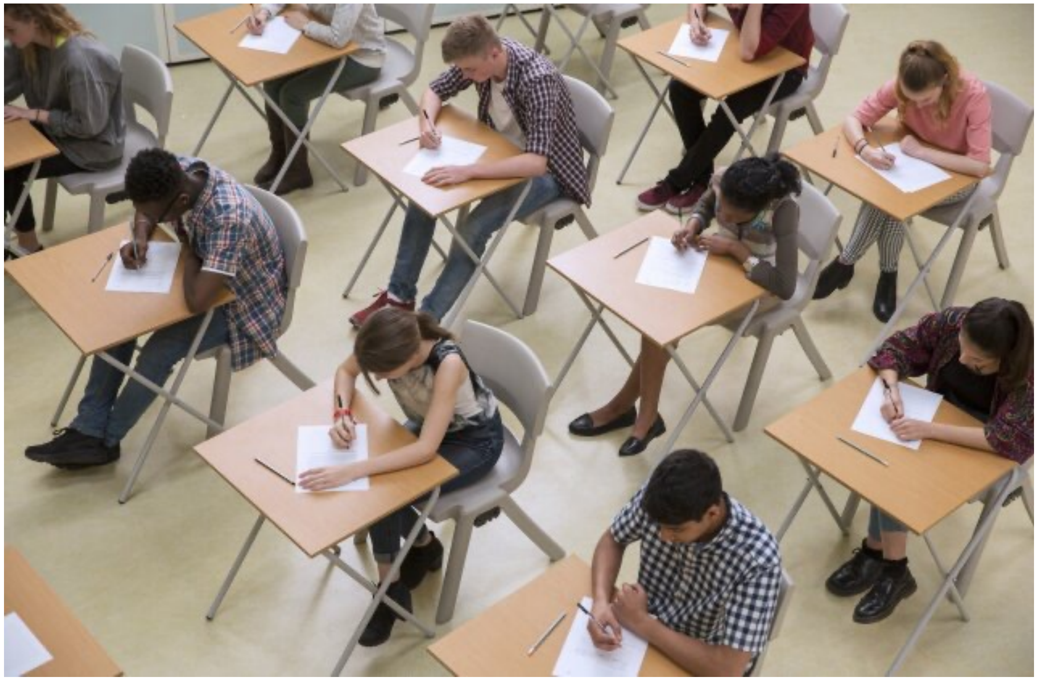
What's in a Question? Using Item Response Data to Better Represent Learning

Jesse Bruhn

joint with Mike Gilraine, Jens Ludwig, and Sendhil Mullainathan

Testing is a major part of education



Ex: standardized testing and prep occupy as much as 18% of instructional time.

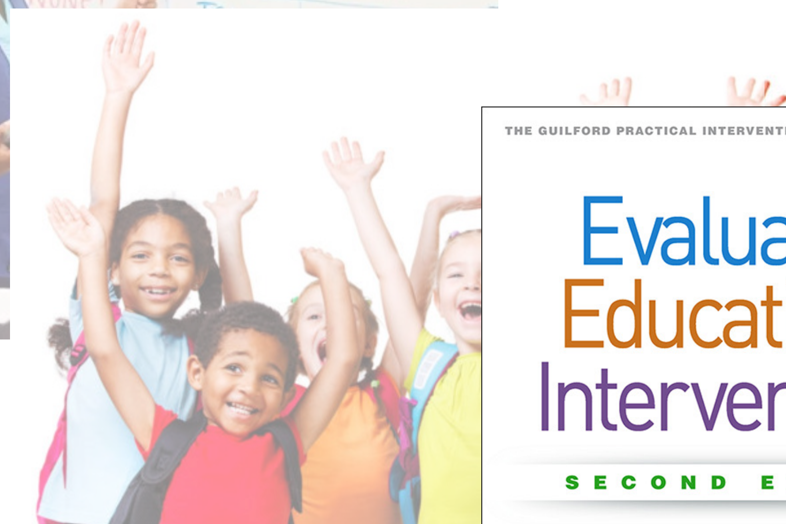
(Nelson et al, 2013)

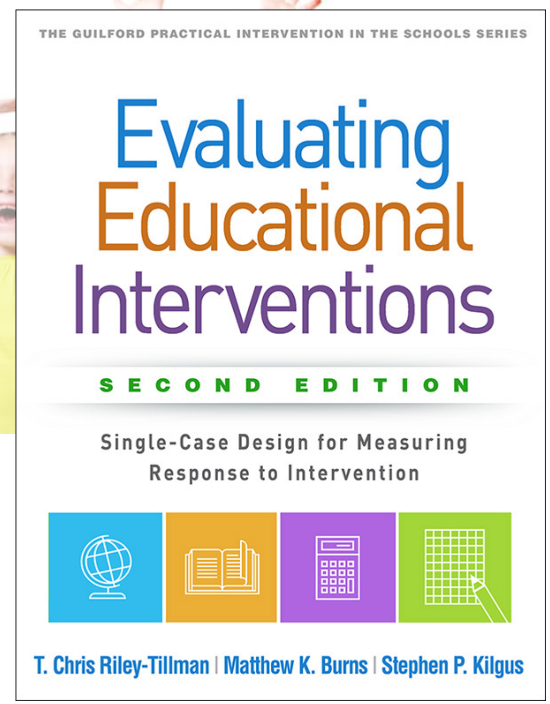
Tests are used to make high-stakes decisions

Tests are used to make high-stakes decisions

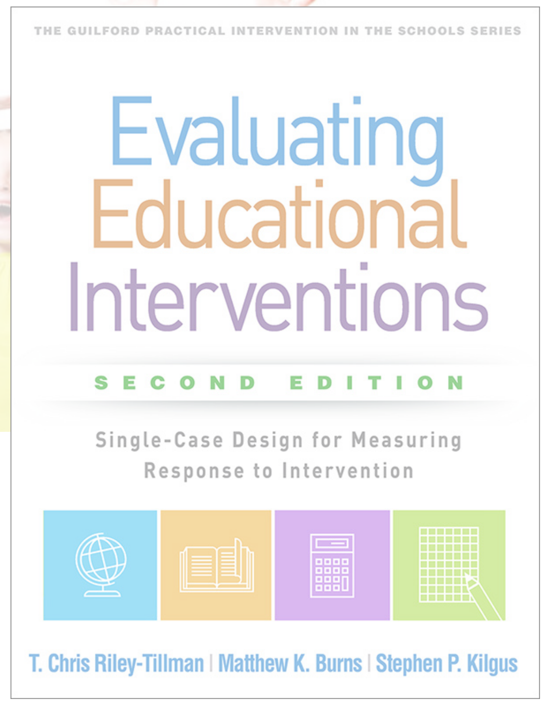
Tests are used to make high-stakes decisions

Tests are used to make high-stakes decisions

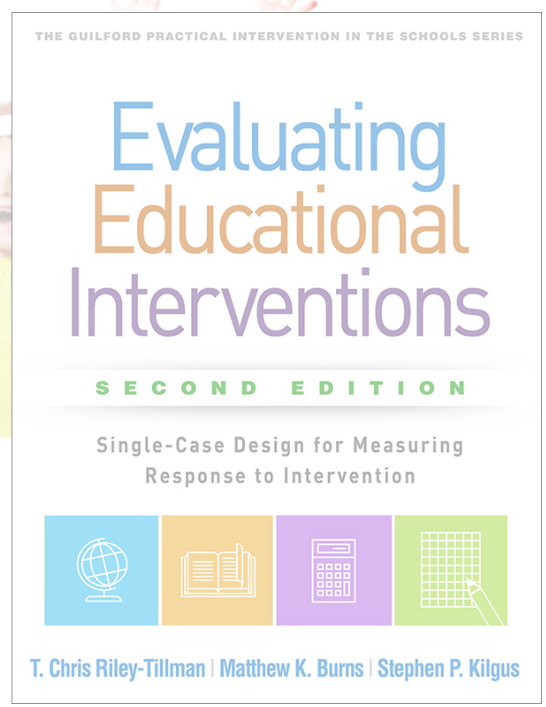




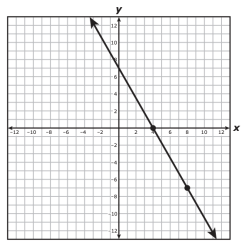
Tests are used to make high-stakes decisions



Test results



15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A $68 > 649$

B $571 > 582$

C $730 < 806$

D $709 < 692$

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F	10	10	10	10	10	10
G	3	3	3	3	3	3
H	3	10		3	10	
J	10			10		

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

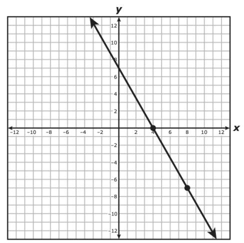
F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F [10 | 10 | 10] [10 | 10 | 10]

G [3 | 3 | 3] [3 | 3 | 3]

H [3 | 10] [3 | 10]

J [10] [10]

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

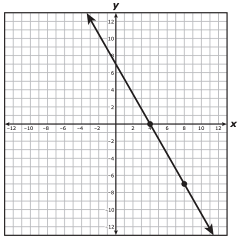
G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$

Aggregator

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F | 10 | 10 | 10 | | 10 | 10 | 10 |

G | 3 | 3 | 3 | | 3 | 3 | 3 |

H | 3 | 10 | | 3 | 10 |

J | 10 | | 10 |

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

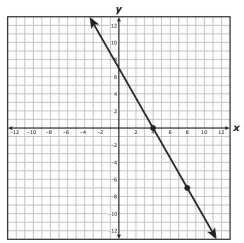
J $d(x) = 0.26x^2 - 3.11x + 1$

Aggregator

Performance Measure

Statistics behind aggregation (IRT, MIRT)

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F | 10 | 10 | 10 | | 10 | 10 | 10 |

G | 3 | 3 | 3 | | 3 | 3 | 3 |

H | 3 | 10 | | 3 | 10 |

J | 10 | | 10 |

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

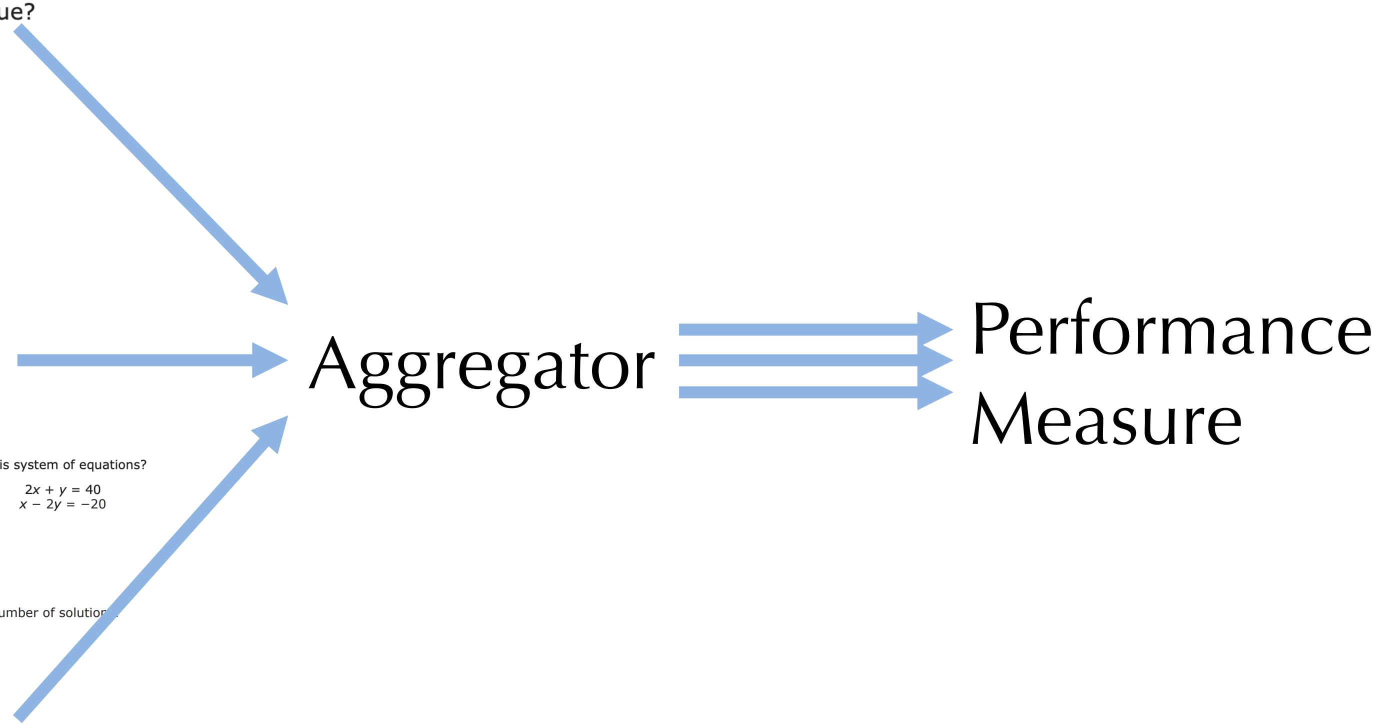
Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

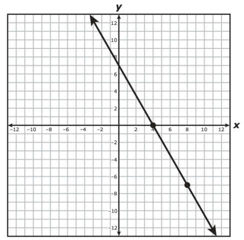
H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$



Statistics behind aggregation (IRT, MIRT)

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F [10 | 10 | 10] [10 | 10 | 10]

G [3 | 3 | 3] [3 | 3 | 3]

H [3 | 10] [3 | 10]

J [10] [10]

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

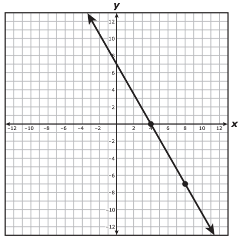
J $d(x) = 0.26x^2 - 3.11x + 1$

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Performance Measure

Statistics behind aggregation (IRT, MIRT)

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F | 10 | 10 | 10 | | 10 | 10 | 10 |

G | 3 | 3 | 3 | | 3 | 3 | 3 |

H | 3 | 10 | | 3 | 10 |

J | 10 | | 10 |

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

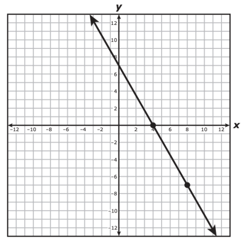
J $d(x) = 0.26x^2 - 3.11x + 1$

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

$$\hat{\theta}$$

Statistics behind aggregation (IRT, MIRT)

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F [10 | 10 | 10] [10 | 10 | 10]

G [3 | 3 | 3] [3 | 3 | 3]

H [3 | 10] [3 | 10]

J [10] [10]

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$

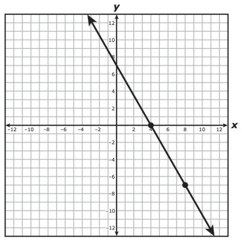
$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Average of items

Statistics behind aggregation (IRT, MIRT)

- Under some assumptions, this will be optimal.

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 $>$ 649

B 571 $>$ 582

C 730 $<$ 806

D 709 $<$ 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F | 10 | 10 | 10 | | 10 | 10 | 10 |

G | 3 | 3 | 3 | | 3 | 3 | 3 |

H | 3 | 10 | | 3 | 10 |

J | 10 | | 10 |

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

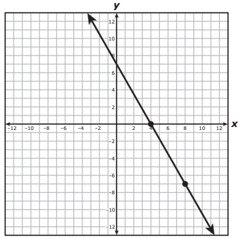
H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Average of items

15 The graph of a linear function is shown on the grid.



Which equation is best represented by this graph?

A $y = -\frac{7}{4}x + 4$

B $y = -\frac{7}{4}x + 7$

C $y = -\frac{4}{7}x + 4$

D $y = -\frac{4}{7}x + 7$

1 Which comparison is true?

A 68 > 649

B 571 > 582

C 730 < 806

D 709 < 692

12 Janet has 2 new games.

- Each game has 3 packs of cards.
- Each pack has 10 cards.

Which model can be used to find the total number of cards Janet has for these 2 games?

F [10 | 10 | 10] [10 | 10 | 10]

G [3 | 3 | 3] [3 | 3 | 3]

H [3 | 10] [3 | 10]

J [10] [10]

5 What is the solution to this system of equations?

$$2x + y = 40$$
$$x - 2y = -20$$

A (12, 16)

B (15, 17.5)

C There is no solution.

D There are an infinite number of solutions.

22 A person dives into a pool from its edge to swim to the other side. The table shows the depth in feet of the person from the surface of the water after x seconds. The data can be modeled by a quadratic function.

Pool

Time, x (seconds)	Depth of Person from Surface of Water, $d(x)$ (feet)
1	−2.85
4	−8.28
6	−9.3
8.5	−7.65
10	−5.1
11.5	−1.38

Which function best models the data?

F $d(x) = 0.05x^2 + 0.74x$

G $d(x) = 0.05x^2 + 0.74x + 9.17$

H $d(x) = 0.26x^2 - 3.11x$

J $d(x) = 0.26x^2 - 3.11x + 1$

Statistics behind aggregation (IRT, MIRT)

- Under some assumptions, this will be optimal.

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}}$$

Average of items

What information do we lose?

What do we lose?

How we evaluate **teachers**

How we evaluate **students**

How we evaluate **interventions**

What do we lose?

Teachers:

Districts use test results to evaluate teachers.

Item data reveals variability obscured by average growth.

As much as 30% of teachers in bottom decile value-add land in the top decile of item performance.

"Good versus bad teachers" is a less accurate model than "different teachers are differentially good at promoting different aspects of achievement."

In total, aggregation destroys ~60-70% of the predictable variation in student performance

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance $\Rightarrow$ Due to teacher comparative advantage

What do we lose?

Students: We use test scores as proxies for later life outcomes.

In total, aggregation destroys as much as 55% of predictable variation in graduation, college attendance, and earnings.

Less than 50% agreement re: "ineffective" educators using predicted student outcomes versus typical aggregates.

Summary statistics using alternative weights lead to different policies and priorities.

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

What do we lose?

Interventions: The impact of pre-K, small class size, and quality teachers "fades-out" on test scores only to reemerge later in life.

Fadeout is heterogeneous item-by-item.

Fade-out is partly an illusion due to changing composition of items across tests

Even very crude alternative weighted averages based on item difficulty can double persistence

Can even find weighted averages that "fade-in"

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

Fadeout: At least 50% of persistence ⟹ Statistical artifact of test composition

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

Fadeout: At least 50% of persistence ⟹ Statistical artifact of test composition

Contribution

Educational measurement:

- e.g. Anaya et al. (2022) Bond & Lang (2018), Cascio & Staiger (2012), Cunha et al. (2008, 2010), Jacob & Rothstein (2016), Kaur et al. (2023), Lang (2010), Nielsen (2019, 2023), Reyes (2023).
- Explore implications of item aggregation for measuring educational performance.

Teacher value-add:

- e.g. Chetty et al. (2014a, 2014b), Gilraine & Pope (2022), Jackson (2018), Mulhern & Opper (2022), Papay (2011), Rose et al (2022), and many others
- Highlight potential for item data to generate new / nuanced TVA measures.

Fadeout:

- e.g. Bailey et al (2017), Cascio & Staiger (2012), Chetty et al. (2011), Currie & Thomas (1995), Ludwig & Miller (2007), Deming (2009), Heckman et al (2013), Puma et al (2010), Gray-Lobe et al (2022).
- New explanation based on the changing composition of item content.

Universe of Texas K-12 students:

- 4.5 million students
- 14 million student-years
- 1.24 billion student-year-test items

Linked to:

- Test scores
- Item responses
- Teachers
- Graduation, college attendance, earnings.

	Full Sample (1)	Teacher-Student Matched Sample (2)
Panel A: Standardized Tests		
# of items on Math Test	52.0	49.0
% Correct on Math Test	57.3	56.8
# of items on English Test	44.0	45.2
% Correct on English Test	65.1	65.9
Panel B: Demographics		
% Hispanic	51.5	51.3
% Black	12.7	13.0
% Free Lunch Eligible	51.1	51.9
Class Size	-	22.0
# of Students	4,495,344	3,644,164
# of Teachers	-	81,628
Observations (student-year)	14,014,753	9,073,848
Observations (student-item-year)	1,240,841,152	855,056,544

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⇉ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⇉ Different priorities from different averages

Fadeout: At least 50% of persistence ⇉ Statistical artifact of test composition

How much information do we lose about teachers?

$$D_{iqt} = \alpha_{qt} + \Gamma X_{it} + \eta_{iqt}$$

D_{iqt} $\Rightarrow$ Takes a value of one if student i correctly answered item q in year t.

α_{qt} $\Rightarrow$ Question fixed effect.

X_{it} $\Rightarrow$ Standard Chetty et al. (2014a,b) vector of teacher value-added covariates, including lagged average score.

How much information do we lose about teachers?

$$D_{iqt} = \alpha_{qt} + \Gamma X_{it} + \eta_{iqt}$$

$var(\eta) \longrightarrow$ Unexplained student performance

$$D_{iqt} = \alpha_{qt} + \delta_{qtj(i,t)} + \Gamma X_{it} + u_{iqt}$$

$\delta_{qtj(i,t)} \Rightarrow$ Teacher $j(i,t)$ by item q in year t fixed effect

How much information do we lose about teachers?

$$D_{iqt} = \alpha_{qt} + \Gamma X_{it} + \eta_{iqt}$$

$var(\eta) \longrightarrow$ Unexplained student performance

$$D_{iqt} = \alpha_{qt} + \delta_{qtj(i,t)} + \Gamma X_{it} + u_{iqt}$$

$var(\eta) - var(u) \longrightarrow$ Explained by teachers

$$D_{iqt} = \alpha_{qt} + \delta_{j(i,t)} + \Gamma X_{it} + \epsilon_{iqt}$$

$\delta_{tj(i,t)} \Rightarrow$ Teacher $j(i,t)$ by year t fixed effect.

- Up to a scaling, equivalent to "standard" TVA for average scores.

How much information do we lose about teachers?

$$D_{iqt} = \alpha_{qt} + \Gamma X_{it} + \eta_{iqt}$$

$var(\eta) \longrightarrow$ Unexplained student performance

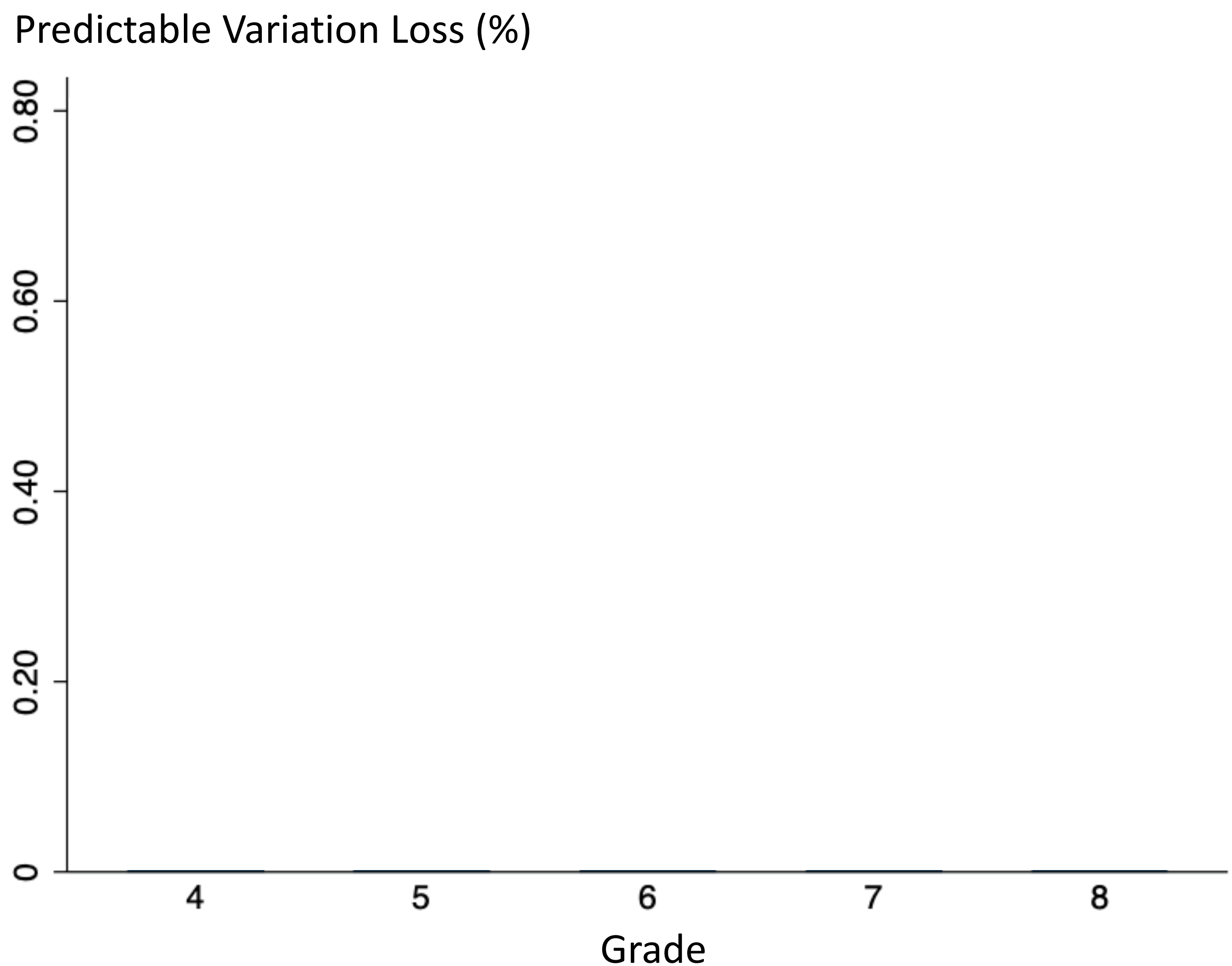
$$D_{iqt} = \alpha_{qt} + \delta_{qtj(i,t)} + \Gamma X_{it} + u_{iqt}$$

$var(\eta) - var(u) \longrightarrow$ Explained by teachers

$$D_{iqt} = \alpha_{qt} + \delta_{j(i,t)} + \Gamma X_{it} + \epsilon_{iqt}$$

$var(\epsilon) - var(u) \longrightarrow$ Lost by averaging.

How much information do we lose about teachers?



What kind of info? Comparative advantage.

Teacher Rank

What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

Fadeout: At least 50% of persistence ⟹ Statistical artifact of test composition

How much do we lose about student outcomes?

$$Y_i = F(X_i, W_i) + \eta_i$$

$$Y_i = G(\bar{X}_i, \bar{W}_i) + \epsilon_i$$

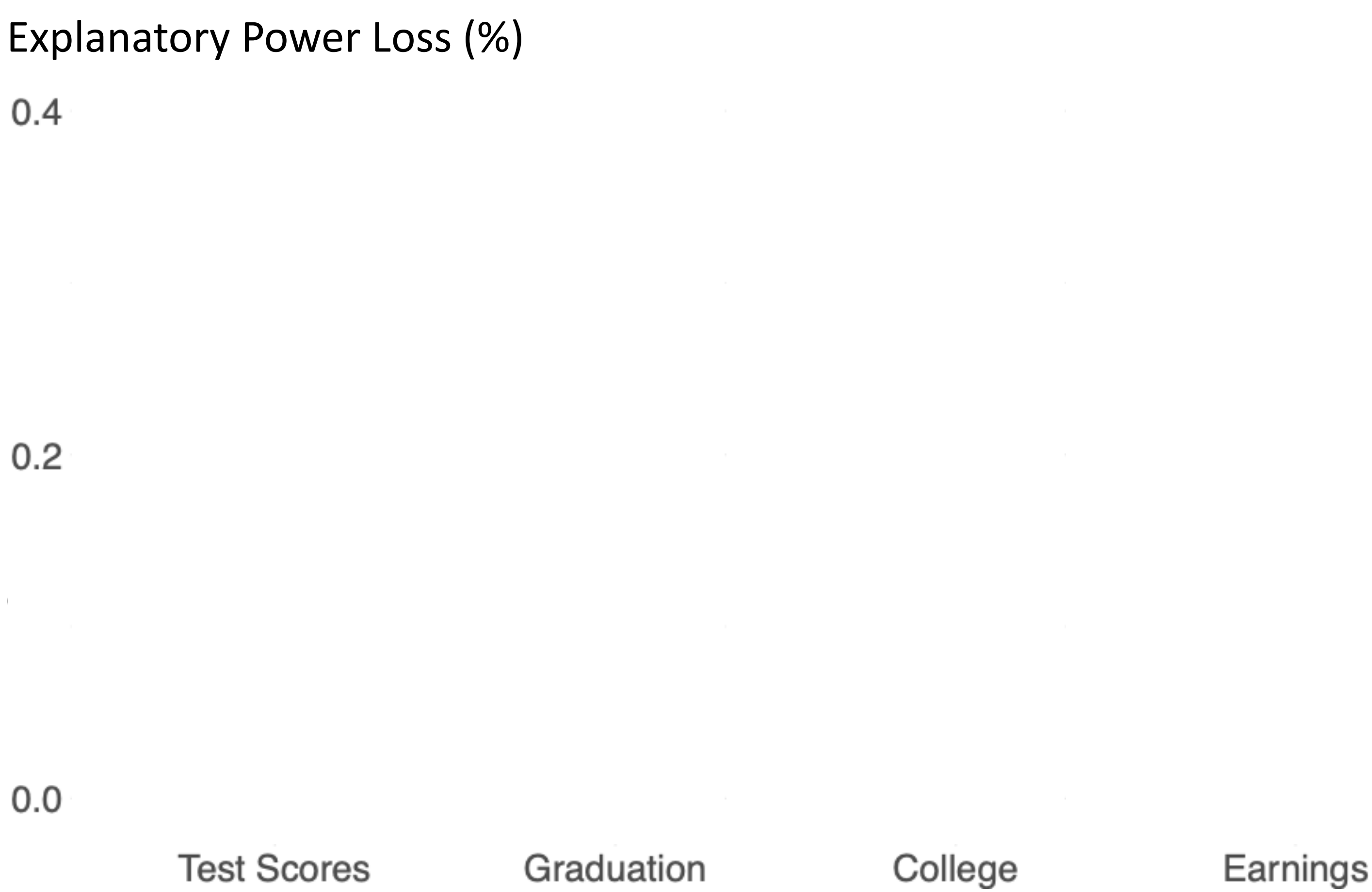
Where:

$X_i = \{x_i\}_{a \in M} \longrightarrow$ Indicator variables denoting *exact answers* (~160 per grade-year) to math items.

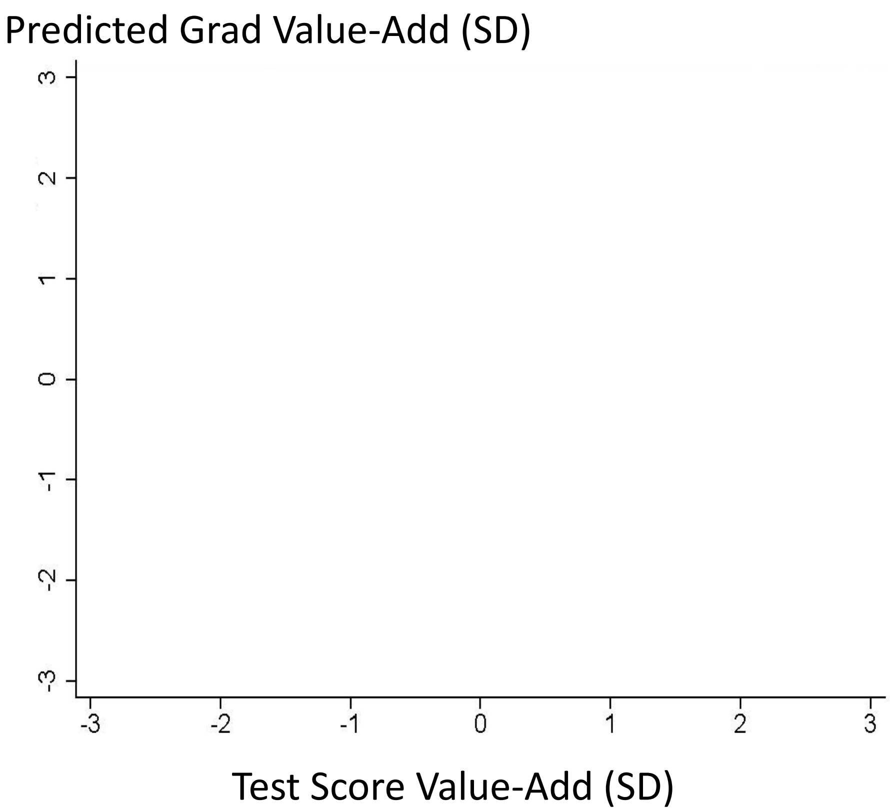
$W_i = \{w_i\}_{a \in E} \longrightarrow$ Indicator variables denoting *exact answers* (~160 per grade-year) to ELA items.

$F()$ and $G() \longrightarrow$ Learned from data using a Gradient Boosted Tree algorithm (Chen & Guestrin, 2016)

How much do we lose about student outcomes?



“Outcome” value-add versus test score value-add



View of ``ineffective” varies with individual item weighting

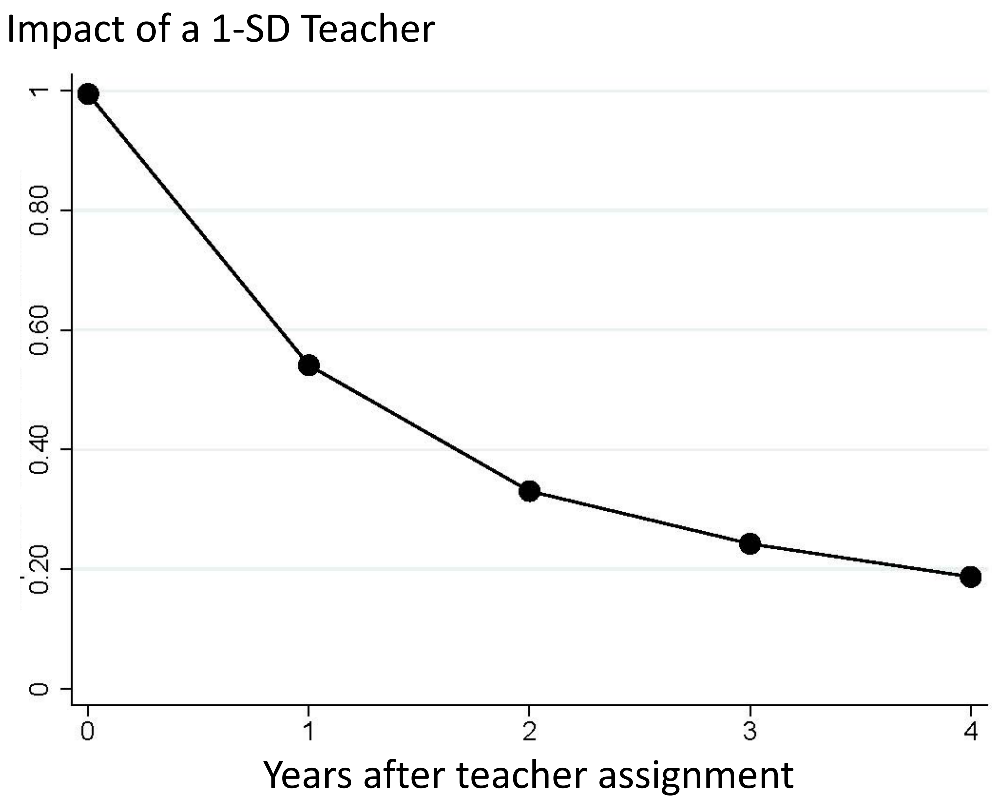
What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

Fadeout: At least 50% of persistence ⟹ Statistical artifact of test composition

Fadeout



Potential explanations:

- Real skill depreciation, similar to fadeout of job training on wages (e.g. Crépon et al., 2013)
- Non-cognitive skills (Heckman et al., 2013)
- Artifact of normalization (Cascio and Staiger, 2012)

But tests aren't like wages…

Different tests measure different concepts.

4th Grade Math item ?? 5th Grade Math item

8 Which equation shows a decimal and a fraction that are equivalent?

F $23.5 = 23\frac{5}{100}$

G $23.55 = 23\frac{55}{10}$

H $23.05 = 23\frac{5}{10}$

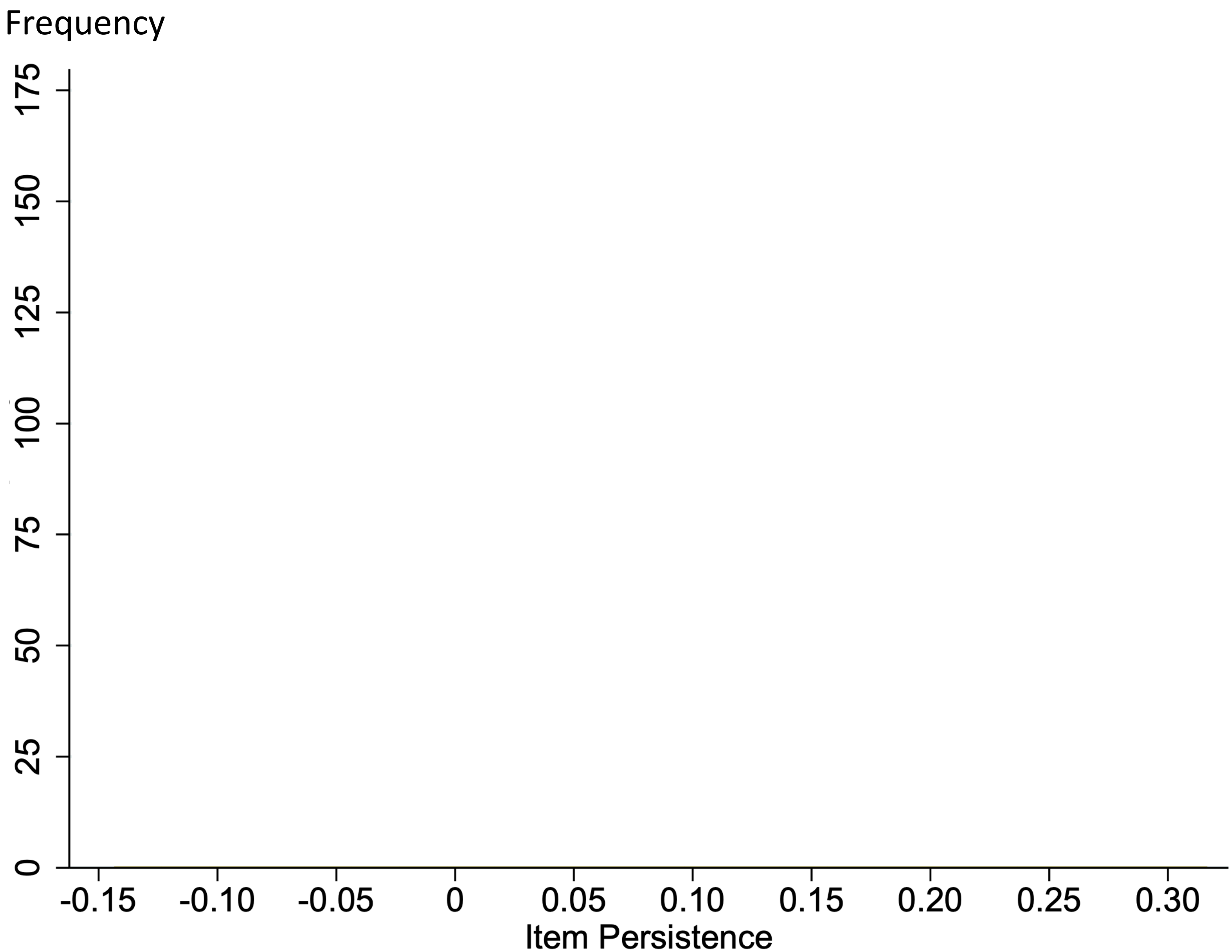
J $23.5 = 23\frac{50}{100}$

2 A worker is building toys at a factory. The relationship between the number of hours the employee works, x, and the number of toys the employee builds, y, is represented by the equation $y = 9x$.

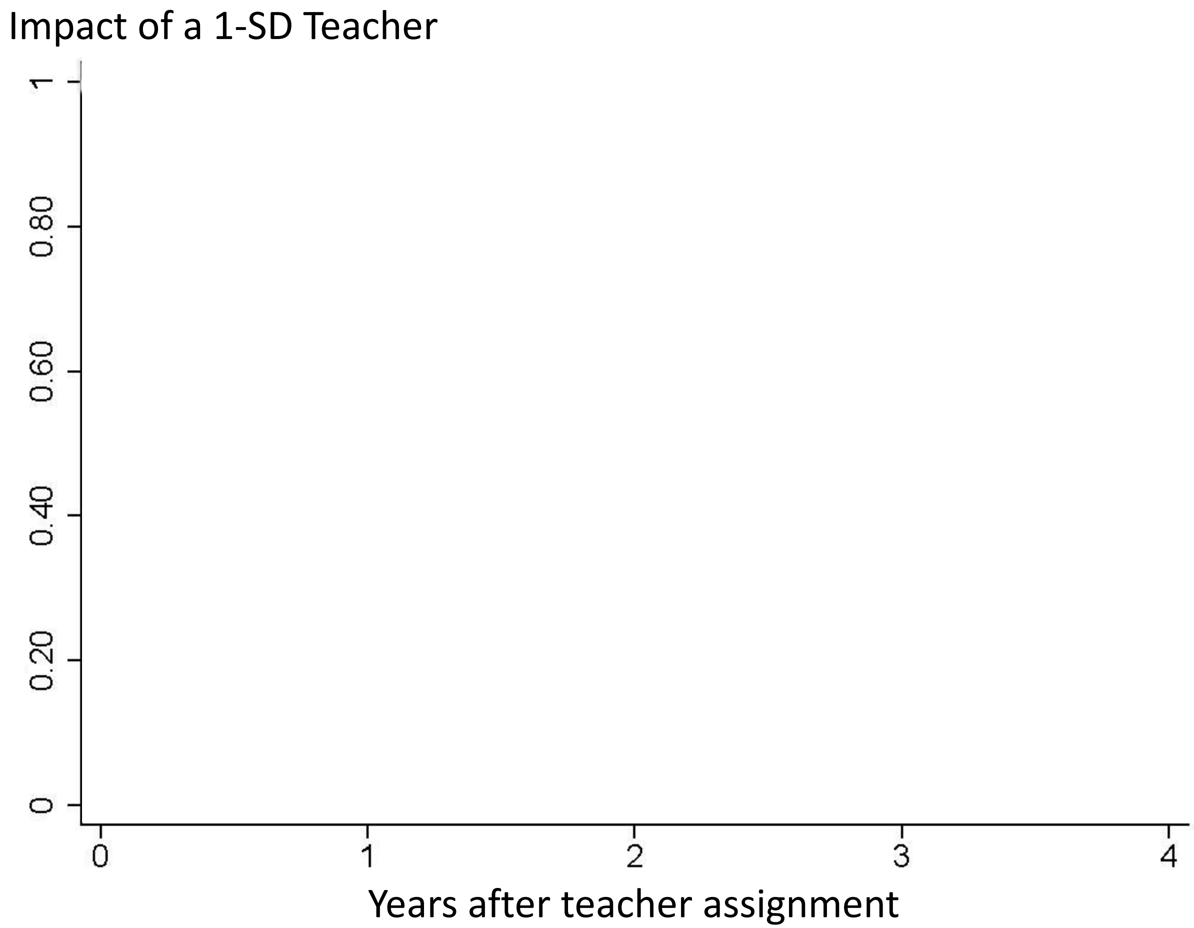
Which graph represents this relationship?

Is fadeout uniform across items?

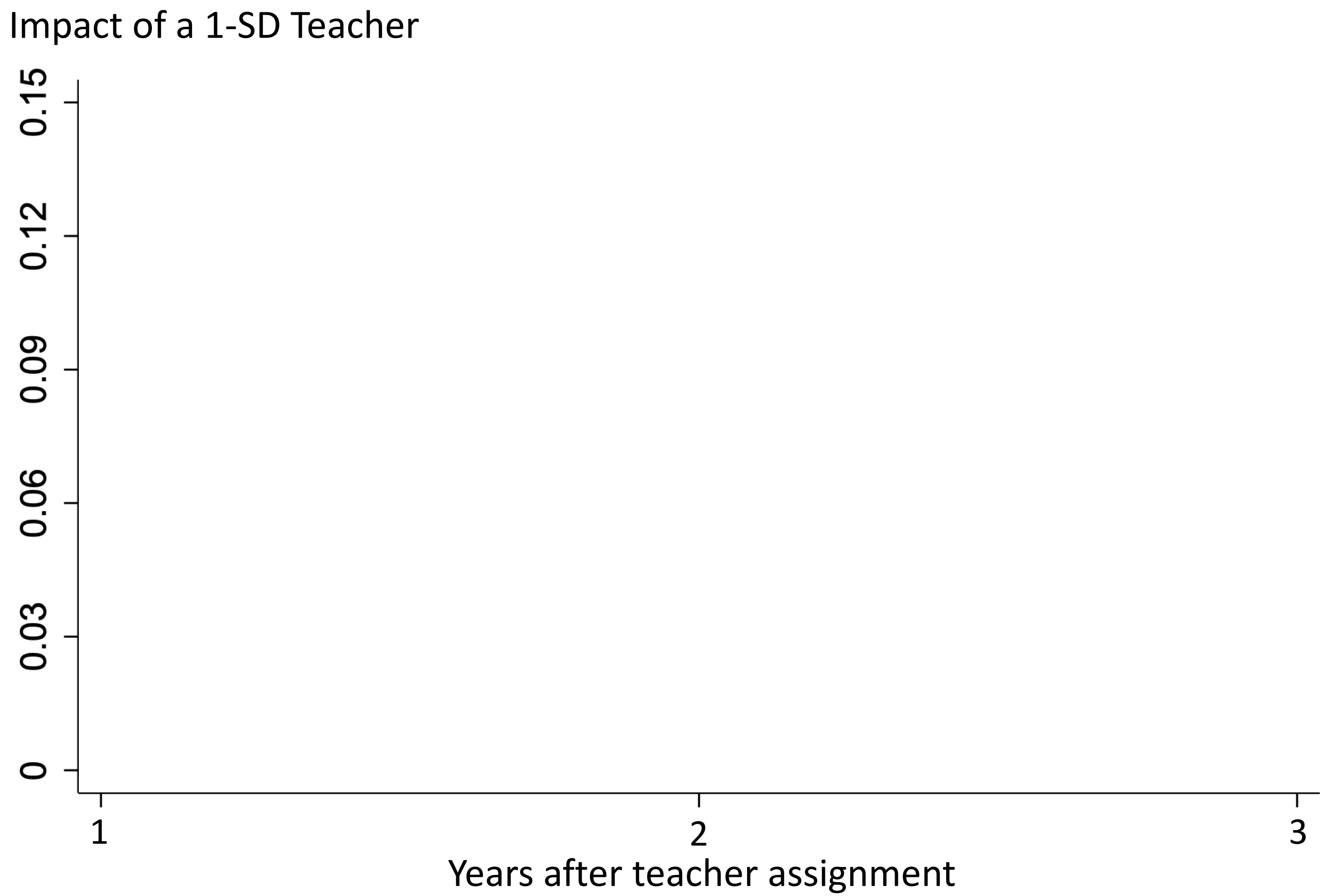
Fadeout is not uniform across items.



Crude reweighting schemes can double persistence.



Even find **fade-in** for certain weighted averages.



What do we lose?

Teachers: ~60-70% of the predictable variation in student performance ⟹ Due to comparative advantage across items

Students: As much as 55% of predictable variation in long-run outcomes ⟹ Different priorities from different averages

Fadeout: At least 50% of persistence ⟹ Statistical artifact of test composition

Thank you!