and does. But given that Koski and Horng (in their response) have directed attention to the technical report, we will address their claim that the report provides the bigger picture and bolsters the article's conclusion of no impact.

The technical report shares the same weaknesses as the article. Consider Table 1 of their response, which lays out the results of all 40 of the report's models. These models are mainly distinguished by different measures of transfer rights (TLST and TLS1) and different samples on which the estimations were carried out. For reasons that we discussed in our own article, both the measures and the sampling criteria are problematic.

First, consider TLST, their "total" measure of seniority-based transfers: this measure is based on six items, some of which *are not measures of the role of seniority in transfers*. There is no theoretical justification for this, and it makes TLST a weak index of the key independent variable. Their TLS1 variable is a weak measure too, because it is based on just one coding item (voluntary transfers). Koski and Horng say that this one-item measure reflects the argument of the original Moe study and the newer Anzia-Moe study as well, but we have never argued for such a measure. We have consistently argued that the best measure is one that combines seniority-based measures of voluntary and involuntary transfers—and in both the Moe and Anzia-Moe studies, those are the measures used.

Now consider the five samples that distinguish the various technical report models in Table 1. As we discussed in our article, it is important *not* to combine elementary, middle, and high schools in the same sample, because teachers typically do not transfer from one type of school to another. That is why we focused our own analysis on elementary schools—plus the fact that these schools are by far the most numerous type within districts. Koski and Horng agree that teachers rarely transfer across different school types, and they exclude high schools from their main sample (Sample A) for this reason. Yet Sample A still includes both elementary and middle schools; it also includes charter schools, which are not covered by district labor contracts. Sample B is the same, except it omits Los Angeles (and charter schools). Sample C focuses just on elementary school districts; but these districts actually contain middle schools as well as elementary schools. A focus on these districts also excludes a large number of elementary schools operating in unified districts; a huge amount of data is thrown away. Sample D includes all unified districts except Los Angeles—and therefore pools different types of schools. Sample E focuses just on high school districts; but some of these districts contain middle schools, and in any event, many of the state's high schools are to be found in unified districts—meaning that, again, a large amount of relevant data is thrown away.

What this brief discussion illustrates is that Koski and Horng's broad frame is essentially a proliferation of models that are not well justified on analytic grounds: every one of their samples is problematic, and each model is based on a weak measure of seniority-based transfers. There is no bigger picture here. And the many insignificant coefficients shown in Table 1 do not serve to somehow justify or bolster the no-impact conclusion of their article.

If the published article is assessed for what it actually says and does, moreover, its no-impact conclusion is unwarranted. As we've argued, their data analysis was based on problematic measures and sampling criteria; but within the context of that analysis, they failed to recognize that their findings actually supported, at a high level of statistical confidence, the hypothesis at issue: that seniority-based transfer rules have a negative impact on the proportion of experienced teachers in disadvantaged schools relative to advantaged schools.

This was not apparent to readers, because the authors presented their findings with no standard errors or t-scores, only asterisks to indicate significance (or not) based on two-tailed tests. We think it is straightforward that, because the hypothesis being tested asserts a negative relationship, the appropriate test is one-tailed rather than two-tailed. But that aside, providing standard errors or t-scores is absolutely essential in any statistical analysis, so that readers can judge for themselves how much confidence to have in the estimates. The standard errors from Koski and Horng's analysis—which we estimated ourselves—show a high level of statistical support for the hypothesis of negative impact.

In their response, Koski and Horng essentially argue that significance tests should always be two-tailed, never one-tailed, regardless of the content of the hypothesis—which doesn't square with the statistical theory of hypothesis testing. They also argue that if a coefficient fails to meet a predetermined level of significance, then the automatic conclusion is one of no impact—end of story—even if the standard errors indicate that the null hypothesis can be rejected at a very high level of confidence. As they see it, judging coefficients is a strict matter of up or down, not a matter of determining what the estimation really has to say about statistical confidence. Most important, Koski and Horng reject the notion that they should present the standard errors or t-scores of their estimates. This is a position that, in our view, cannot be justified. Yet it is reflected once again in Table 1 of their response, which provides readers with the estimated coefficients from 40 models—with no standard errors or t-scores.

In writing our own article, we sought to highlight these concerns with Koski and Horng's article, but criticism was not our purpose. Our purpose was to be constructive in helping to build and encourage a new literature. We aimed to bring evidence to bear on the impact of seniority-based transfer rules—and in the process, to clarify the key theoretical and methodological issues involved in carrying out this kind of research, to explain what decisions on these issues seem to make good analytic sense, and to try to move the literature forward on a firm analytic footing. Koski and Horng criticize these efforts as leading to a narrow analysis. But the decisions we made were specifically designed to mitigate basic problems, and in so doing to allow for better models, measures, and tests.

Details aside, three of these analytic decisions stand out. First, we restricted our sample to elementary schools for reasons we've just discussed. Second, we declined to use the Koski-Horng transfer measures—for reasons we've also discussed—and we built our own analysis around a composite index that, because it captures both voluntary and involuntary transfers and because it is only based on measures that explicitly code the role of seniority, is better suited to the job. Note that these two decisions on samples and measures are efforts to improve upon the Koski-Horng analysis, given their data set and coding scheme (within which we were operating)—but they are hardly permanent. In future work, scholars can surely carry out analyses on samples of middle schools and samples of high schools; and they can create other transfer measures based on new data sets and coding schemes. We are not wedded to the specifics of our own analysis. Our emphasis is on the criteria and logic underlying them—and on mitigating problems.

Our third decision was to make teacher experience the dependent variable, and not to include teacher credentials as well. Our reasons are straightforward. Seniority-based transfer rules are explicitly about experience. They are not about credentials. So if these rules do have an impact, it should show up most directly in how experienced teachers get distributed across schools—and exploring this connection should be the initial focus of research. We can only emphasize, moreover, that this is a very new literature, and already there is a dispute about whether these rules have any impact at all. The way to settle it is to put the focus on fundamentals. Experience is fundamental. Credentials are not.

Attention to fundamentals sheds a still brighter light on Koski and Horng's own findings. They make much of the fact, for example, that 29 of their 40 technical report models yield insignificant impacts (by their calculations and decision rules), but they don't point out that 20 of the 29 insignificant coefficients arose in their credentials models (with 20 out of 20 producing insignificant coefficients)—whereas a majority of coefficients in their experience models were actually significant. The difference is quite striking and of obvious theoretical interest, yet Koski and Horng do not recognize or discuss it in explaining what their statistical findings mean.

We agree with Koski and Horng that the larger goal is to understand how these seniority rules might affect teacher quality. But this is just another reason not to put the focus on credentials. Basic teaching credentials are only tenuously connected to quality, at best, and cannot be regarded as a well-supported proxy. Experience, on the other hand, is not only fundamental to the theory, but also a better indicator of quality. The evidence shows that teachers in their first year or two on the job do not perform as well, on average, as their more experienced colleagues. Experience is properly the main focus of analysis.

Much remains to be done in exploring the connection between seniority rules and teacher quality, and more generally, in determining how contract rules affect the behavior of teachers, the organization of schools, and the quality of government service provision. Our own work is just the beginning of what we hope will prove a dynamic program of research going forward, involving many scholars. Our core theme, both here and in the article, is less about specific findings than about that process. It is about the importance of thinking clearly, systematically, and rigorously about the fundamentals that need to undergird this line of research. And it is about encouraging progress based on analytic focus and coherence.

An Addendum on Cohen-Vogel, Feng, and Osborne-Lampkin (2013)

As our article was being copy edited, we learned that a new article on seniority-based transfer rules by Cohen-Vogel, Feng, and Osborne-Lampkin (2013) had just been published in this same journal. This new article cites our own and was written in knowledge of its contents, because we sent it to the authors long ago. We did not know about their paper, however, until it was actually published. Had we known, we would have included it in our assessment of the literature. The best we can do here is to offer brief comments.

The thrust of their article is that, based on an analysis of Florida data, the Koski-Horng conclusion of no impact is confirmed. We don't have access to their data, and it is possible that, for Florida, seniority-based transfer rules don't have any effect—although that would surprise us. But the greater surprise is that the authors cited our paper and yet ignored its arguments and evidence. We go to great lengths, for example, to explain that the Koski-Horng data analysis in their published article actually *supports* the hypothesis that seniority-based transfer rules have an impact on the distribution of experienced teachers across advantaged and disadvantaged schools. This argument about the proper interpretation of Koski and Horng's findings is of great relevance to Cohen-Vogel et al.'s work and how they think it squares with the rest of the literature. If they disagree with our argument, they should say so and explain why. That they simply ignore it is yet another source of confusion in this body of research, and a missed opportunity to move the literature forward.

Empirically, the key finding of our own analysis is that the impacts of seniority-based transfer rules are substantial in large school districts but near zero in small districts. Moe (2009) also found dramatic differences across large and small districts in studying the impact of contract "restrictiveness" on student achievement. There are good theoretical reasons, moreover, for expecting such differences: formal rules are much more likely to be followed in large bureaucratic settings. Yet Cohen-Vogel et al. don't mention these findings—and they don't consider the possibility that their own findings might also be very different in large and small districts. Here again is a missed opportunity for moving this literature forward.

We have various concerns about the details of their study, but due to space restrictions we can only highlight a few basic issues. First, their analysis suffers from the same problems of measurement and sampling as Koski and Horng's does. Their models are based on the Koski-Horng transfer variables, TLST and TLS1, which are weak measures for reasons that we've addressed—but that the authors don't discuss. They also don't discuss sampling issues. Their analysis is based on a sample that appears to simply pool together all types of Florida schools: elementary schools, middle schools, high schools, charter schools, and more. This approach is highly problematic. It is compounded by the fact that, in analyzing these schools' teachers, they included part-time teachers—who are not affected by seniority based transfers in the way full time teachers are.

Second, they frame their analysis as a test of the Moe and Koski-Horng models on Florida data, but their own model is in fact very different. Among other things, they do away with the "minority" variable so central to the earlier models and replace it with three variables—black, Hispanic, free-and-reduced-price lunch—that are each interacted with transfer rules. They also introduce a vast number of new control variables—including one for English language learners, which is doubtless highly correlated with percent Hispanic. The resulting model is exceedingly more complex than the originals. This complexity greatly magnifies the likelihood that coefficient estimates are weakened by multicollinearity problems; it imposes enormous new burdens on both the HLM (especially) and fixed-effects models in arriving at valid estimates; and it hobbles any efforts to relate this analysis to the earlier ones.

Third, as might be expected, Cohen-Vogel et al.'s complex model mainly generates insignificant coefficients for the full range of variables—and when it does produce significant coefficients, their meaning and import are theoretically unclear. One finding is that "the teacher experience and certification gap between schools with higher and lower percentages of Hispanic students may be smaller" when seniority based transfer rights are stronger (p.339). Yet transfer rights don't have the same effect for African-American students or free-and-reduced-price lunch students. What theory could possibly explain these results? Here and in other cases, the authors do not attempt to make theoretical sense of their findings. They simply report them. But theory is of the essence in all empirical work—and when findings are incongruous and don't add up to a coherent theoretical whole, that is a red flag that something is amiss. In this case, we have reason to think that is so.

Our bottom line is that we have concerns about the Cohen-Vogel et al. analysis, and we don't know what the Florida evidence would actually show given a more appropriate research design. That said, we take their interest in the topic of seniority based transfers as a very positive development. We hope that many more scholars will be drawn to this line of research—and more generally, to the study of collective bargaining and its impacts on America's schools.

References

Cohen-Vogel, L., Feng., L., and Osborne-Lampkin, L. (2013). Seniority provisions in collective bargaining agreements and the 'teacher quality gap.'" *Educational Evaluation and Policy Analysis*, 20(1), 1-20.

Moe, T.M. (2009). Collective bargaining and the performance of the public schools. *American Journal of Political Science*, 53(1), 156-174.