SCREEN TWICE, CUT ONCE: ASSESSING THE PREDICTIVE VALIDITY OF APPLICANT SELECTION TOOLS
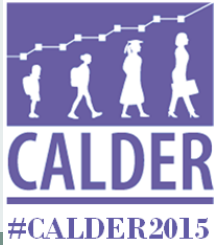
Dan Goldhaber
Cyrus Grout
Nick Huntington-Klein
Center for Education Data & Research

January 2015

#CALDER2015

About the Project

- Applicants at the Doorstep: Improving Hiring Practices through a Better Understanding of the Link between Applicant Information and Teacher Quality

- Institute of Education Sciences funded project September 2013- September 2015

- Researcher-Practitioner Grant Program
  - Advance the relevance and usability of education research for the day-to-day work of education practitioners and policy makers
  - Build research capacity in practitioner settings and support the use of rigorous research-based evidence in decision-making
  - Foster greater use of state and district-level administrative data for education research

What Is Known About Applicant Selection Tools & Teacher Hiring?

- Some evidence suggests value in use of structured selection protocols (McDaniel, Schmidt, & Hunter, 1988)

- Mixed evidence about whether school systems hire the best available applicants (Ballou, 1997; Boyd et al., 2011, 2013; Hinrichs, 2013)

- Rockoff et al. 2011 is probably the most comprehensive study: some non-traditional information has been shown to be modestly predictive of teacher quality

    - "…districts may be able to gain some traction in selecting more effective teachers by using broader sets of information during recruitment…the variation of predicted value-added with an expanded set of data on new teachers has only about 12 percent of the variance of the expected distribution of teacher effectiveness."

AMERICAN INSTITUTES FOR RESEARCH®

Focus of this Study

- Analysis of the relationship between two teacher selection rubrics used during the teacher hiring process in Spokane Public Schools (SPS) and three teacher outcomes:

  1. Value-added measures of effectiveness

  2. Teacher absence behavior

  3. The likelihood of attrition

- We observe *all* applicants:

  - Allowing for assessment of applicants hired into non-SPS WA public schools

  - Allowing us to see if the rubric distinguishes between teachers SPS wants to hire and those SPS doesn't

Quick Questions and Answers

- Are the screening instruments associated with…

  - Teacher value added? Yes, one SD increase (7-8 points out of 60) in 60-point screening instrument associated with 0.06 SD increase in math, marginally significant 0.03-0.04 SD increase in reading

  - Teacher absences? No, not even close

  - Teacher attrition? Yes, one SD increase associated with 3 % point decrease

- 21-point score is less significant, but still has some predictive power

- There are some ways to improve the rubrics and their use

AMERICAN INSTITUTES FOR RESEARCH®

Data Overview

- Spokane Public Schools: applicant profile data, screening scores, record of job applications and status, and teacher absence data from 2008-09 to 2012-13 SYs (4 years of applicant data)

- Washington State OSPI

  - S-275 personnel records for certificated public school positions

  - PESB certification and licensure exam records

  - Core Student Record System: student achievement, demographic and assignment information

- NCES: school and district-level data

- WSIPC: teacher absence data for other districts (date and type)

The Applicants We Observe
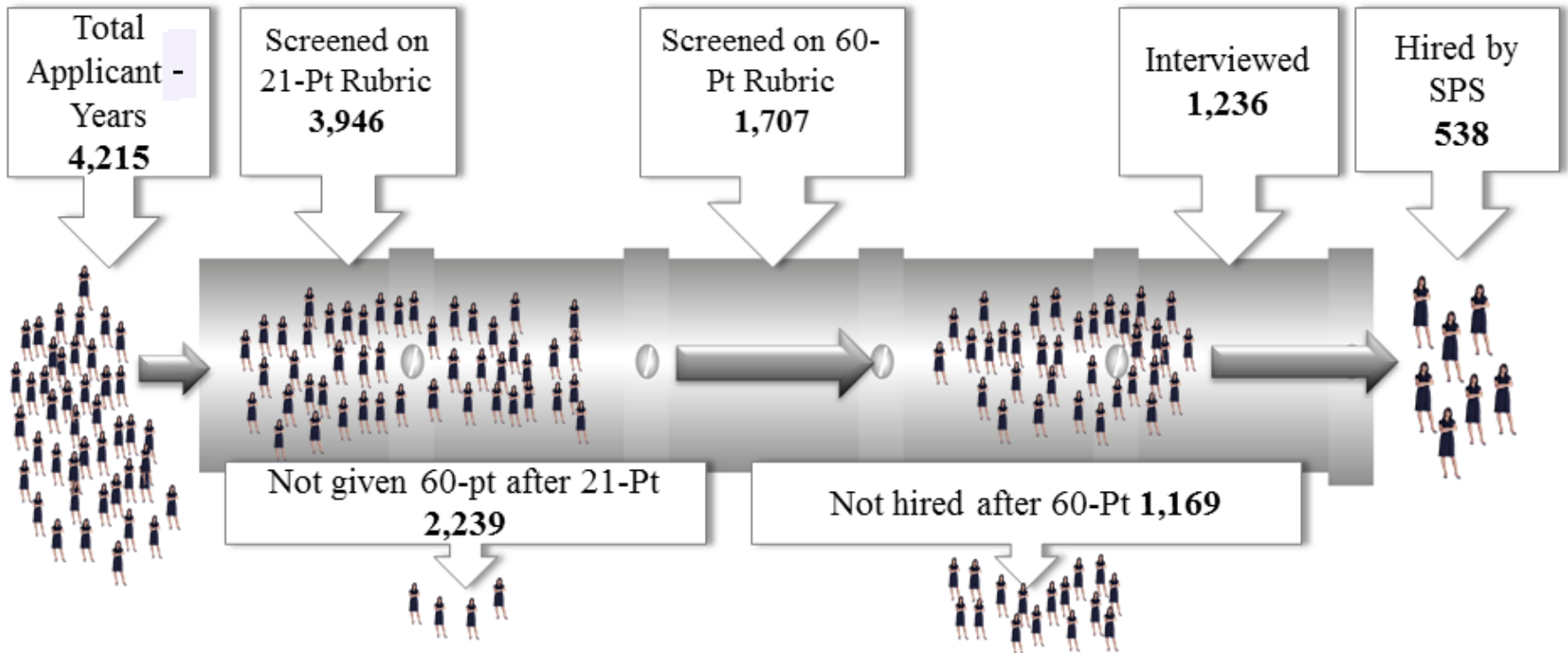
21-Point Pre-Screening Rubric
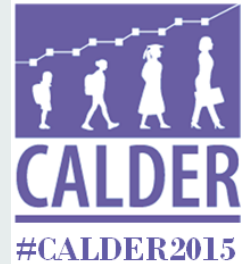
Applicant Name:	Position:	**CERTIFICATED**
PID:		
Date:	HR Official: Angela R. Brown	

NEW / RESCREENING / CORRECTION   Delete previous screening (if appl.):   YES / NO

HUMAN RESOURCES PRESCREENING

DOES APPLICANT MEET BASIC QUALIFICATIONS FOR THIS POSITION: YES / NO

Notes:

> 1 – 2 Some evidence to support this as an area of strength
> 3 – 4 Satisfactory evidence to support this as a area of strength
> 5 – 6 Strong evidence to support this as an area of strength

EXPERIENCE related to position   1   2   3   4   5   6

Notes:

Look fors: years of experience, type of experience, type of school/district, gaps in teaching experience

DEPTH OF SKILLS related to position   1   2   3   4   5   6

Notes:

Look fors: evidence of strong content knowledge, strong classroom management, differentiates instruction, engages parents/families, strong rapport with students and colleagues, commitment to the school as a community, socially just practices, experience with diverse student populations, makes learning relevant, engages students in active learning, elem level currently seeking those with experience using Fosnot, Calkins, GLAD strategies, Response to Intervention

QUALITY OF RECOMMENDATIONS   1   2   3   4   5   6

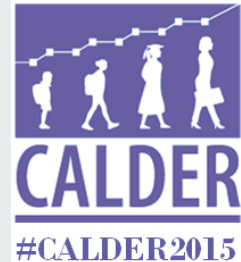Notes: X 1.5 =   1.5   3   4.5   6   7.5   9

Look fors: all items noted in above categories, does writer recommend/strongly recommend, personal or professional recommendation, does the writer regularly evaluate teachers (preference of letter from principal, asst principal, instructional coach, supt.)

TOTAL:

QUALIFIED TO SUBSTITUTE IN THIS AREA: YES / NO / NA

Other Notes:

54-Point Screening Rubric

CERTIFICATED APPLICANT - PRINCIPAL / SUPERVISOR SCREENING		
APPLICANT NAME:		
Job # / Position Title:		
Screened by:		
SCREENING TOTAL:	54	DATE:

SCREENING CRITERIA	RATING (1-6) 5 - 6 Strong evidence to support this as an area of strength 3 - 4 Satisfactory evidence to support this as an area of strength 1 - 2 Some evidence to support this as an area of strength	NOTES
CERTIFICATE AND EDUCATION		*Note completion of course of study; certificate held (current or pending); education*
Washington State Certificate	Yes / No	
Required Endorsement	Yes / No	
Rating (1 - 6)	6	
TRAINING		*Look for quality, depth and level of candidates additional training relating to the position.*
Rating (1 - 6)	6	
EXPERIENCE		*Note degree to which experience supports the prediction of success .... not just the number of years. A beginning candidate could be rated highly*
Rating (1 - 6)	6	
CLASSROOM MANAGEMENT		*Look for specific references to successful strategies. This may not mean quiet and orderly but planned and directed. Effectively handles large /small or ethnically/socioeconomically diverse groups; develops routines and procedures to increase learning, establishes clear parameters, and responds appropriately.*
Rating (1 - 6)	6	
FLEXIBILITY		*Note multiple endorsements, activity, coaching interests, student, building or district, or community support. Willing to learn new concepts and procedures, successfully teaches a variety of assignments, effectively uses various teaching styles.*
Rating (1 - 6)	6	
INSTRUCTIONAL SKILLS		*Look for specific references in support of skill in this area – plans, implements, evaluates, relates to students, creative, multiple approaches, monitors and adjusts, uses culturally responsive strategies appropriate to age, background and intended learning of students.*
Rating (1 - 6)	6	
INTERPERSONAL SKILLS		*Develops and maintains effective working relationships with diverse staff, students, parents/guardians, and community.*
Rating (1 - 6)	6	
CULTURAL COMPETENCY		*Look for specific references to successful strategies for building and maintaining a relationship with each student and their family. This may not be explicitly mentioned, but the following strategies offer some evidence of cultural competency: specific instructional strategies providing each student access to a rigorous curriculum, inclusive/respectful language about students and families, a belief that all children can achieve at high levels, mention of conflict resolution/restorative practices, specific instructional strategies for integrating culturally responsive materials which are also rigorous, and appropriate statements about their work with diverse populations. Note relevant training, course work, authors/book titles listed.*
Rating (1 - 6)	6	
PREFERRED QUALIFICATIONS AS INDICATED ON POSTING		
Rating (1 - 6)	6	

Pair-wise Correlations of Applicant Screening Rubric Total Adjusted Summative Rating and Individual Components
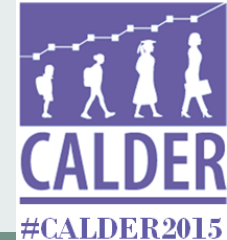
	21-Point Pre-Screening Rubric				60-Point Screening Rubric									
	Total	Exper.	Skills	Rec's	Total	Cert.	Train.	Exper.	Class.	Flex.	Instr.	Interp.	Cult.	Pref.
21-Point Pre-Screening Rubric														
Total Summative Rating	1.00													
Experience	0.57	1.00												
Skills	0.83	0.37	1.00											
Recommendations	0.86	0.22	0.71	1.00										
60-Point Screening Rubric														
Total Summative Rating	0.20	0.15	0.21	0.15	1.00									
Certificate & Edu.	0.03	0.05	0.02	0.03	0.41	1.00								
Training	0.13	0.14	0.13	0.09	0.72	0.21	1.00							
Experience	0.29	0.30	0.13	0.13	0.75	0.25	0.62	1.00						
Classrm. Mgmt.	0.21	0.08	0.24	0.18	0.77	0.21	0.44	0.49	1.00					
Flexibility	0.16	0.07	0.20	0.13	0.79	0.23	0.47	0.51	0.72	1.00				
Instructional Skill	0.23	0.13	0.24	0.17	0.83	0.21	0.55	0.60	0.74	0.69	1.00			
Interpersonal Skill	0.19	0.09	0.22	0.16	0.79	0.25	0.45	0.50	0.66	0.73	0.69	1.00		
Cultural Comp.	0.11	0.09	0.11	0.10	0.68	0.14	0.46	0.45	0.49	0.53	0.52	0.56	1.00	
Preferred Qual.	0.03	0.03	0.04	0.03	0.62	0.24	0.43	0.42	0.42	0.36	0.38	0.34	0.32	1.00
Letters of Rec.	0.11	0.02	0.11	0.08	0.44	0.08	0.18	0.24	0.24	0.31	0.33	0.34	0.25	0.21

Summary Statistics of Input Variables by Progression through Hiring Process

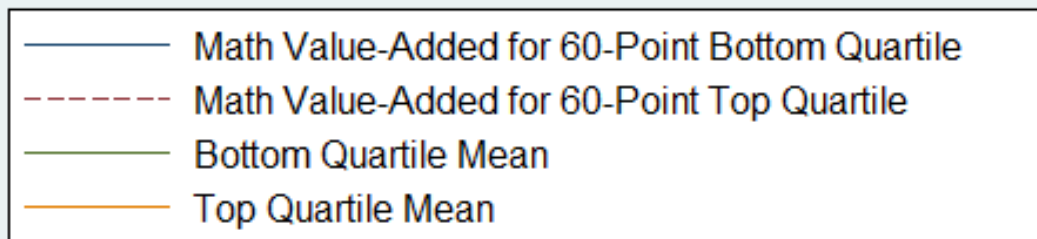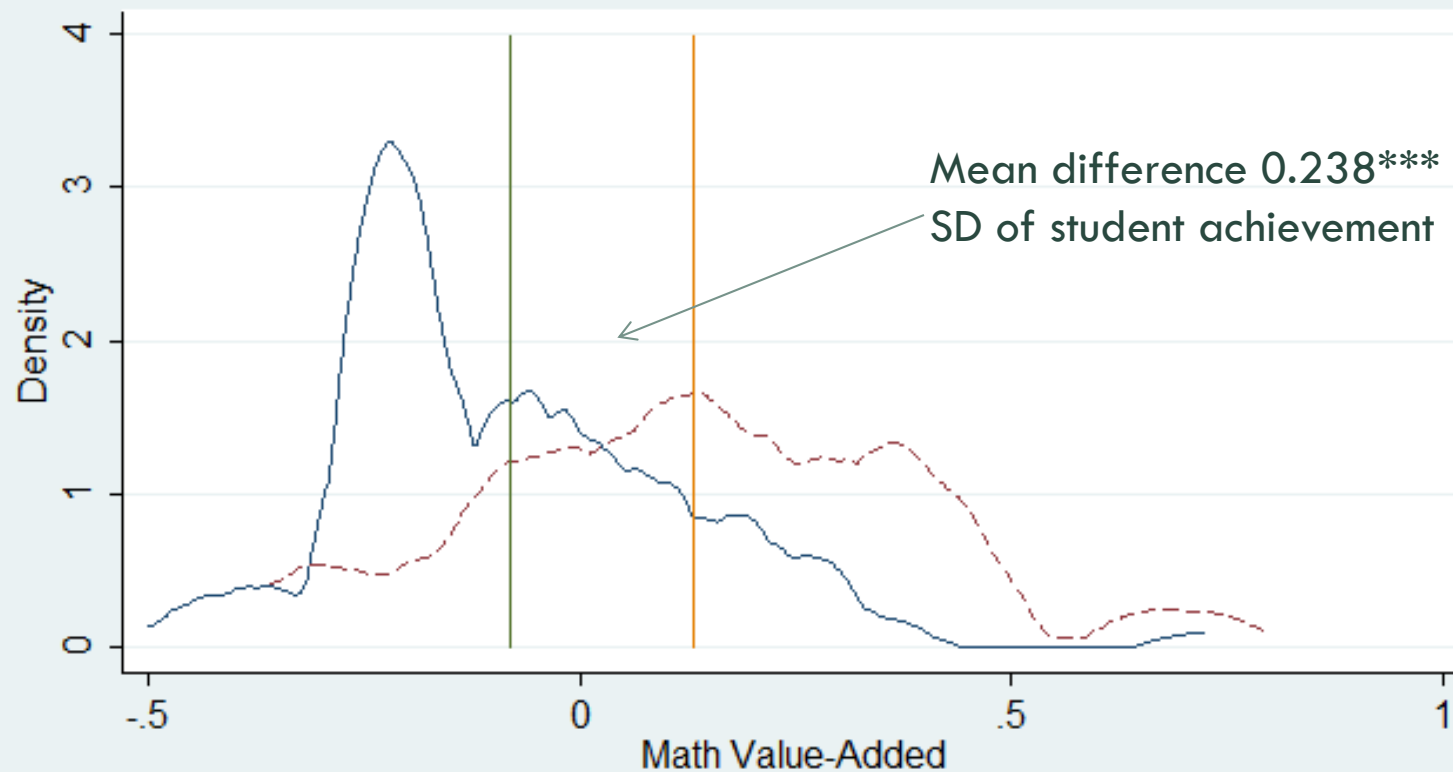|	All	21-Pt Pre-Screening Summ. Rating	60-Pt Screening Summ. Rating	Interview	Hired/ Offered	Hired Elsewhere
Total Observations	4,215	3,946	1,707	1,236	538	496
Total Proportions	1.00	0.94	0.40	0.29	0.13	0.12
Applicant Information						
Certificated Employment						
Experience in Year Applied						
No Experience	0.83	0.84	0.69	0.63	0.49	0.53
SPS District	0.11	0.09	0.22	0.28	0.43	0.03
Other District	0.07	0.07	0.09	0.09	0.08	0.44
Calculated Experience	3.18	3.22	3.85	3.70	3.23	4.41
Student Teaching in SPS? (Y/N)	0.36	0.37	0.40	0.42	0.47	0.29
21-Point Pre-Screening Rubric Summative Rating	NA	16.10 (2.38)	16.99 (2.22)	17.16 (2.20)	17.27 (2.20)	16.46 (2.24)
60-Point Pre-Screening Rubric Summative Rating	NA	NA	41.31 (7.29)	43.60 (6.13)	45.61 (5.75)	40.09 (6.76)
WESTB Average (Standardized over state)	-0.07 (0.75)	-0.07 (0.75)	-0.03 (0.74)	-0.02 (0.75)	0.02 (0.70)	-0.04 (0.75)

AMERICAN INSTITUTES FOR RESEARCH®

Overview of Empirical Approach

- Estimate relationship between screening instruments and teacher effectiveness in two steps:

  1. Estimate basic value-added models (covariates include prior test scores, student covariates, etc.)

  2. Regress value added measures against screening scores (21- and 60-point scores, and various subcomponents)

  - *Supplemental models account for sample selection*

- Estimate logit models of 1- and 3-year attrition probabilities

- Estimate linear models of days absent/year (& Monday and Friday absences)
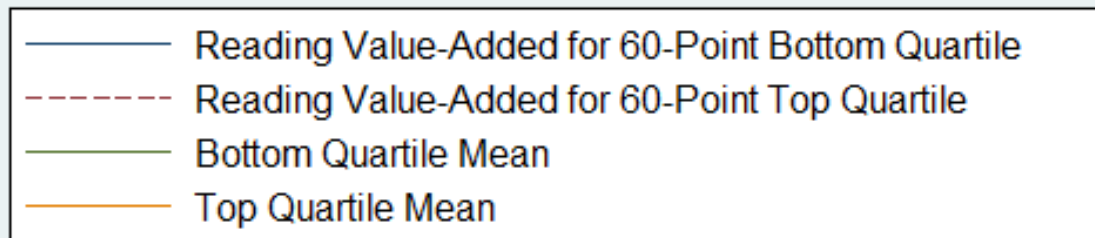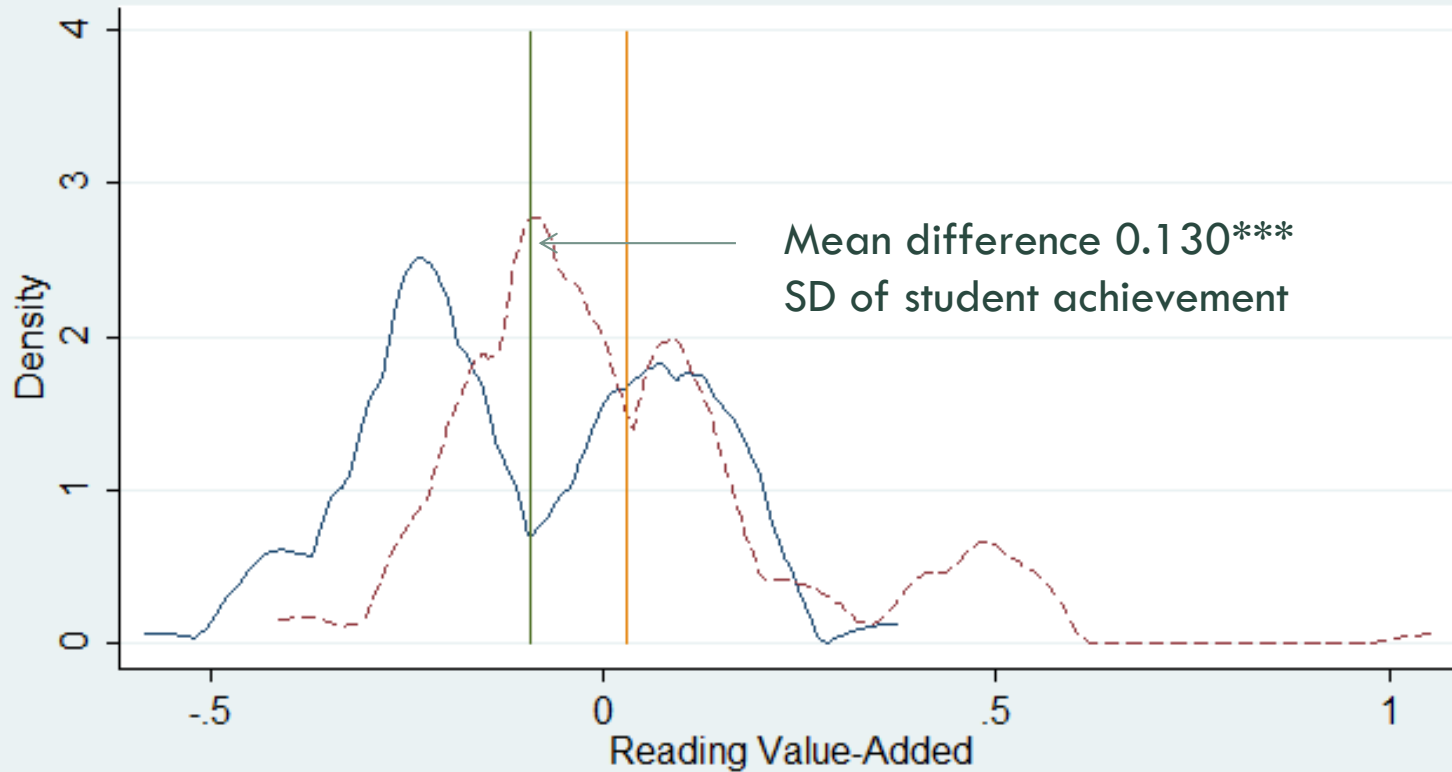
Predictors of Teacher Outcomes

Panel A	Math		Reading	
		School FE		School FE
(Specification 1)	N = 220 (184 clusters)		N = 229 (189)	
21-Point Score	0.032	0.022	0.024	0.016
	(0.022)	(0.019)	(0.015)	(0.018)
(Specification 2)	N = 152 (127)		N = 151 (126)	
60-Point Score	0.064**	0.039	0.033	0.048
	(0.028)	(0.031)	(0.025)	(0.029)
	Panel B: Annual Absences		Panel C: District Attrition	
		School FE		School FE
(Specification 1)	N = 453 (335)		N = 1,210 (617)	N = 1,020 (545)
21-Point Score	0.416	0.263	-0.019*	-0.018
	(0.300)	(0.326)	(0.011)	(0.013)
(Specification 2)	N = 287 (213)		N = 1,265 (633)	N = 1,063 (558)
60-Point Score	-0.083	-0.050	-0.030***	-0.037***
	(0.508)	(0.626)	(0.011)	(0.014)

Value Added Findings: Math

Mean difference 0.238***
SD of student achievement

Value Added Findings: Reading

Mean difference 0.130***
SD of student achievement

For Some Perspective

0.049 = average difference between 1st-year teachers and those with at least 4 years of experience

-0.031 = average for teachers in 1st year of teaching

+0.018 = average for teachers having at least 4 years experience

Student Achievement in Standard Deviation Units

Novice Teachers
Seasoned Teachers

Absences

Mean difference -3.69 days absent
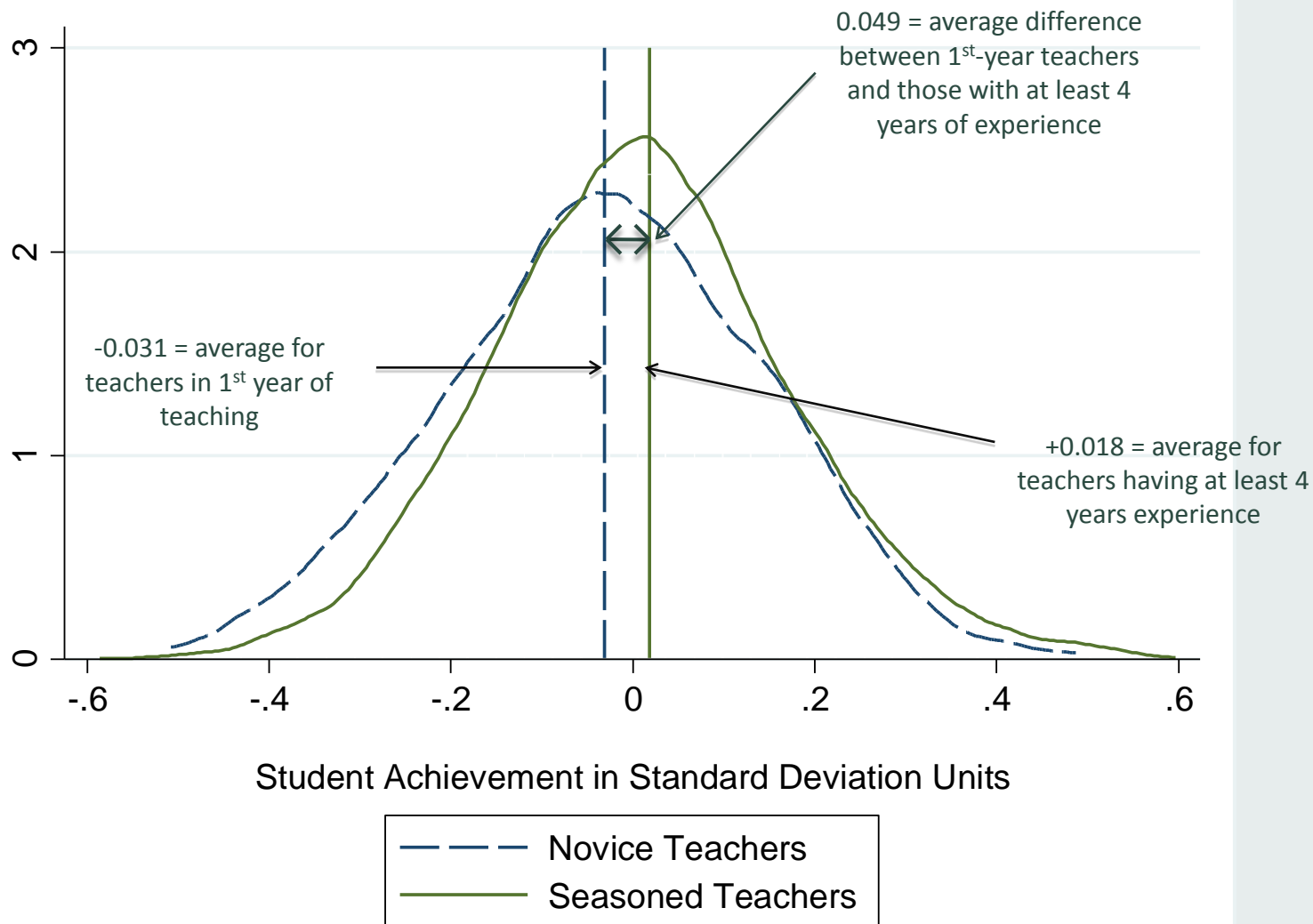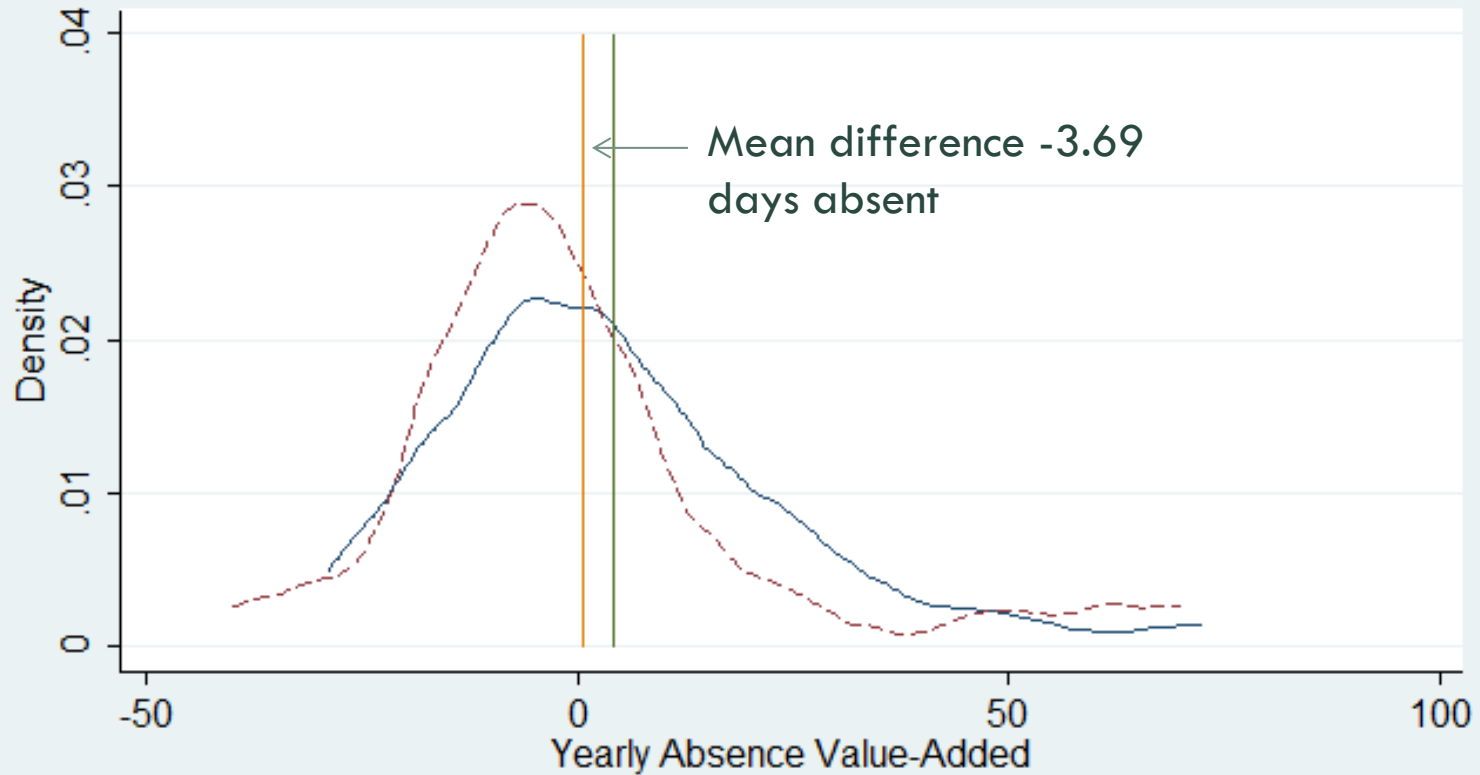
Legend:
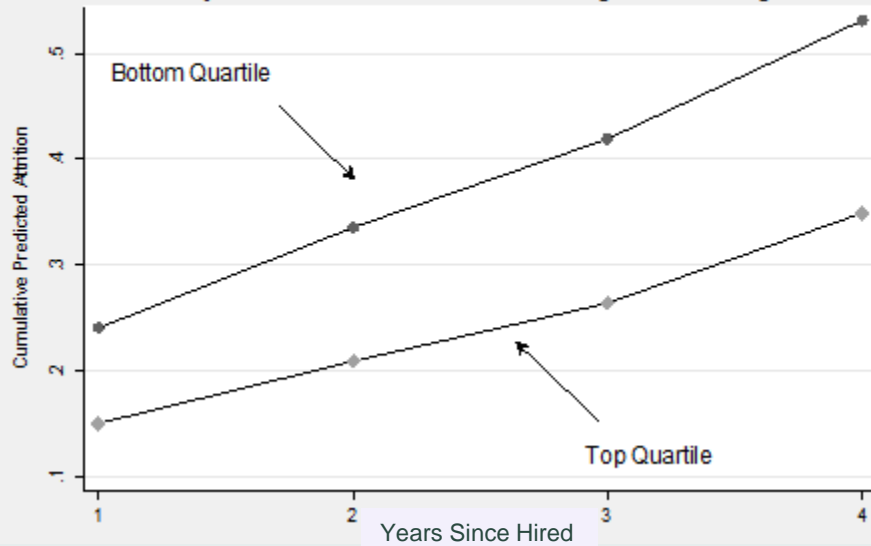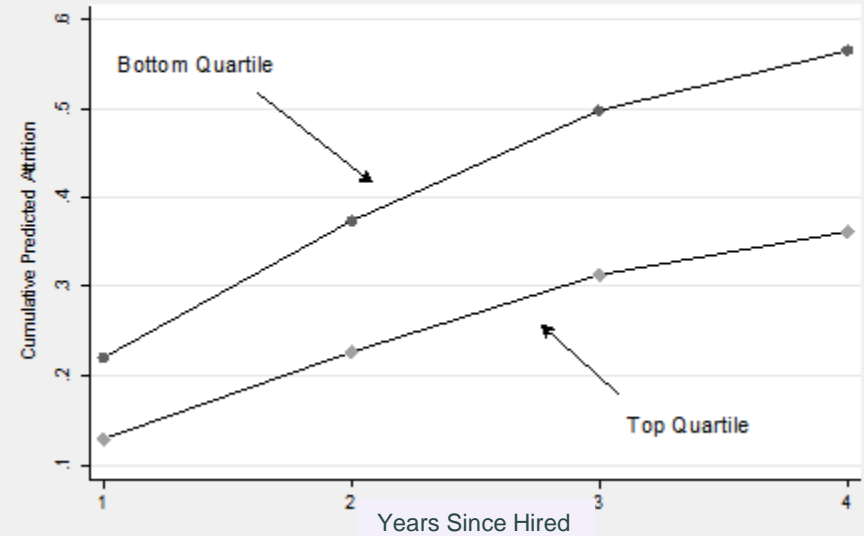- Yearly Absence Value-Added for 60-Point Bottom Quartile
- Yearly Absence Value-Added for 60-Point Top Quartile
- Bottom Quartile Mean
- Top Quartile Mean

Attrition of Top/Bottom Quartiles

Baseline 1-year attrition: 0.183
Baseline 3-year cumulative attrition: 0.313

Notes on Summative Results

- The 60-Pt score predicts large and significant effects for math achievement and district attrition

- Attrition results are similar if we instead look at attrition from the building or the profession

- Rubrics do somewhat better inside Spokane (as expected) and in middle school

- These results take the screening scores as they're used – the current system is capable of picking good teachers!

- But…

  - Should we trust the results?

  - How could we improve the rubrics?

AMERICAN INSTITUTES FOR RESEARCH®

Trusting the Results: Sample Selection
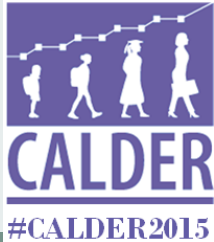
- Potential problem: we observe those who work in WA public schools (SPS and elsewhere), but not those who end up elsewhere

  - If there's something different about the qualities of these people that we can't measure, this will bias our results

- Sample Selection Correction model (Heckman, 1979)

  - We need variables that teacher hiring, but not performance: we use the quality of the competition, arithmetic errors in computing the 21-point scores

  - Selection models validate our findings

Conclusions

- SPS' hiring rubrics strongly predict student achievement and teacher attrition; findings are robust to sample selection correction

  - Suggests potential for meaningful improvements in teacher workforce and, perhaps, hints about teacher preparation (e.g. classroom management findings)

- There are ways for Spokane to improve the process of scoring/use of the subcomponents such that the relationship with the outcomes *we assess* is strengthened

  - Talking w/ Spokane about an experiment to test these

BACKUP SLIDES

Practical Improvements on Rubrics (1)

- Scorers make procedural errors in scoring applicants
    - Some subcomponents (Letters of Recommendation, Preferred Qualifications) skipped entirely for all applicants to a job
    - Erroneous skipped or 0-point subcomponent scores for particular applicants
- These issues could be avoided by better training or clearer instructions on rubric administration
- Or by a computerized entry system that doesn't allow errors

Practical Improvements on Rubrics (2)

- Computerization could also help cut down on arithmetic errors. We compare correctly added subcomponents to reported totals
  - 18.8% of 21-point scores are totaled incorrectly
    - Of these, 38% are addition errors
    - 7% forget to multiply Recommendations by 1.5
    - 56% make an error in multiplying
  - 8-9% of 60-point scores have addition errors
- Addition of decimal point for better differentiation

Procedural Improvements on Rubrics

- Because teachers apply to many schools each year, there is redundancy in scoring

  - About five *applicants* per job, but 135 *applications* per elementary job, 34 per non-elementary job

- Have multiple raters on the 21-point rubric (randomization of applicant to rater would allow testing of rater effects)

- If subcomponent scores are reliable and consistent across jobs (research on this is coming down the pike) and education levels, could save time by reusing all or some scores

Where Do the Rubrics Predict Best?

		Math	Reading	Annual Absences	District Attrition
		(1)	(2)	(3)	(4)
Main Model	60-Point Rubric Summative Score	0.064**	0.039	-0.083	-0.030***
		(0.028)	(0.031)	(0.508)	(0.011)
	Observations	152 (127)	151 (126)	287 (213)	1,265 (633)
	R-squared	0.162	0.089	0.169	0.073
Split Sample	**Inside Spokane**				
	60-Point Rubric Summative Score	0.068**	0.032	0.236	-0.027**
		(0.033)	(0.031)	(0.555)	(0.012)
	Observations	117 (94)	116 (94)	198 (146)	966 (469)
	R-squared	0.115	0.099	0.191	0.158
	Outside of Spokane				
	60-Point Rubric Summative Score	0.058	-0.024	-0.606	-0.018
		(0.047)	(0.035)	(1.119)	(0.030)
	Observations	35 (34)	35 (32)	89 (67)	299 (171)
	R-squared	0.542	0.214	0.272	0.266

AMERICAN INSTITUTES FOR RESEARCH®

Where Do the Rubrics Predict Best?

		Math	Reading	Annual Absences	District Attrition
Main Model	60-Point Rubric Summative Score	0.064** (0.028)	0.039 (0.031)	-0.083 (0.508)	-0.030*** (0.011)
	Observations	152 (127) [a]	151 (126)	287 (213)	1,265 (633)
	R-squared	0.162	0.089	0.169	0.073
Split Sample	**Elementary School**				
	60-Point Rubric Summative Score	0.044 (0.032)	0.022 (0.027)	-0.617 (0.595)	-0.016 (0.014)
	Observations	114 (94)	113 (93)	145 (106)	711 (348)
	R-squared	0.079	0.052	0.315	0.066
	Middle School				
	60-Point Rubric Summative Score	0.084* (0.049)	0.031 (0.056)	1.253 (1.172)	-0.062** (0.027)
	Observations	38 (34)	38 (36)	62 (47)	204 (109)
	R-squared	0.557	0.316	0.418	0.246
	High School				
	60-Point Rubric Summative Score			0.599 (0.866)	-0.032 (0.028)
	Observations			80 (63)	314 (174)
	R-squared			0.214	0.164

Table 1. Teacher Applicant Data

Applicant Background	Applicant Experience
Application Date	Student Teaching Experience
Full Name	Start Date
Employee ID	End Date
Address	State
Overall GPA	District
Race/Ethnicity	Subject
Degree Type	Title I Funded? (Y/N)
Degree Major	Teaching Experience
Degree College	Start Date
Degree Date	End Date
Degree Start Date	State
Degree End Date	District
Certification Area	Position Title
Certification Field	Subject
Certification Type	Reason Left
Certification State	Non-Teaching Experience
	Start Date
Pre-Screen Score (21 Pt Rubric)	End Date
Screening Score (54 Pt Rubric)	Company
	Title
	Hours
	City
	State
	Reason Left

Note: The applicant data accommodates multiple degrees, student teaching experiences, teaching experiences, and non-teaching experiences.

21- and 60-Point Rubrics Together

Panel A	Math		Reading	
		School FE		School FE
(Spec. 3) 21- and 60-Point Scores	N = 130 (106)		N = 128 (104)	
21-Point Score	0.016	0.014	0.029	0.015
	(0.025)	(0.028)	(0.021)	(0.028)
60-Point Score	0.030	0.000	0.003	0.008
	(0.035)	(0.039)	(0.032)	(0.044)
	Panel B: Annual Absences		Panel C: District Attrition	
		School FE		School FE
(Spec. 3) 21- and 60-Point Scores	N = 272 (205)		N = 1,092 (560)	N = 912 (491)
21-Point Score	0.415	-0.138	-0.026**	-0.023
	(0.699)	(0.713)	(0.012)	(0.015)
60-Point Score	-0.126	0.146	-0.027**	-0.032**
	(0.534)	(0.654)	(0.013)	(0.016)

Errors and Competition as Hiring Predictors

	Hired	Placebo (Hired Elsewhere)
21-Pt Screen	0.008***	0.028***
	(0.001)	(0.008)
Excluded Variables:		
Error in Teacher's Favor	0.014***	0.002
	(0.002)	(0.017)
21-Pt Screen Competition	-0.012***	-0.005
	(0.002)	(0.014)
Observations	41,866 (3,937) [a]	18,236 (1,329)
F(Excluded Variables)	103.58***	0.938

Improving the Rubrics: Subcomponents

21-Point Subcomponents	Math	Reading	Yearly Absences	District Attrition
	N = 220 (184)	N = 229 (189)	N = 453 (335)	N = 1,210 (617)
Experience	0.013	0.009	0.410*	-0.075
	(0.020)	(0.015)	(0.246)	(0.102)
Skills	0.020	0.009	0.102	-0.154*
	(0.019)	(0.014)	(0.297)	(0.086)
Recommendations	0.041*	0.029*	0.186	-0.132
	(0.023)	(0.016)	(0.264)	(0.094)

Each coefficient is from its own regression

AMERICAN INSTITUTES FOR RESEARCH®

Improving the Rubrics: Subcomponents

60-Point Subcomponents	Math	Reading	Yearly Absences	District Attrition
	N = 152 (127)	N = 151 (126)	N = 287 (213)	N = 1,265 (633)
Certificate & Education	0.025	-0.000	0.298	0.003
	(0.040)	(0.029)	(0.549)	(0.012)
Training	0.062**	0.040	0.150	-0.020*
	(0.030)	(0.025)	(0.547)	(0.012)
Experience	0.037	0.010	1.116**	-0.028**
	(0.034)	(0.027)	(0.453)	(0.011)
Classroom Management	0.131**	0.043*	-0.291	-0.025**
	(0.032)	(0.026)	(0.500)	(0.010)
Flexibility	0.090**	0.032	-0.175	-0.026**
	(0.032)	(0.029)	(0.602)	(0.011)
Instructional Skills	0.059*	0.033	-0.410	-0.031***
	(0.033)	(0.026)	(0.574)	(0.011)
Interpersonal Skills	0.037	0.010	-0.553	-0.037***
	(0.037)	(0.028)	(0.474)	(0.011)
Cultural Competency	0.016	-0.004	-0.004	-0.012
	(0.026)	(0.023)	(0.477)	(0.011)
Preferred Qualifications	0.028	0.041	0.369	-0.025**
	(0.031)	(0.026)	(0.644)	(0.012)
Letters of Recommendation	-0.062	-0.070**	-0.297	-0.009
	(0.045)	(0.023)	(0.425)	(0.013)

Each coefficient is from its own regression

AMERICAN INSTITUTES FOR RESEARCH®

Improving the Rubrics: Subcomponents

- ## Subcomponents reweighted to maximize outcomes

Component Weights Which Maximize Correlation with Outcomes				
	Math Value-Added	Reading Value-Added	Absences	1-Yr District Attrition
60-Point Rubric Component Weights — Certificate & Education	0.109	0.139	0.286	0.171
Training	0.102	0.127	0.063	0.053
Experience	0	0	0	0.069
Classroom Management	0.512	0.158	0	0.068
Flexibility	0	0.117	0.034	0.022
Instructional Skills	0.043	0	0.031	0
Interpersonal Skills	0.075	0.025	0	0.236
Cultural Competency	0	0	0.141	0
Preferred Qualifications	0.009	0.354	0.060	0.178
Letters of Recommendation	0.150	0.080	0.385	0.203
Coefficient of Weighted 60-Pt Screening Score	0.136** (0.042)	0.051 (0.036)	-0.545 (0.570)	-0.049** (0.022)
Standard (equally)-weighted Model Coefficients	0.064** (0.028)	0.033 (0.025)	-0.083 (0.508)	-0.030*** (0.011)

AMERICAN INSTITUTES FOR RESEARCH®