



What Do We Know About the Tradeoffs Associated with Teacher Misclassification in High Stakes Personnel Decisions?

Dan Goldhaber

Susanna Loeb

May 8, 2013

Introduction

- Measures of performance are imperfect
 - Not always reflecting true performance
- Yet, personnel decisions and teacher classification is unavoidable (whether explicit or implicit)
 - Who is eligible to teach
 - Who teaches in each classroom
 - Who receives tenure
 - How teachers are promoted or compensated
- A range of imperfect information used
 - Traditionally: teaching experience, academic degrees, administrator evaluations
 - More recently: structured classroom observation ratings, “value-added” estimates
- Errors are inevitable, highlighted by the use of formal evaluation measures
 - Improper sanctioning of good teachers can harm those teachers, and may discourage capable people from entering the teaching
 - Improper classify of poorly performing teachers as higher performers who continue to teach without additional supports can harm students
 - The potential harm to students is often overlooked

Questions for this Brief

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

Questions

- 1. What do we currently know about the misclassification of teachers based on value added?**
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

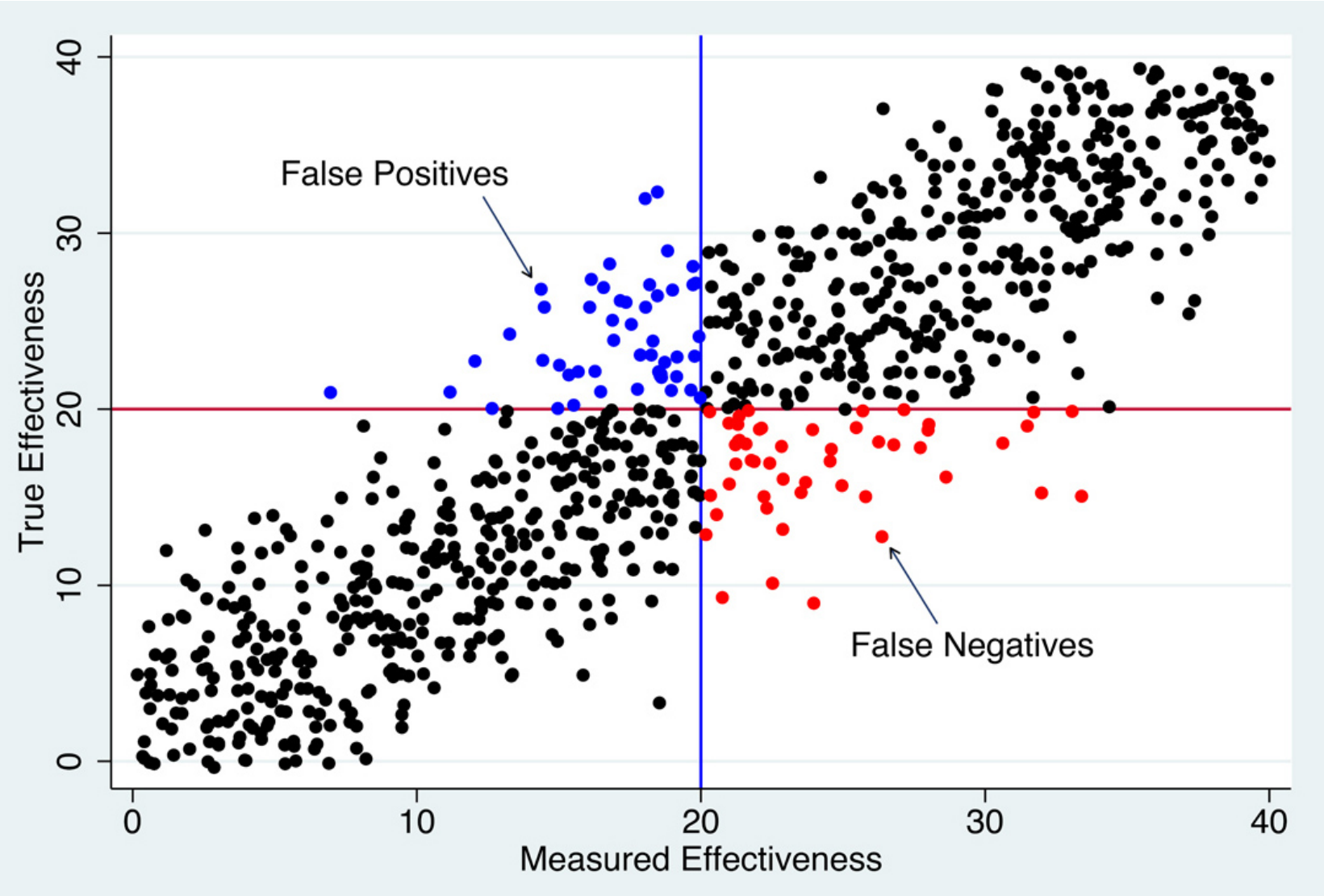
Start Conceptually

False Positives and False Negatives

An Example: Identifying Low Performers

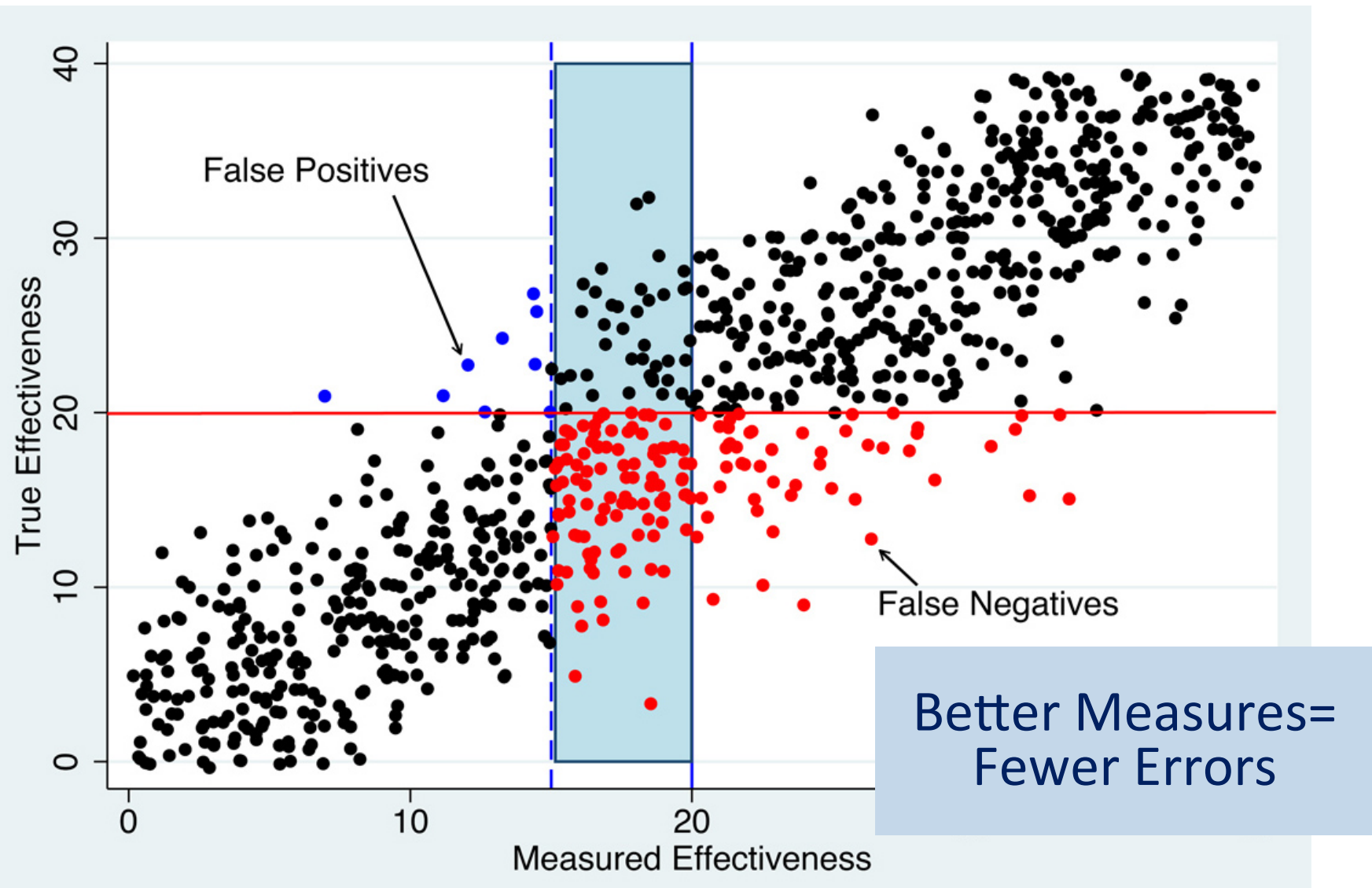
	Teacher is Truly Ineffective	Teacher is <u>Not</u> Truly Ineffective
Performance <i>measure</i> suggests teacher is ineffective	Correct outcome, True positive	Incorrect outcome, False positive
Performance <i>measure</i> suggests teacher is not ineffective	Incorrect outcome, False negative	Correct outcome, True negative

Graphically: Identifying Low Performers

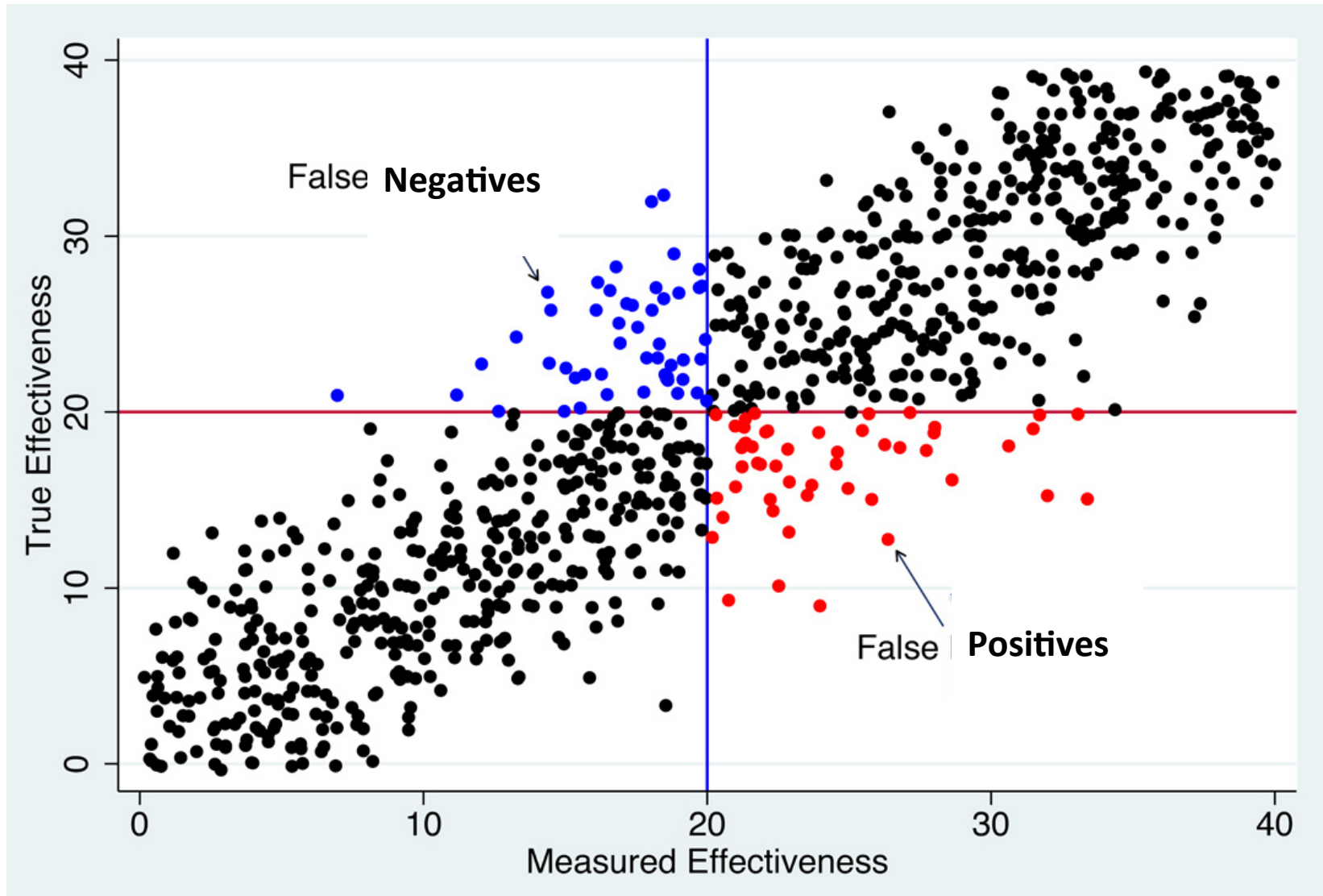


Inherent Tradeoff with Existing Measures

Reducing Risk to Teachers
Increases Risk to Students



Similar if Identifying High Performers



Empirical Evidence on Tradeoffs

Now vs. Future

Table 2: Quintile Transitions from Initial Performance to Future Performance by Subject.

Pre-tenure quintile rank	Post-Tenure Quintile (percent)					Total teachers
	Bottom	Second	Third	Fourth	Top	
<i>Panel A. Using first two years of performance to predict post-tenure performance</i>						
Bottom	32	23	19	16	11	57
Second	27	14	27	18	14	56
Third	21	23	30	18	7	56
Fourth	16	27	18	18	21	56
Top	5	13	5	30	46	56
Total Teachers	57	56	56	56	56	281

Source: Goldhaber and Hansen (2008)

Using only bottom: 68% false positives, 68% false negatives

Using 2 rows: 71% false positives, 42% false negatives

Significant Misclassification

VAM Compared to What?

- When the outcome in question is student achievement on test scores, misclassification of teachers according to value added is high
- But, other measures are also imperfect, often worse at predicting future value-added.
 - e.g. misclassification if we use licensure tests to measure quality is likely much higher.
- These comparisons are with future value-added. We know little about how measures compare relative to other student outcomes of interest.

Questions

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

Questions

1. What do we currently know about the misclassification of teachers based on value added?
- 2. What more needs to be known about misclassification?**
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

What More Needs to be Known?

- Misclassification is an inescapable part of decision-making. Yet, we can reduce the harm of misclassification by...
 1. better measurement to reduce misclassification
 2. balancing the two types of errors
 3. considering carefully the consequences of classification and implementation of reforms
- Measurement: How to improve performance measurement?
 - Better underlying exams
 - Better construction of value-added estimates
 - Combination with other teacher performance measures
- Balancing the two types of error: How would identification of low-performing teachers affect schools and teachers?
 - Will pre-service preparation improve to meet the new demands?
 - Will those rated ineffective be more inclined to seek out high-quality professional development?
 - Will identification systems affect the appeal of teaching as a career?

What More Needs to be Known?

- Implementation / Consequences: How can reforms be structured to benefit schools and students?
 - What effects are reforms having? How does this vary by choices such as how to combine value-added with other measures and what consequences to link to evaluation?
 - How will administrators and teachers respond to policies that attach high stakes to teacher performance measures?
 - Are there ways that we can reduce the consequences for students of false negatives?
 - Are there ways to allow for false positives that reduce risks for teachers (e.g. follow-ups) while still gaining some of the benefits for students of identification?
- How will the legal system handle challenges to the use of *student growth measures* for high-stakes teacher evaluation?

Questions

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

Questions

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. **What can't be *directly* resolved by empirical evidence on misclassification?**
4. What are the practical implications of the research on misclassification for decision making?

What Can't Be Resolved Based on Empirical Evidence?

- What is the “right” tradeoff between false positives and false negatives
 - Even if evidence showed the impact of such policies on observable outcomes, such as student test scores, good teachers produce learning gains in areas that go beyond tested academic subjects
- What are the “right” times for consequential decisions and what should the stakes be in those decisions

Questions

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. What are the practical implications of the research on misclassification for decision making?

Questions

1. What do we currently know about the misclassification of teachers based on value added?
2. What more needs to be known about misclassification?
3. What can't be *directly* resolved by empirical evidence on misclassification?
4. **What are the practical implications of the research on misclassification for decision making?**

What Are the Practical Implications of the Research on Decision-making?

- Personnel decisions are inevitable
- Information used for these decisions will be imperfect, resulting in
 - False positives (those that should not be identified, but are) and
 - False negatives (those who should be identified, but are not)
- The magnitude of false positives **and** false negatives are important to consider
 - False positives for low performance may harm teachers
 - False negatives for low performance may harm students
- Currently, likely reducing false positive by allowing a high rate of false negatives.
- Tradeoff is inherent, unless measures become more accurate
- System structures can reduce the potentially negative consequences of misidentification
 - Multiple measures
 - Student access to multiple teachers...
- Little current information on best practices



What Do We Know About the Tradeoffs Associated with Teacher Misclassification in High Stakes Personnel Decisions?

Dan Goldhaber

Susanna Loeb

May 8, 2013