CALDER
#CALDER2014

The Efficiency Implications of Using Proportional Evaluations to Shape the Teaching Workforce

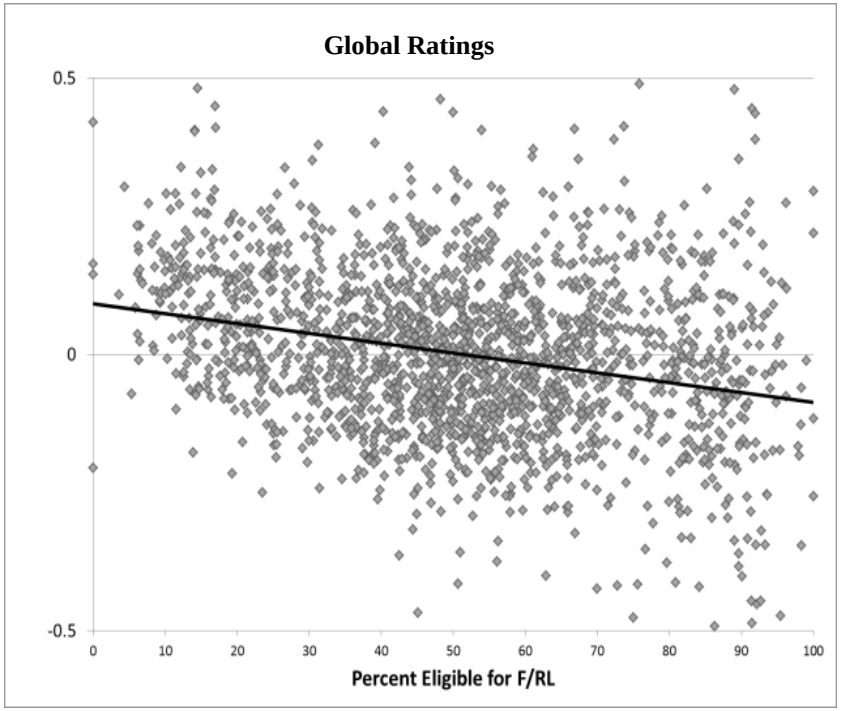
Cory Koedel & Jiaxi Li
University of Missouri

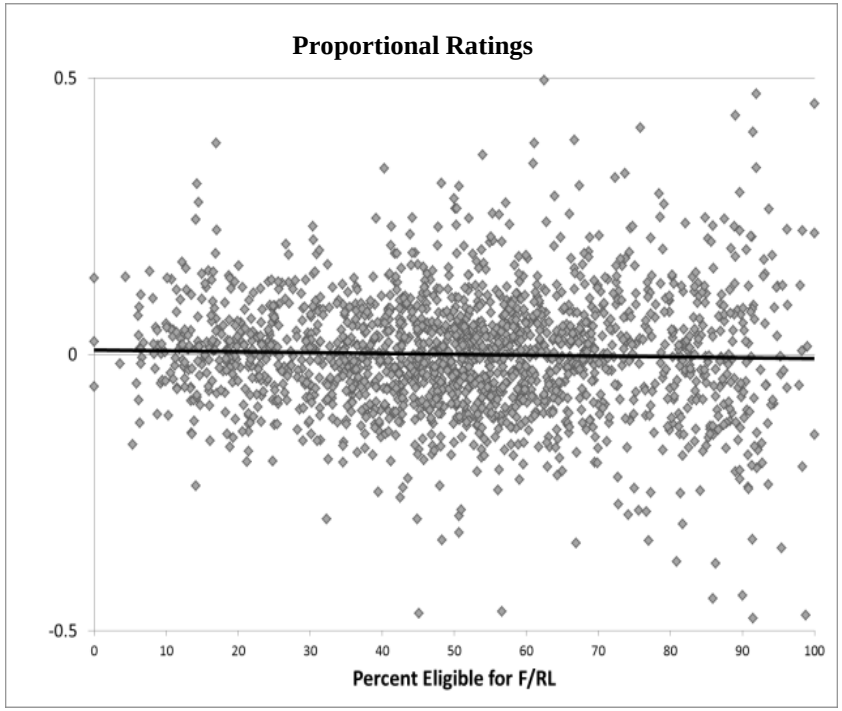
Background and Motivation

CALDER
#CALDER2014

- "Proportional" Evaluation System:
  - Ratings are such that if *x percent* of the teaching population teaches in schooling environment *y* (e.g., in high-poverty schools), then *x percent* of any subset of teacher rankings (e.g., the top quintile) includes teachers who teach in schooling environment *y.*
  - Proportional teacher rankings can conflict with teachers' global quality rankings. As a practical matter, they will tend to favor teachers in disadvantaged schools.
- "Global" Evaluation System:
  - Attempts to compare teachers across schooling environments. For example, if a higher percentage of more-effective teachers (as estimated by the model) work in schooling-environment *y*, then a disproportionate share of top-quintile teachers based on global ratings can work in schooling environment *y*.

Example of Proportional and Global Ratings (school-level; from Ehlert et al., 2013)





Background and Motivation

CALDER
#CALDER2014

- Last year we presented a paper that provides a substantive argument for the value of the proportionality property in teacher/school evaluations. We identify the following benefits of proportionality:
  - Proportional performance signals are useful in that they (a) allow high- and low-poverty schools/teachers to understand their context-specific performance and (b) facilitate productive educator-to-educator learning (note that proportional performance measures can be reported in conjunction with global-reference benchmarks if desired).
  - They elicit optimal effort from educators.
  - They avoid exacerbating labor-market inequities between high- and low-poverty schools.

- Also: Proportional models can be used to help address concerns from labor groups and community organizations about fairness and diversity in teaching (Polikoff et al., forthcoming; Vaznis, 2013, Boston Globe).
  - At the time of the Boston Globe article, black teachers in Boston were three times more likely than white teachers to be placed on a “direct growth plan” or “improvement plan.” Both plans can lead to termination.

Background and Motivation

CALDER
#CALDER2014

- But…
  - There has been some hesitation about the idea that a removal policy might purposefully not target the "globally" least effective teachers in schools; e.g., a proportional policy.
    - The sentiment seems to be that such a feature would not be conducive to improving student learning.
      - A symmetric argument could be applied to a retention-bonus policy.

Background and Motivation

CALDER
#CALDER2014

- Example removal policy comparison:
  - Assume that teachers in low-poverty schools are, on average, more effective than teachers in high-poverty schools (e.g., see Arcaira et al., 2013; Goldhaber, Walch and Gabele, 2013; Isenberg et al., 2013; Sass et al., 2012).
  - Suppose that we can remove 10 percent of teachers based on either a global or proportional policy and that we have perfect information.
    - Global policy result: more teachers in high-poverty schools will be removed
    - Proportional policy result: an equal number of teachers in high- and low-poverty schools will be removed (forced equally-circumstanced comparisons).
  - Which policy is more efficient?

Key Result

- When average teacher quality differs across different types of schooling environments, and under a wide range of evaluation circumstances, **the proportional removal policy is more efficient in terms of increasing overall student achievement (and certainly no less efficient) than the global alternative**.
  - This is true:
    - Despite the fact that the proportional policy will result in the removal of some teachers who are globally more effective than other teachers who are not removed. The reason becomes apparent when we take direct account of the link between observed teacher quality and the quality of teacher replacements.
    - Even when we put the global measures in their best possible light by assuming we can fully account for schooling context with the model that is used to estimate teacher effectiveness – no bias from uncontrolled aspects of challenging educational environments.

Research Design

- Simulation analysis:
  - Simulation framework built on the foundation provided by Winters and Cowen (2013).
  - Real data for retrospective evaluation of system design will not be available for many years.
    - This should not prevent rigorous inquiry now because mistakes prior to the backward-looking studies will be costly.
- The key feature of the simulation setup is differentiated schools.
  - Type-A, low-poverty schools with higher-quality teachers
  - Type-B, high-poverty schools with lower-quality teachers

Proof of Concept

- Initially Specify:

$$q_j^A \sim N(0.05, 0.15)$$

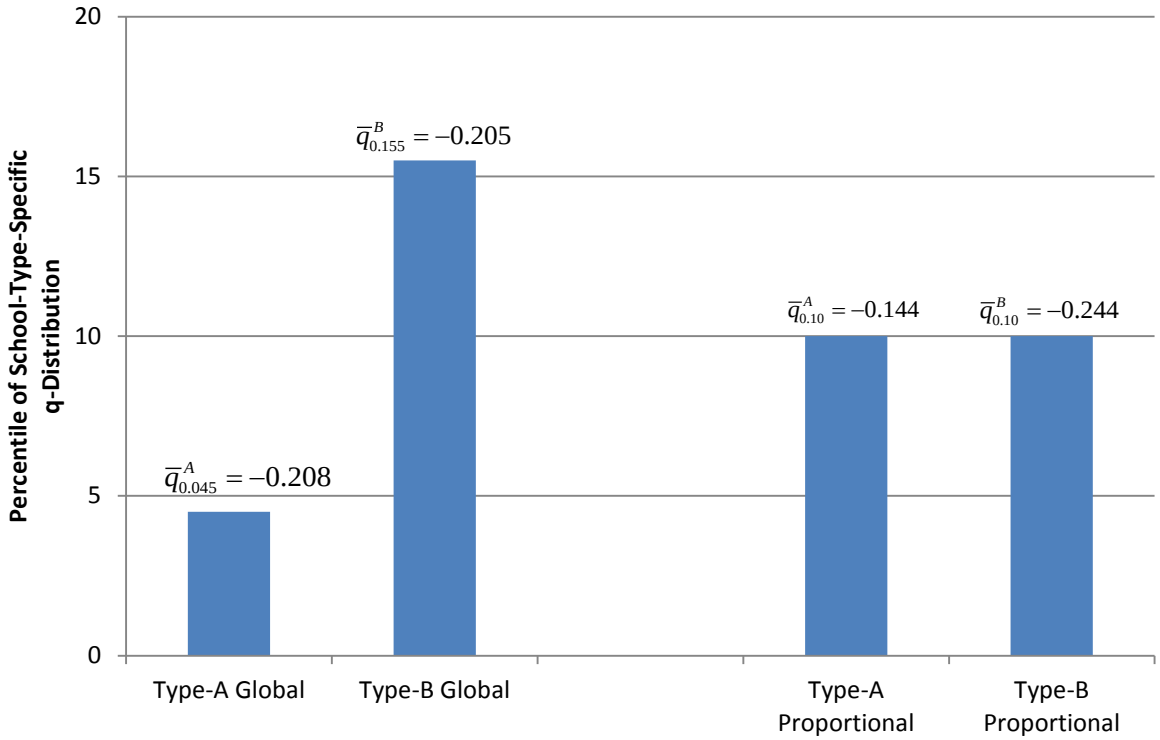
$$q_j^B \sim N(-0.05, 0.15)$$

- The expected gap in quality between low- and high-poverty schools is 0.10 standard deviations of student test scores.
  - These distributions imply what is perhaps an implausibly large gap in quality across school types (although this may depend on how you define "high poverty" and "low poverty"). We consider more moderate gaps in the paper – intuition carries over.

- **Replacement teachers in each school type are drawn from the same distributions (more on this later).**

Proof of Concept

- Average Percentile Rankings and *q*-Values for Marginally Removed Teacher at Each School Type Using Global and Proportional Rankings. Based on 1,000 Simulations.



Note: The expected value of *q* for replacement teachers in the figure is +0.05 at type-A schools and -0.05 at type-B schools

Proof of Concept

- Global policy

	Type-A (0.045)	Type-B (0.155)
Expected Quality of Replacement	+0.050	-0.050
Quality of Marginally Removed Teacher	-0.208	-0.205
Expected Gain from Marginal Removal	**+0.258**	**+0.155**

- Proportional Policy

	Type-A (0.10)	Type-B (0.10)
Expected Quality of Replacement	+0.050	-0.050
Quality of Marginally Removed Teacher	-0.144	-0.244
Expected Gain from Marginal Removal	**+0.194**	**+0.194**

The Link Between Observed- and Replacement-Teacher Quality

CALDER
#CALDER2014

- The efficiency rationale hinges on replacement teachers at each school-type coming from the same distribution as current teachers.
- Is this reasonable? What can account for estimated gaps in teacher quality across school types that have been documented in the literature? We offer four possible explanations:
  - The gaps reflect applicant pool quality (disadvantaged schools face labor-market challenges: Boyd et al., 2005; Clotfelter et al., 2006; Jacob, 2007; Reininger, 2012).
  - The gaps reflect differences in the quality of leadership across school types (Koedel et al., 2011).
  - The gaps reflect differences in access to instructional strategies for teachers across school types (Ehlert et al., 2013; Raudenbusch and Willms, 1995).
  - The gaps reflect bias in estimation (e.g., from student sorting), in which case they would not be real.
- It turns out that no matter the root cause of the gap (among explanations we can think of), whatever we see for current teachers will translate to their replacements absent direct policy intervention to remedy the problem (assuming the gap is real).

Caveats and Complications

- Caveat : There is a strong efficiency rationale to support proportionality when teacher quality differs across school types on average; however, under practical evaluation conditions the efficiency gains from proportionality are close to zero.
- In the remainder of the paper, we show that the efficiency gains from proportionality are attenuated in practice by:
  - Estimation error in estimated teacher effects (which lowers the average gain in quality from removals – the removal policy is clumsier)
  - Allowing for natural teacher attrition (which decreases the number of forced removals from a fixed-rate removal policy)
  - Allowing for smaller (and more reasonable) gaps in quality across school types.

- Also, if the **variance** in teacher-quality is higher in one sector relative to the other, the proportional policy need not be the most efficient.
  - Sass et al. (2012) find evidence of more variance in teacher quality in high-poverty schools.
  - We parameterize simulations around their distributional estimates and find that the proportional policy is roughly efficiency neutral.

Caveats and Complications

CALDER
#CALDER2014

- Conversely,
  - The efficiency gains from proportionality increase if global teacher-quality estimates are biased in the most-likely direction (to favor teachers in low-poverty schools)
  - In the simulations we only allow the proportional policy to improve efficiency by affecting which teachers are replaced. We do not allow teachers to respond behaviorally to the proportional policy.
    - Ehlert et al. (2013) argue that proportionality can have efficiency-improving behavioral effects.

Caveats and Complications

CALDER
#CALDER2014

- The issue of equity is really complicated.
  - The proportional policy moves removals out of high-poverty schools and into low-poverty schools. Because each removal below the 50th percentile has a positive effect on achievement in expectation, this lowers achievement at high-poverty schools.
    - Put differently, the policy that would result in the largest achievement gains for high-poverty schools would be to remove only teachers at those schools. This seems inconsistent with larger policy objectives (at least stated).
    - Thus, proportionality as an equity cost in this way.
  - A symmetric proportional retention policy (not considered in our study) will increase efficiency for the same reason as the removal policy and will shift retention efforts toward high-poverty schools. Thus, a proportional retention policy is efficiency enhancing and it raises achievement in high-poverty schools.
  - Teacher turnover is costly, and more costly at high-poverty schools (Ronfeldt et al., 2011). When we incorporate turnover costs the proportional policy is less inequitable because it shifts costly removals away from high-poverty schools.
  - Ehlert et al. (2013) discuss the potential for equity-enhancing behavioral benefits of proportionality. Recall that we do not account for any behavioral responses in our study.
- In summary, the question of whether the proportional policy is more or less equitable is complicated. Purely within the context of the simulated removal policy, it has a net effect of lower student achievement in high-poverty schools because it lowers the forced exit rate at those schools.

Concluding Remarks

CALDER
#CALDER2014

- We examine the efficiency implications of using proportional evaluations to shape the teaching workforce and show that under a wide range of evaluation circumstances, the proportional removal policy is not less efficient than the global alternative (and is more efficient under some conditions).
  - Although one might initially think that removing teachers in a way that is not consistent with a "global" quality ranking would harm efficiency, this is not generally the case.

- Overall, the efficiency gains from proportionality are likely to be close to zero in real-world evaluation contexts.
  - Given the other benefits that proportionality offers (as discussed in Ehlert et al., 2013), which include the possibility of positive behavioral responses from educators, the fact that proportionality can be achieved without efficiency costs in our simulations is notable.
  - Our findings point to proportional evaluations as being a viable alternative for educational administrators charged with developing and implementing teacher evaluation systems.