Helping Districts Hire Better Teachers

A Case Study Evaluating the Effectiveness of a District's Teacher Hiring Practices

Mollie Rudnick

This document was submitted as a dissertation in December 2015 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Matthew W. Lewis (Chair), Laura S. Hamilton, and Jennifer L. Steele.

PARDEE RAND GRADUATE SCHOOL

For more information on this publication, visit http://www.rand.org/pubs/rgs_dissertations/RGSD365.html

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2016 RAND Corporation

RAND® is a registered trademark

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.html.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Abstract

This dissertation looks at the teacher hiring practices in Huntsville City Schools (HCS), Alabama. The research uses a combination of qualitative and quantitative methods to answer the following questions: 1) What are Huntsville City Schools' (HCSs') current hiring practices?; 2) How well does Huntsville City Schools' existing interview process for rating teacher candidates predict the effectiveness of new hires?; and 3) What are the costs and advantages of alternative instruments as compared to the current process? Both the literature on teacher hiring practices and teacher effectiveness and the findings from the quantitative analyses suggest that there is not one clear measure or tool that predicts later teacher effectiveness. Consequently, I also explore the trade-offs across various hiring measures and teacher characteristics that might inform district decision making about teacher hiring. While this research is specific to HCS, the findings can be broadly applied to other schools and districts with an interest in changing their hiring processes.

Table of Contents

Abstract.....	iii
Figures.....	ix
Tables.....	xi
Acknowledgments.....	xiii
Abbreviations.....	xv
1. Introduction.....	1
Policy Issue.....	1
Context: Huntsville City Schools.....	2
Research Aims and Questions.....	2
2. Background/Literature Review.....	4
Introduction.....	4
Acquire Teachers.....	5
Partner with Teacher Suppliers Responsive to District.....	5
Detect and Forecast Personnel Needs.....	6
Build a Pool of High Quality Applicants.....	7
Teacher Screening Methods.....	8
Interviews.....	11
Written Documentation.....	11
Observations.....	12
Extent of Centralization.....	12
Commercial Teacher Interviews.....	13
Teacher Characteristics.....	14
Teaching Practice.....	17
Content Knowledge and Academic Qualifications.....	17
Interpersonal Skills.....	18
General Personal Attributes.....	18
Attitudes, Beliefs, and Values.....	18
Match based on Content, Grade, Pedagogy, Connection to Community.....	19
Teacher Evaluation.....	19
Standardized Test Scores (Student Growth Percentiles).....	20
Classroom Observation Scores.....	21
Principal Ratings.....	22
Student Surveys.....	23
Attendance.....	23
Discipline Referrals.....	23
3. Methods.....	25
Data sources.....	25

Qualitative Methods	25
What are Huntsville City Schools' (HCS) current hiring practices?	25
Quantitative Methods	26
How well does HCS's existing interview process for rating teacher candidates predict the effectiveness of new hires?	26
Matrices of costs and advantages of instruments	30
What are the costs and advantages of alternative instruments as compared to the current process? ..	30
4. RQ1: Background on Huntsville City Schools	31
Introduction	31
Huntsville City Schools (HCS) Context	31
Huntsville City Schools' Hiring Process	33
Prior Hiring Process	33
Background on Hiring Process	33
Determining Personnel Needs in the District	34
Personnel Involved in Hiring Process	35
Application Process	35
Description of Interview Process	36
Principal Role in Hiring Process	38
Descriptive Statistics on Interview Process	40
Huntsville City Schools' Teacher Evaluation Process	41
Background on Teacher Evaluation Process	41
Description of Teacher Evaluation Process	42
Descriptive Statistics on Teacher Evaluation Process	48
Teacher Retention	56
Conclusion	56
5. RQ2: Quantitative Analyses	58
Introduction	58
Overall Interview Score and Overall Probationary Review Score	58
Interview Score and Year 1 Review	58
Interview Score and Year 2 Review	60
Overall Interview Score and Retention in Year 2	62
Overall Interview Score and Probationary Review Components	64
Principal Ratings	64
Observation Scores	65
Attendance	65
Disciplinary Actions	66
Student Survey	67
SGP Scores	67
Teacher Interview Factors and Teacher Effectiveness Outcomes	68
Factor Analysis	68
Relationship between Revised Interview Score and Outcomes	69
Conclusion	69

6. RQ3: Alternatives to HCS’ Current Hiring Process	71
Introduction	71
Teacher Screening Tools and Associated Teacher Characteristics	72
Screening Tools Huntsville City Schools (HCS) is Already Using or Has Access to	73
Instruments HCS Is Not Currently Using.....	80
Conclusion.....	82
7. Conclusion and Recommendations.....	84
Improvements to HCS Teacher Hiring Process.....	84
Tailor rubrics to each interview question	84
Identify questions and characteristics of interest.....	85
Field and test questions to see how they work	85
Improvements to HCS Teacher Evaluation Process.....	86
Improve scoring for probationary teacher review components	87
Track and record actual scores on probationary review components.....	87
Consider the weighting of the elements of the probationary review process	88
Find alternative measures for teachers who do not receive all review components.....	88
Use teacher evaluation data to inform the hiring process.....	89
Suggestions for Future Data Collection in HCS.....	89
Track teacher retention and conduct exit interviews to better understand attrition.....	89
Track data on all candidates	90
References.....	93
Appendix A. Interview Protocol for March 2014 Site Visit.....	101

Figures

Figure 2.1 Teacher Human Capital Framework taken from Myung, et al., 2013.....	5
Figure 4.1 Process map depicting HCS screening process	34
Figure 4.2 Distribution of scores on interview questions	41
Figure 4.3 Point distribution of 2012-2013 probationary teacher review.....	43
Figure 4.4 Point distribution of 2013-2014 probationary teacher review.....	43
Figure 4.5 Distribution of retention recommendations on probationary teacher review	49
Figure 4.6 Distribution of principal rating scores on probationary teacher review	50
Figure 4.7 Distribution of attendance scores on probationary teacher review	51
Figure 4.8 Distribution of discipline scores on probationary teacher review	52
Figure 4.9 Distribution of points on probationary teacher review from SGP scores.....	53
Figure 4.10 Distribution of student survey scores on probationary teacher review	54
Figure 4.11 Distribution of observation scores on probationary teacher review.....	55
Figure 5.1 Standardized Year 1 probationary review scores and interview scores by cohort	59
Figure 5.2 Interview score percentage and Year 2 probationary review scores	61
Figure 5.3 Interview score percentage and Year 1 probationary review scores by retention	63
Figure 7.1 Revised process map for HCS hiring process	91

Tables

Table 1.1 Audiences for research findings and areas of impact	3
Table 2.1 Overview of Descriptive Studies of Teacher Hiring Practices within Districts	9
Table 2.2 Tools used in the teacher screening process	10
Table 2.3 Personnel with whom teacher candidates interviewed	11
Table 2.4 Characteristics districts in descriptive studies used to select candidates.....	16
Table 2.5 Negative characteristics districts in descriptive studies used to rule out candidates	17
Table 4.1 Descriptive statistics by major teacher suppliers.....	32
Table 4.2 Interview questions and “look-fors” by subject areas and grade level	38
Table 4.3 Distribution of scores on interview questions.....	40
Table 4.4 Rubric for assigning points for principal ratings in the teacher review.....	45
Table 4.5 Rubric for assigning points for absences in the teacher review.....	45
Table 4.6 Rubric for assigning points for discipline referrals in the teacher review	46
Table 4.7 Rubric for assigning points for SGP scores in the teacher review.....	46
Table 4.8 Rubric for assigning points for student survey scores in the teacher review.....	47
Table 4.9 Rubric for assigning points for observations in the teacher review.....	48
Table 4.10 Frequencies of principal rating scores on the probationary teacher review	50
Table 4.11 Frequencies of attendance scores on the probationary teacher review.....	51
Table 4.12 Frequencies of discipline scores on the probationary teacher review	52
Table 4.13 Descriptive statistics on available SGP scores	52
Table 4.14 Frequencies of points on the probationary teacher review from SGP scores	53
Table 4.15 Frequencies of student survey scores on the probationary teacher review.....	54
Table 4.16 Frequencies of observation points on the probationary teacher review.....	55
Table 5.1 Coefficients (and standard errors) from of Year 1 probationary review scores on candidate interview scores	60
Table 5.2 Coefficients (and standard errors) from OLS regressions of Year 2 probationary review scores on candidate interview scores	62
Table 5.3 Odds ratios (and standard errors) from logistic regressions of Year 2 retention on candidate interview scores	64
Table 5.4 Odds ratios (and standard errors) from ordinal logistic regressions of principal ratings on candidate interview scores, by teaching year.....	65
Table 5.5 Odds ratios (and standard errors) from ordinal logistic regressions of observation scores on candidate interview scores, by teaching year.....	65
Table 5.6 Odds ratios (and standard errors) from ordinal logistic regressions of attendance scores on candidate interview scores, by teaching year.....	66

Table 5.7 Odds ratios (and standard errors) from ordinal logistic regressions of disciplinary scores on candidate interview scores, by teaching year.....	66
Table 5.8 Odds ratios (and standard errors) from ordinal logistic regressions of student survey scores on candidate interview scores, by teaching year.....	67
Table 5.9 Odds ratios (and standard errors) from ordinal logistic regressions of SGP Scores on candidate interview scores, by teaching year.....	67
Table 5.10 Correlation matrix between interview questions	68
Table 5.11 Overall and item-level alpha coefficients	69
Table 6.1 Matrix of costs and advantages for teacher selection tools	72
Table 6.2 Characteristics measured by each selection tool.....	73

Acknowledgments

I would like to extend great appreciation to my dissertation committee: Matthew Lewis, Laura Hamilton, and Jennifer Steele. In addition to reading and commenting on countless drafts, they also provided invaluable guidance on both my dissertation work and conducting research in general. This guidance has served me well during my time at Pardee RAND and will continue to do so in my future career.

This dissertation would not have been possible without the support of Huntsville City Schools (HCS) for providing both financial support and access to the data to conduct this research. Specifically, I would like to thank the HCS superintendent, Casey Wardynski, for his support of this project and for serving in the role of outside reader of my dissertation. I would also like to thank the district personnel who provided ongoing support and access: Debbie Miser, Matthew Creel, Cathy McNeal, and Linda Talley.

This dissertation was also supported by the Doris Dong, Anne and James Rothenberg, and Susan Way-Smith Dissertation Awards. I am thankful for these awards and the work that they enabled me to accomplish and would like to thank the donors and the Pardee RAND Graduate School.

Finally, I would like to thank the faculty, administration, and staff at Pardee RAND Graduate School. The last four years have been an invaluable experience that I could not have received anywhere else. The combination of classroom learning and practical application has provided me with a set of tools and way of thinking that will serve to influence my future work.

Abbreviations

AYP	Adequately Yearly Progress
GAO	Government Accountability Office
HBCU	Historically Black College or University
HCS	Huntsville City Schools
OLS	Ordinary Least Squares
TFA	Teach For America

1. Introduction

Policy Issue

Students in the United States, particularly low-income and minority students, are struggling academically (Gardner, 1983; Hanushek, Peterson, & Woessmann, 2011, 2012; Ladson-Billings, 2006). The issue is not just about struggling students; many of these students are concentrated in schools that are struggling to raise achievement. A 2007 Government Accountability Office (GAO) report found that approximately eight percent of Title I schools failed to make Adequate Yearly Progress (AYP) on annual benchmarks for school improvement for more than four years and that these schools tended to have high percentages of low-income and minority students than other Title I schools (Government Accountability Office, 2007). While there are numerous factors that policymakers have implemented to improve academic performance among students and schools (e.g., teacher-student ratio, school funding, length of the school day, use of educational technologies in the classroom), research has shown that teacher quality plays a large role in influencing student outcomes (Nye, Konstantopoulos, & Hedges, 2004; Rockoff, Jacob, Kane, & Staiger, 2011). Given that teachers make a significant difference, placing effective teachers in schools that are struggling provides a promising avenue for raising student achievement, especially among the lowest performing schools and students. In fact, studies have found that high quality teachers have a greater effect on student achievement at low income schools than they do at high income schools because of larger differences in the quality of teachers at low income schools (Nye et al., 2004). Yet, in spite of the potential of quality teachers to help low achieving students, there are fewer experienced, fully certified teachers with degrees in the content area they are teaching at low achieving, disadvantaged schools than there are at more affluent, high achieving schools (Akiba, LeTendre, & Scribner, 2007; Loeb, Kalogrides, & Béteille, 2012; Peske & Haycock, 2006). On the other hand, a recent study conducted in four school districts across the US found that teachers with more low-income and minority (LIM) students tended to have higher value-added scores, but that access to high value-added teachers varied more within schools than across schools (Steele, Baird, Engberg, & Hunter, 2014). Taken together, these studies indicate that teachers can have an impact on student achievement, particularly if distributed in a way that will help all students.

Giving all students access to quality teachers to help them succeed, especially students who could benefit the most, is part of enacting the U.S. Department of Education's commitment to providing all students with access to a high quality education (Office of Civil Rights, 2012). In order to provide such access to quality teachers, administrators must be able to determine which teachers are most effective. While test scores and observations of instructional practice are frequently used by researchers and policymakers to judge teacher quality, schools and districts

do not have such information when novice teachers enter the system and district and school administrators have to rely on a variety of other tools to measure characteristics of teacher candidates (Rutledge, Harris, Thompson, & Ingle, 2008). Yet there is hope that research can yield indicators and measures that will predict future teacher effectiveness. Therefore, this study focuses on the teacher hiring process in one school district, which can help inform the research on teacher hiring more broadly.

Context: Huntsville City Schools

This research is being conducted on the Huntsville City Schools (HCS), a school district in Huntsville, Alabama. HSC has approximately 1,700 teachers and 23,374 students across 39 schools¹. The school district is undertaking many innovative programs, including a digital one-to-one initiative in which all students in the district have a tablet (grades k-2) or laptop (grades 3-12) and use digital curricula. The district is also focused on using data to inform practice. The directors of the various departments at the district-level (e.g., Talent Management, Assessment and Accountability, Transportation, IT) hold weekly meetings in which they present work their department has done, along with any relevant data, and receive input and guidance from others about future actions. The status of the teacher-hiring process, among other topics, are discussed during these meetings. Similarly, principals meet with the superintendent and district staff to discuss the performance data of their teaching staff on a regular basis.

HCS also has large disparities in achievement across its schools and has been under a desegregation order by the U.S. Department of Justice since 1965. Initial conversations with the HCS Talent Management Department indicated that they didn't explicitly take the desegregation order into consideration when hiring teachers and felt that improving their hiring process, selecting the most effective teachers, and placing them into schools as positions arose would address the issue. Fifteen of the 39 HCS schools are Title I schools that receive additional federal funding because they have high percentages of students from low-income families.

Research Aims and Questions

While the results of this study will provide specific guidance to HCS, the results will, more importantly, fill the gap in the existing research by providing information about which measures and characteristics of prospective teachers are related to later indicators of teacher effectiveness, and consequently, which measures and characteristics should be included in the hiring process. Furthermore, identifying these teacher hiring characteristics and measures will inform decisions about teacher placement in schools with the most need. By recognizing which teachers have the most potential to help students achieve, policymakers and administrators, both nationally and

¹ <http://www.huntsvillecityschools.org/?PN=AboutUs>

within HCS, can focus on most equitably distributing quality teachers to schools. Moreover, given the lack of clarity in the best methods for teacher hiring, I will also explore the costs and advantages of various teacher screening methods, which can aid school districts in making choices about how to structure their hiring processes, given their specific needs. Table 1 lists the audiences for the findings of this research, along with the areas of concern for each audience that the research will inform.

Table 1.1 Audiences for research findings and areas of impact

Target Audience	Possible Areas of Impact
HR personnel in school districts	<ul style="list-style-type: none"> • Methods for evaluating and choosing new teachers • Placement of teachers across schools to raise achievement at low-performing schools
Faculty/staff at schools of education	<ul style="list-style-type: none"> • Emphasis of skills in teacher education programs and curriculum
Parent Teacher Associations (PTAs)	<ul style="list-style-type: none"> • Who will sit on hiring committees/contribute to hiring process • New teacher qualifications/hiring process requirements
Teacher Unions	<ul style="list-style-type: none"> • Who will sit on hiring committees/contribute to hiring process
Researchers	<ul style="list-style-type: none"> • Hiring measures and teacher characteristics needing further research

This research will fulfill these goals by focusing on the following three research questions:

1. What are Huntsville City Schools' (HCSs') current hiring practices?
2. How well does Huntsville City Schools' existing interview process for rating teacher candidates predict the effectiveness of new hires?
3. What are the costs and advantages of alternative instruments as compared to the current process?

First, I will provide background on the literature on teacher hiring and the literature on teacher effectiveness, as it relates to the teacher evaluation process in HCS (Chapter 2). I will then detail the methods used to answer the three research questions (Chapter 3). Chapters 4, 5, and 6 will present findings from each of the three research questions. Chapter 4 answers Question #1 by providing an overview of the teacher hiring and evaluation processes used in HCS, along with descriptive statistics for each of those processes. Chapter 5 answers Question #2 by providing the results of the quantitative analyses of data from HCS' teacher hiring and evaluation processes. Chapter 6 addresses Question #3 by exploring the costs and advantages of potential teacher screening methods that HCS and other districts might use to hire teachers. Finally, I present specific recommendations for HCS, along with broader conclusions that are applicable to the teacher hiring process (Chapter 7).

2. Background/Literature Review

Introduction

This chapter reviews the literature on teacher hiring and teacher evaluation. Since teacher hiring processes are positioned within a much larger human resource process, this review is guided by a Teacher Human Capital Framework, shown below in Figure 1 (Myung, Martinez, & Nordstrom, 2013, p. 8). That framework entails four components, which feed into the ultimate aim of creating a stronger teacher workforce: 1) acquire, 2) develop, 3) sustain, and 4) evaluate teachers. This review focuses on the first component, acquiring teachers and the components within that process. After presenting the framework, I will review the literature on teacher hiring. Since this study uses probationary teacher review data as the outcome of interest, the review then includes a brief summary of the teacher evaluation and effectiveness literature. The other two components, developing and sustaining teachers, while equally important to a stronger teacher workforce, are beyond the scope of this review.

Figure 2.1 Teacher Human Capital Framework taken from Myung, et al., 2013

Acquire Teachers

The Human Capital Framework identifies five components within the realm of acquiring teachers: 1) Partner with teacher suppliers responsive to district needs (recruitment); 2) Detect and forecast personnel needs; 3) Build a pool of high quality applicants (incentives); 4) Enact strong hiring practices (hiring); and 5) Match based on content, grade, pedagogy, connection to community (candidate fit) (Myung et al., 2013). The review below is guided by those five components and the research relevant to each of those areas is summarized.

Partner with Teacher Suppliers Responsive to District

Increasing the pool of applicants, particularly highly qualified applicants, is essential to increasing the likelihood of finding and identifying effective teachers. As such, a portion of the literature on teacher hiring practices has focused on recruitment. Some studies have looked at recruitment practices of districts. Balter and Duncombe, in surveying superintendents across New York State, found that most districts work with local colleges in a variety of capacities (e.g.,

supervising student teachers, posting job notices, talking to college faculty) in addition to using a variety of other methods to increase the local teacher supply (2005). They later explored the relationship between the other methods to increase the local teacher supply (i.e., use of substitute teachers, retired teachers, alternatively certified teachers, or paraprofessionals) and teacher quality (i.e., a rating comprised of certification test scores, certification status, and ranking of college attended) and found that these district efforts were statistically insignificant, but negatively related to teacher quality within the district (Balter & Duncombe, 2006).

While there is little to no research on whether or not these recruitment practices are effective, more attention has been focused on the relative effectiveness of various teacher preparation programs. In fact, the National Council on Teacher Quality publishes annual rankings of teacher preparation programs in the United States based on 19 standards for teacher quality (Greenberg, McKee, & Walsh, 2014). Information on the effectiveness of specific teacher preparation programs, including traditional programs at universities and alternative certification programs such as Teach for America, can help inform which programs schools and districts partner with. Such studies produce varied findings. Some indicate that traditional teacher preparation programs only account for small differences in teacher effectiveness (Koedel, Parsons, Podgursky, & Ehlert, 2012). Others suggest that the selectivity of the university a teacher attended has some impact on teacher effectiveness, which could be attributed to a variety of other factors, including the cognitive ability of the teacher (Rice, 2003). In the realm of alternative certification programs, one study found that the students of Teach for America teachers, when compared with teachers prepared through traditional programs, have lower achievement in math and reading (Lackzo-Kerr, 2002), while others found that they have higher achievement in either reading or math, but not both (L. Darling-Hammond, Holtzman, D.J., Gatlin, S.J., and Vasquez Heilig, J., 2005; Glazerman, 2006; Kane, Rockoff, & Staiger, 2008; Raymond, 2001). Given the great variability in findings across the different types of teacher preparation programs, more research needs to be done on both the effectiveness of potential suppliers of teachers, as well as the models that the programs use.

Detect and Forecast Personnel Needs

Timing plays an important role in the teacher hiring process. Districts cannot begin the process of hiring teachers for specific positions until they become aware of personnel needs. Even after vacancies are determined, the district must carry out the often time consuming screening and hiring process. Throughout this process, as time goes on and teachers begin to get positions, the pool of candidates begins to deplete. Consequently, the quality of teacher candidates may dwindle as the higher quality candidates receive job offers earlier in the year. Despite this issue, studies that look at teacher hiring practices have found that many districts hire teachers late in the year (i.e., during late summer or after the school year begins) (Liu & Johnson, 2006; Wise, Darling-Hammond, Berry, & Berliner, 1987).

Wise, Darling-Hammond, Berry, and Berliner conducted case studies of the teacher selection process in six U.S. school districts and found that frequently, the logistics of the hiring process appeared to influence the quality of the staff hired as much as the actual screening mechanisms used (1987). Liu and Johnson suggest that the timing of the hiring process affects the ability of the districts to conduct an in-depth screening process. The authors surveyed a random sample of newly hired teachers in a random sample of schools across four states. In two of the states, one in three teachers was hired after the start of the school year. In three of the four states, the majority of teachers were hired within 30 days of when they would begin teaching. Liu and Johnson also noted that the state with the lowest percentage of late hires also had teachers who reported getting a more accurate picture of their school, indicating that the earlier timeframe of hiring teachers allowed for a more thorough screening process (2006).

Build a Pool of High Quality Applicants

While identifying the most effective candidates is a large part of the teacher hiring process, enticing those highly qualified candidates to take the offered positions is an equally important part. This is especially relevant when considering positions in areas where there is a higher need for teachers, including mathematics, science, and special education. Little research has been conducted on the practice of offering new teachers incentives, including salaries, signing bonuses, and insurance, and the role it plays during the hiring process, much less for high need positions. Wise et al., after reviewing the hiring practices across six districts, recommended teacher salaries and policies that promote interstate mobility of teacher certification as two potential incentives (1987). Those studies on the hiring process that mention incentives only do so descriptively. Balter and Duncombe surveyed superintendents throughout New York State about their hiring practices. They found that most districts offered either compensation for extracurricular activities or gave credit for teaching outside their districts as recruitment incentives. A smaller number of districts used other incentives, including compensation for National Board Certification, subsidized college tuition, and credit for non-teaching work experience. Furthermore, they found that 90% of the districts had some mechanism in place to increase the teacher supply (e.g., recruiting substitute teachers, using alternatively certified teachers, offering paraprofessionals support to get their teaching credentials) (2005). In a later study, Balter and Duncombe looked at the relationship between such recruitment incentives and teacher quality (i.e., a rating comprised of certification test scores, certification status, and ranking of college attended) within the New York districts. They found that district incentives such as subsidized tuition at local colleges and extra compensation for National Board Certification, participation in extracurricular activities, and experience in other districts or non-teaching occupations were significantly related to having more qualified teachers in the district (Balter & Duncombe, 2006).

Beyond this exploratory study, few studies have been conducted that look at the role incentives play in teachers' decision making. A 2005 review of literature by the Education

Commission of the States found that there was limited evidence that incentives impacted teacher recruitment. The five cited studies, conducted in the 1980s and 1990s, were mainly descriptive and focused on the role that salaries played in whether or not individuals went into teaching or another profession (e.g., engineering) (Allen, 2005). Stinebrickner used data from the National Longitudinal Study of the High School Class between 1972 and 1986 to model the potential effects of salary policies on the composition of the teacher labor supply. He found that policies that increase the salaries of all teachers or increase the salaries of teachers based on academic ability (i.e., test scores) both lead to an increase in the labor supply of teachers. However, the latter policy served to increase the supply of teachers of higher academic ability (2001). While this study focuses on academic ability as measured by test scores, it is not clear how that is related to teacher effectiveness. Nonetheless, such a study indicates that increasing compensation for teachers may serve to increase the number of people who enter into teaching, and can also be used to increase the supply of teachers with certain characteristics.

Teacher Screening Methods

Numerous studies have been conducted in the last ten years that focus on addressing the gap in the literature on teacher hiring practices. The majority of the studies describe teacher hiring practices in various school districts and the processes they use, ascertained through surveys and interviews (Balter & Duncombe, 2005; Kersten, 2008; Liu, 2003; Rutledge et al., 2008; Wise et al., 1987). The studies that described teacher hiring practices are included in Table 2.1, along with the locations, methods, and populations for each study. While these studies describe trends within districts, variation frequently exists within a district, especially as schools become more involved in the hiring process. While principals across schools may use different methods to hire teachers, there is even variation in what individual schools do or are allowed to do, depending on the vacant positions and resources available at the time (Rutledge et al., 2008; Wise et al., 1987).

Such studies that describe the range of methods for teacher hiring are a precursor to understanding which of those strategies lead to hiring the most effective teachers. While beyond the scope of this study, additional descriptive information about hiring practices at schools could be found from nationally representative samples, such as the Schools and Staffing Surveys (2012).

Table 2.1 Overview of Descriptive Studies of Teacher Hiring Practices within Districts

Study	Location(s)	Methods	Population
Balter and Duncombe (2005)	NY	Survey	Superintendents
Liu (2003); Liu and Johnson (2006)	CA, MI, MA, FL	Survey	New Teachers
Kersten (2008)	IL	Survey	Principals
Wise et al. (1987)	Mesa, AZ; Montgomery County, MD; East Williston, NY; Hillsborough, FL; Rochester, NY; Durham, NC	Case Study – Interviews, Document Review	District's Director of Personnel; Superintendent; Central Office Staff; Teacher Union Officials; Principals; Newly Hired Teachers; Senior Teachers
Mason and Schroeder (2010)	Southeastern, WI	Survey	Principals
Rutledge et al. (2008)	Hillyer County, FL	Interviews; Hiring Practice Ranking Activity	Principals; District Administrators

The districts included in the descriptive studies in Table 2.1 used a variety of tools and methods to screen prospective candidates. Table 2.2 summarizes the screening methods districts used. While these studies have been carried out over a twenty-year timespan, not much has changed from the time of the 1987 Wise et al. study to the 2009 Kersten study.

While most districts used some type of interview, they tended to use less intensive selection methods (e.g., application form, letters of recommendation/references) along with the interview. Methods that required more time and personnel on the part of the district (e.g., class observations, sample lesson, writing sample, portfolio) were used much less frequently. Rutledge et al. (2008) found that schools tend to use tools that are less time consuming, less costly, and more convenient. However, the tools schools use can be related to a district's capacity to carry out the selection process. In their review of school district hiring processes in the state of New York, Balter and Duncombe (2005) found that the larger districts with higher capacity tended to have more complex processes, largely because of the need to hire larger numbers of teachers. Thus, having the personnel, time, and monetary resources to use various screening procedures is necessary to carry out the more intensive procedures.

Table 2.2 Tools used in the teacher screening process

Study	Location	Application Form	Resume	Refs/Letters of Rec	Transcript	Cover Letter	Proof of Certification	Portfolio	Test or Test Scores	Writing Sample	Lesson Plan	Number of Interviews	Sample Lesson	Class Observation	Observe Faculty Meeting
Balter and Duncombe (2005)	NY	Almost all	Almost all	Almost all	Almost all	Almost all	Almost all	30%	Most	Most	No	2-3 avg.	50%	No	No
Liu (2003)	CA, MI, MA, FL	No	99.3%	92.3%	-72.3% under -41.2% grad	67.7%	No	40.5 %	27.8 %	24.4 %	20.0 %	1.64 mean	7.9%	35.1 %	13.2 %
Kersten (2008)	IL	-online 55.7% - paper 44.3%	82%	41.0%	77%	52.5%	46.7%	4.1%	No	No	No	1-2	7.4%	No	No
Wise et al. (1987)	Mesa, AZ	Yes	No	Yes	Yes	No	No	No	Yes	No	No	1	No	No	No
Wise et al. (1987)	Montgomery County, MD	Yes	No	Yes	Yes	No	Yes	No	Yes	No	No	1	No	No	No
Wise et al. (1987))	East Williston, NY	Yes	Yes	Yes	No	No	Yes	No	No	No	No	2	No	Yes	No
Wise et al. (1987)	Hillsborough, FL	Yes	No	Yes	No	No	No	No	Yes	No	No	1	No	No	No
Wise et al. (1987)	Rochester, NY	No	No	No	No	No	Yes	No	No	Yes	No	2	No	No	No
Wise et al. (1987)	Durham, NC	Yes	No	Yes	No	No	Yes	No	Yes	No	No	2	No	No	No
Mason & Schroeder (2010)	Southeastern, WI	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	NA	No	No	No
Rutledge et al. (2008)	Hillyer County, FL	Yes	Yes	Yes	No	No	Yes	No	yes	No	No	1-2	No	No	No

Interviews

Across studies, all districts used some sort of interview and many districts had multiple interviews. Candidate interviews in Mason and Schroeder's sample of principals were conducted in teams 91.57% of the time. The personnel included in the interviews in each of the studies are noted in Table 2.3 The majority of sites included building administrators and district personnel. While only about half of the sites included teachers in the interview process, many researchers noted the importance of including them in the hiring process to help gauge a candidate's fit with the school (Liu, 2003; Wise et al., 1987).

Table 2.3 Personnel with whom teacher candidates interviewed²

Study	Location	Building Admin	Teachers	District personnel	Dept. Chair at School	Parents	Superintendent	School Board	Students
Balter and Duncombe (2005)	NY	Majority (if 1 int.)	Majority (if 1 int.)	Final if multiple	Majority (if 1 int.)	No	No	No	No
Liu (2003)	CA, MI, MA, FL	-80.1% principal -33.2% other	45.6%	34.9%	14.7%	9.0%	9.0%	No	.1%
Kersten (2008)	IL	82%	79.5%	28.7%	No	2.5%	No	2.5%	.8%
Wise et al. (1987)	Mesa, AZ	Yes	Not often	No	No	No	No	No	No
Wise et al. (1987)	Montgomery County, MD	No	No	Yes	No	No	No	No	No
Wise et al. (1987)	East Williston, NY	Yes	No	No	No	No	Yes	No	No
Wise et al. (1987)	Hillsborough, FL	No	No	Yes	No	No	No	No	No
Wise et al. (1987)	Rochester, NY	Yes	No	Yes	No	No	No	No	No
Wise et al. (1987)	Durham, NC	Yes	Yes	Yes	No	No	No	No	No
Mason and Schroeder (2010)	WI	17% ³	55% ⁴	17%	No	17%	13%	No	No

Written Documentation

Principals in Mason and Schroeder's sample reported using tools that focus on professional attributes, such as letters of reference and resumes, first in order to narrow down the number of candidates (2010). Most districts across the other studies also required candidates to submit a

² Cells with percentages indicate the percent of district respondents within a state who said that each personnel type was involved in interviews. Cells with "yes" or "no" indicate that that district said they did or did not involve that personnel type in interviews.

³ 17% of surveyed principals said the assistant principal, parents, or central office staff were involved in interviews

⁴ 55% of surveyed principals said grade level and content area teachers are involved in interviews, 42% said additional teachers were involved, and 20% said special education teachers were involved

variety of written documentation, such as resumes, cover letters, transcripts, and references. Balter and Duncombe, in their survey of superintendents in New York, found that while most districts consider a candidate's certification, college major, and references, few considered measures of academic success (e.g., exam scores, GPA, caliber of college, quality of portfolio). They also found that the written documentation asked for depended on the characteristics of the district. For example, districts who had hiring managers were more likely to look at writing samples and ask consider the caliber of the college a candidate attended, whereas urban districts were more likely to emphasize prescreening tests (2005).

Within the realm of submitted written documentation, there was variation in how frequently other tools were used. The principals in Mason and Schroeder's study reported that references, both verbal and written, were the most important source of information on teacher candidates. On a scale of 1 to 5 of level of importance, principals gave a mean rating of 4.41 for verbal references and 3.61 for written references (2010). On the other hand, not many districts used portfolios. Mason and Schroeder's sample also rated the mean level of importance of portfolios lower, with a mean of 2.53 for portfolios and a mean of 2.44 for e-portfolios (2010).

Observations

Observations of candidates teaching, or candidates observing school practices, were not frequently used across all of the districts in the included studies. Many of the studies cited time constraints as the main reason for this (Liu, 2005; Mason & Schroeder, 2010). As Liu and Johnson point out, while an observation of teaching practice may provide rich information about a candidate, it is time consuming, requires many logistical considerations, and is therefore rarely used (Liu, 2005). However, half of the districts in Balter and Duncombe's New York sample reported using sample lessons to screen teachers, most usually observed by principals or teachers (**Balter & Duncombe, 2005**).

Extent of Centralization

Many of the studies made a distinction between districts that use a centralized or decentralized hiring process (Liu, 2003; Rutledge et al., 2008; Wise et al., 1987). On the extreme end of centralization, the school district does all of the screening, selects the candidates, and then places them in schools. On the decentralized end, individual schools do all of the screening and selection. Finally, many districts are somewhere in the middle, where the district does some amount of screening and then schools participate in additional screening and/or candidate selection. Wise et al. (1987) identified the tendency for the more centralized districts in their study to have formal procedures in place for screening candidates. They found that decentralized districts, on the other hand, can use informal procedures (e.g., selecting teachers known to the school) to ensure that there is a good fit between the teacher and the school context. Naper further explored outcomes of the centralization of teacher hiring practices in Norway. She found that school districts that have more decentralized teacher hiring practices have higher educational

efficiency, but suggests that decentralization may only be useful when there is excess teacher supply which allows schools to choose among multiple teachers (2010).

The majority of districts included in the descriptive studies had decentralized, although not completely so, hiring processes. Fifty-eight percent of the Illinois principals that Kersten surveyed reported being able to make the ultimate decision when hiring, whereas in 13% of the cases, someone at the district made that decision (2008). About three-quarters of teachers Liu and Johnson surveyed reported experiencing either highly decentralized or decentralized hiring processes (2006).

However, having a decentralized process in which schools participate in hiring decisions doesn't mean that candidates get sufficient information about the school. In fact, the study found that while the hiring process is decentralized, teacher candidates still frequently viewed it as impersonal (Liu, 2003; Liu & Johnson, 2006). Thus, a decentralized system is not sufficient in order to ensure the best fit between candidates and schools; schools must enact a hiring process that allows teachers to learn about the school and the school to select teachers based on important characteristics.

Commercial Teacher Interviews

In addition to the screening methods described above, there are multiple commercial teacher interview instruments that as many as 15% of districts in the United States use (Metzger & Wu, 2008). Common instruments include the Gallup Teacher Perceiver Interview (TPI), the Gallup Urban TPI, the Gallup TeacherInsight system, and the Urban "STAR" Teacher Selection Interview. While the Gallup Teacher Perceiver Interviews and STAR Interview are administered as face-to-face interviews, the TeacherInsight is a web-based questionnaire. The face-to-face interviews are rated by district personnel (e.g., administrators, teachers), whereas teachers rate a series of statements on a 5-point Likert scale for the TeacherInsight. All interviews measure specific constructs that are said to be drawn from research on teacher effectiveness; in the case of the Gallup Urban TPI and STAR Interview, they are constructs connected to effective urban teachers (Haberman, 1995; Metzger & Wu, 2008).

Many of the constructs measured by the commercial teacher interviews measure a variety of teacher characteristics with varying evidence-bases. Haberman identified 14 constructs related to quality teachers in urban areas through conducting a series of interviews with teachers who were identified as outstanding by their school community; seven of those constructs are measured in the Haberman STAR Urban Teacher Interview. The seven measured attributes include: persistence (i.e., ability to continuously find ways to engage students and help them learn), protecting learners and learning (i.e., emphasize and focus on love of learning over any impediments), application of generalizations (i.e., ability to apply general theory into practice and see relationships amongst day-to-day practice), approach to "at-risk" students (i.e., sees own role in ability to help at-risk students), professional versus personal orientation to students (i.e., separate affect for students from ability to teach them), burnout: it's causes and cures (i.e., work

within the system to be able to best do their job), and fallibility (i.e., recognize your mistakes). Furthermore, Haberman identified the following seven characteristics that are not measured: organizational ability (i.e., planning and gathering materials), physical/emotional stamina (i.e., persistence in tough situations), teaching style (i.e., coaching over directive teaching), explanations of success (i.e., effort over ability), basis of rapport (i.e., approach to student involvement), and readiness (i.e., approach to prerequisite knowledge) (Haberman, 1995).

The Gallup Teacher Perceiver (TPI) interview measures twelve themes: mission (i.e., goal is to contribute to student growth), empathy (i.e., responds to students' thoughts and feelings), rapport drive (i.e., promotes warm relationships with students), individualized perception (i.e., considers interests and needs of each student), listening (i.e., listens to student feelings), investment (i.e., satisfaction comes from learner response), input drive (i.e., searches for new ideas and experiences for students), activation (i.e., motivates students to think, respond and feel in order to learn), innovation (i.e., implements creative new ideas), gestalt (i.e., tends toward perfectionism but works from individual to structure), objectivity (i.e., responds to the total situation), and focus (i.e., uses models and goals to select activities) (Metzger & Wu, 2008). Metzger and Wu conducted a meta-analysis of 24 studies that looked at the relationship between the Gallup TPI with various measures of teacher effectiveness. The 24 studies found a moderate relationship with teacher quality (i.e., average effect size of .28), with the highest correlations between the TPI and administrator ratings of teachers and teacher absenteeism. Furthermore, two studies that conducted regression analyses of TPI scores on student outcomes found no significant relationship between the two. While the study authors only looked at the relationship between overall TPI scores and teacher effectiveness, they suggest that there is value in looking at the relationship between individual themes and teacher outcomes (2008).

Teacher Characteristics

In order to hire the most effective teachers, schools and districts must pay attention to the characteristics they desire in a teacher and ensure that their measures are aligned with those qualities (Little, Goe, & Bell, 2009; Wise et al., 1987). While much of the research on the teacher hiring process has focused on describing the process and the tools used, fewer studies have focused on the specific characteristics that are being measured during that process. Table 2.4 includes those studies that identified teacher characteristics and notes the characteristics that each district measured. Table 2.5 includes any negative characteristics that districts specifically noted they would use to rule out teachers. While many of these were not characteristics districts would explicitly measure, when candidates exhibited them, they were less likely to be hire. Many of the characteristics were not well-defined or operationalized, or definitions varied across studies, therefore there is overlap between the different characteristics and how they are categorized. For example, while one study listed interpersonal skills, cooperation, and communication as separate constructs, others considered cooperation and communication to be a

subset of interpersonal skills. Furthermore, while the characteristics are presented in categories for ease of discussion, districts generally screened for multiple characteristics across each of the categories.

While the table provides an overview of the types of characteristics districts looked for, the studies do not necessarily include a comprehensive list of characteristics. In some cases, individual principals who make hiring decisions may use different criteria and look for characteristics not described in the studies. Additionally, while some studies included open-ended questions to gather information about which characteristics schools and districts looked for (Kersten, 2008), others described the characteristics in the context of the tools used throughout the process (Wise et al., 1987), and others included closed-ended questions with a finite number of characteristics (Mason & Schroeder, 2010). Each of these methods will not likely yield a complete list of characteristics. Furthermore, as Wise, et al. noted was the case for their study, there was great variation in the weight given to each characteristic; the inclusion of each characteristic does not indicate its level of importance.

Table 2.4 Characteristics districts in descriptive studies used to select candidates

Study	Location	Teaching Practice	Academic Qualifications/ Knowledge	General Interpersonal Skills	Communication Skills	Cooperation	Professionalism/ Work Ethic	Appearance	Willingness to Learn	Other Personal Attributes	Sensitivity to Needs of Parents/Students	Passion/ Enthusiasm for Job/Working with Children	Work in Extracurriculars	Fit with Culture
Kersten (2008)	IL		22.3 %	10.5 %	7%	7.2%	13.7 %		4.3%	Student-focused (9.4%)		4.6%		
Wise et al. (1987)	Mesa, AZ		Yes	Yes						*Persistence *Creativity *Nurturing	Yes	Yes	Yes	Yes
Wise et al. (1987)	Montgomery County, MD		Yes			Yes				*Persistence *Excitement				
Wise et al. (1987))	East Williston, NY		Yes			Yes	Yes				Yes		Yes	Yes
Wise et al. (1987)	Hillsborough, FL	Yes	Yes	Yes			Yes	Yes		*Role model *Motivation				
Wise et al. (1987)	Rochester, NY	Yes			Yes	Yes		Yes		*Poise *Flexibility		Yes		Yes
Wise et al. (1987)	Durham, NC		Yes			Yes				Manage diversity		Yes	Yes	
Mason and Schroeder (2010)	Southeastern, WI		15%		15%	10%	13%	20%	10%	Confidence (20%) Excitement (25%)		17%		

Table 2.5 Negative characteristics districts in descriptive studies used to rule out candidates

Study	Location	<i>Poor Appearance</i>	<i>Poor communication</i>	<i>Poor preparation</i>	<i>Lack of Professionalism/ Arrogance</i>	<i>Punctuality</i>	<i>Lack of Confidence</i>	<i>Lack of Depth of Response to Questions</i>	<i>Lack of Content Knowledge</i>
Kersten (2008)	IL		10.7%		3.4%				15.1%
Mason and Schroeder (2010)	Southeastern, WI	32%	15%	13%	12%	10%	10%	17%	10%

Teaching Practice

While teaching practice is difficult or costly to measure for new teachers, it is not surprising that only one study mentioned districts or schools that looked for teaching experience or teaching ability when hiring teachers. Two of the five case study districts in that study screened teachers for teaching practice (what they referred to as “teaching ability”), one of which specifically looked at classroom management and the ability to meet the needs of individual students (Wise et al., 1987).

Content Knowledge and Academic Qualifications

School districts in the descriptive studies did rely on more traditionally used measures to hire teachers, including academic qualifications, content knowledge, and teaching ability. Many school districts and schools look for teachers with sufficient academic qualifications and content or pedagogical knowledge (Kersten, 2008; Mason & Schroeder, 2010; Wise et al., 1987). Others also noted lack of depth of responses to questions and lack of content knowledge as reasons districts did not hire teachers (Kersten, 2008; Mason & Schroeder, 2010).

Interpersonal Skills

Many of the districts identified interpersonal skills as important, both generally and in the form of specific skills. Kersten (2008) noted the sample of principals in Illinois identified friendliness, caring, tact, empathy, and sense of humor, as well as a candidate that is outgoing, people-centered, and personable. Other districts mentioned cooperation and teamwork (Kersten, 2008; Mason & Schroeder, 2010; Wise et al., 1987) and communication skills (Kersten, 2008; Mason & Schroeder, 2010; Wise et al., 1987) as important teacher characteristics. On the other side, districts saw poor communication skills (Kersten, 2008; Mason & Schroeder, 2010) as a red flag that they screened teachers out on the basis of.

General Personal Attributes

Districts also identified a number of general personal attributes that they look for in prospective teachers, including work ethic (Kersten, 2008), personal appearance (Mason & Schroeder, 2010), confidence (Mason & Schroeder, 2010), willingness to learn (Mason & Schroeder, 2010), professionalism (Mason & Schroeder, 2010), persistence (Wise et al., 1987), flexibility (Wise et al., 1987), creativity (Wise et al., 1987), nurturing (Wise et al., 1987), excitement (Wise et al., 1987), role model (Wise et al., 1987), motivation (Wise et al., 1987), poise (Wise et al., 1987), and ability to manage diversity (Wise et al., 1987). Additionally, districts identified poor appearance, poor preparation, punctuality, lack of confidence (Mason & Schroeder, 2010), and lack of professionalism (Kersten, 2008; Mason & Schroeder, 2010) as red flags. Teach for America, one alternative certification program, uses mathematical models to refine their selection process and identify teacher characteristics that are related to their most effective teachers (i.e., high student achievement). Information about these characteristics are measured through an extensive selection process that includes an application and essays, online activities, a phone interview, and a day-long in-person interview. Many of the characteristics they look for, that presumably have some relationship to teacher effectiveness, are personal attributes such as personal responsibility, critical thinking, organizational ability, motivational ability, respect for others, and commitment to the Teach for America's mission (Glazerman, 2006).

Attitudes, Beliefs, and Values

The districts in many of the descriptive studies screened to some extent for the values, beliefs, and attitudes of prospective teachers (Kersten, 2008; Mason & Schroeder, 2010; Wise et al., 1987). This encompasses a wide range of values, beliefs, and attitudes related to teaching and learning, as well as behaviors that may be indicative of those beliefs, including a passion for and commitment to children (Kersten, 2008; Mason & Schroeder, 2010; Wise et al., 1987), being

student-focused (Kersten, 2008), and being sensitive to the needs of parents and students (Wise et al., 1987).

Match based on Content, Grade, Pedagogy, Connection to Community

While there are certain characteristics that most districts would want to screen for, some characteristics will depend on the context in which the teacher is hired under. The fit between a candidate and school is another important factor in the teacher hiring process. Borrowing from the literature on hiring in areas outside of teaching, Rutledge et al. identify three ways to think about the organization of work and hiring: 1) the person-job fit between a candidate's knowledge, skills and abilities and job requirements; 2) the person-organization fit between a candidate and the organization's culture and values; and 3) the person-group fit between a candidate and the workers she will work with. Such a wide range of characteristics leaves the process open to a wide range of local preferences, and consequently a wide range of tools that can be used and characteristics that can be measured to select on those preferences (2008).

While there is no empirical evidence to support the claim, Liu also identified the fit between a candidate and school as important to consider during the hiring process. He, however, looks at this issue in terms of how well the candidate is informed about the school and the extent to which selection processes actually provide information to the candidates. In the states studied, teachers felt they had moderately accurate pictures of their schools and had, on average, a "good" fit with their position and a "moderate" to "good" fit with their school. Liu concluded by speculating that designing hiring activities so that teachers get an accurate picture of their school and can gauge the extent to which they fit in their schools and positions may influence teachers' satisfaction with their jobs and subsequent retention (Liu, 2003; Liu & Johnson, 2006).

Teacher Evaluation

While the methods above focus on hiring new teachers, it is worth briefly reviewing additional methods for evaluating teacher effectiveness that can be used to evaluate new teachers or when hiring experienced teachers. Since the focus of this study is on the teacher hiring process and the teacher effectiveness measures will be used mainly to validate those hiring practices, this section will only provide a brief summary of those methods that are used in HCS.

These methods include using standardized test scores (e.g., Student Growth Percentiles), classroom observation scores, student perception surveys, principal evaluations, teacher attendance, and student discipline referrals.

Researchers recognize that there are multiple ways to evaluate teachers, each method having its own tradeoffs (Baker et al., 2010; Little et al., 2009). In the absence of a single perfect measure of teacher effectiveness, much of the literature on the subject focuses on the trade-offs

between using various measures. The Gates Foundation Measures of Effective Teaching (MET) project looked at what combinations of classroom observation scores, student perception survey scores, and student test score gains could reveal about teacher effectiveness. They found that including all three measures yields more consistent ratings over time than test scores alone. Furthermore, using teachers' scores with all three measures from the previous school year, researchers were able to accurately predict student performance for the following school year. They also found that various combinations of those measures produced trade-offs between teacher evaluation scores accurately predicting future student achievement, and those scores varying dramatically from year-to-year. While teachers on the high or low end of the spectrum are less likely to vary largely each year, those in the middle of the spectrum tend to be more volatile. While there was not one clear best weighting scheme, ultimately, the study indicated that the optimal weighting scheme of the various evaluation measures will vary and depend on the needs of individual districts (MET Project, 2013).

Little, Goe, and Bell also recommend that any teacher evaluation system considers the context in which teachers work in, informs and improves instruction, and uses measures in a way consistent with how they were validated. Additionally, they recommend that multiple measures are incorporated that include the perspectives of multiple stakeholders (Little et al., 2009). Thus, we would expect that while school districts might use a variety of teacher effectiveness measures, the specific measures and weight they give each measure would likely vary.

Standardized Test Scores (Student Growth Percentiles)

Recent literature on teacher effectiveness has focused on a variety of statistical models that look at the growth that individual or groups of students make over time on standardized tests. Growth models, as opposed to models that focus solely on achievement levels, take into account that students may have started at different achievement levels by focusing on the amount of progress students make over time. While many of these models don't measure an individual student's growth from one time period to the next, they measure a student's current achievement relative to the current achievement of students with similar past scores. Castellano and Ho reviewed various growth models that serve the purpose of describing growth, predicting growth, or identifying the extent to which teachers and schools contribute to growth. Each method they reviewed has different underlying statistical foundations and different data requirements, and can also lead to different interpretations about groups of students, different standards for what constitutes adequate growth, and different misinterpretations based on the data (Castellano & Ho, 2013). The purpose for the model, along with the relative merits of each model, must be considered when deciding which to use.

Two frequently used models are student growth percentiles (SGP) and value-added models (VAM). SGP measures compare a student's current test score with the current score of students with similar score histories. It then assigns a percentile to where that student is relative to the

other students (e.g., a score of 75 means a student scored as well as or better than 75% of students with similar past scores). The teacher is then typically assigned the median of the SGPs for all of that teacher's students. VAM measures estimate the contribution of a teacher on student performance, controlling for differences in prior student achievement and, in some cases, student demographics.

The Center for Education and Data Research (CEDR) compared the results from different model specifications of teacher effectiveness using VAM and median SGP measures with teachers in North Carolina. They found that with the overall sample, there was a high level of agreement (i.e., .90 for math and .80 for reading) between the ratings of teacher effectiveness produced by each method. However, they found that the ratings of teacher effectiveness vary across models when analyzing teacher effectiveness within subsets of classroom characteristics (e.g., prior year average test score, percent free and reduced price lunch, percent black students) (D. Goldhaber, Gabele, & Walch, 2012). Thus, it is important to recognize that using specific models will lead to differential ratings of teacher effectiveness for different groups of teachers.

Beyond issues with VAM and SGP measures, there are multiple issues with the use of standardized test scores to measure teacher effectiveness. The tests create incentives for teachers to focus on test preparation that will increase test scores without real learning, avoid working with the neediest students, narrow the curriculum to focus on tested content, and not work with other teachers (despite the fact that evidence suggests teacher collaboration can positively impact student achievement). Furthermore, the tests aren't vertically scaled (i.e., don't measure the same content across years) and usually only measure basic content through multiple-choice items. Recognizing all of these issues with student test scores, Baker et al. recommend that test scores are only used as a small component of a teacher evaluation system and that districts have the autonomy to test out a variety of methods.

Classroom Observation Scores

Classroom observations are another widely used method of evaluating teacher practice. The Gates' Foundation MET study looked at the validity and reliability of five widely used classroom observation protocols - Framework for Teaching (FFT), Classroom Assessment Scoring System (CLASS), Protocol for Language Arts Teaching Observation (PLATO), Mathematical Quality of Instruction (MQI), and UTeach Teacher Observation Protocol (UTOP). They found that all five protocols were positively related to student achievement gains, but required scores from multiple raters over multiple lessons to achieve reliable ratings (Kane & Staiger, 2012). The National Comprehensive Center for Teacher Quality reviewed the research on FFT and CLASS and found that both protocols had positive relationships with student outcomes. While FFT had moderate correlations to student outcomes that varied across studies, the Kindergarten through 5th grade version of CLASS was generally correlated with student achievement gains, as well as other student outcomes (Little et al., 2009). Ing and Webb (2012) looked at the use of an observation protocol to measure teacher practice in mathematics and found that the focus of the observation

(individual students, whole class or small group), tracking individual students, observing on one versus multiple days, and tracking more specific elements of student discussion lead to different conclusions about classroom practice (Ing & Webb, 2012). Such changes in the way an observation protocol is used can thus lead to very different inferences about a teacher's effectiveness.

Other researchers have focused on identifying what good teaching is using observations within specific domains. Each of these studies points to specific elements of teacher effectiveness, as well as tools that can potentially measure those elements. The Learning Mathematics for Teachers Project found that many of the numerous observation instruments in education are mainly focused on generic pedagogical techniques. They therefore created an instrument, the Mathematical Quality of Instruction (MQI), that was also included in the MET study, that measures the nature of mathematical content in an observed lesson, postulating that such a construct may relate to effective teaching and subsequent learning. They found that the observed constructs measured in the instrument correlated to teachers' mathematical knowledge, as measured by a test of mathematical knowledge ("Measuring the mathematical quality of instruction," 2011). While further research is needed to explore the relationship between such a construct and student outcomes, such findings indicate that there may be elements of teacher practice that are specific to math, measurable, and may impact student outcomes.

Principal Ratings

While most classroom observations used for teacher evaluations are conducted by principals, there is also additional research on principal evaluations that are more holistic and are not solely tied to observations. Some of these studies indicate that principals can identify effective teachers. Jacob and Lefgren had 13 principals in a school district rate 201 of their teachers in terms of a variety of characteristics related to teacher effectiveness. By comparing those ratings with value-added scores, they found that principals could generally identify the most and least effective teachers in their schools, but had a more difficult time identifying teachers in the middle of the distribution. Furthermore, they found that while VAM measures better predicted future student achievement gains, principal evaluations and VAM measures were equally effective at identifying the most and least effective teachers (Jacob & Lefgren, 2008). Such findings support the idea there is value to using multiple methods to measure teacher effectiveness and that there is value to include principal evaluations as one of those components. While the Jacob and Lefgren study involved evaluating teachers based on holistic ratings in multiple categories, other principal evaluations are more structured and based on ratings from classroom observation protocols (as discussed above).

Student Surveys

Many researchers have recognized the value of using student surveys in rating teacher practice. While researchers have found that student surveys are at least moderately correlated with other measures of teacher effectiveness, they also recognize that are sources of possible bias that the survey responses are subject to (Burniske & Meibaum, 2011; Ferguson, 2012; Greenwald, 1997). The Southwest Education Development Laboratory conducted a review of research on student perception surveys. The studies they reviewed largely found that student survey ratings are correlated with student achievement and that students can differentiate between effective and ineffective teaching (Burniske & Meibaum, 2011). Similarly, the Gates Foundation's MET project found that the Tripod survey, a widely used student perception survey, had a positive relationship with student achievement gains (MET Project, 2010). The Tripod survey measures "the 7 Cs" (i.e., care, control, clarify, challenge, captivate, confer, consolidate), as well as items related to student engagement (Ferguson, 2012; MET Project, 2010). Ultimately, student surveys introduce another point of view for measuring teacher effectiveness, but should be used recognizing their limitations and potential shortcomings.

Attendance

Prior research has found that students whose teachers are absent less perform better (Clotfelter, Ladd, & Vigdor, 2007; Miller, Murnane, & Willett, 2008; Woods & Montagno, 1997). One study in an urban school district found that ten days of teacher absence reduced the achievement of fourth graders by 3.2% of a standard deviation on a standardized math test, and to a lesser extent in reading. Additionally, they found that unexpected absences had a more negative effect on student achievement than anticipated absences (Miller et al., 2008). Another study found that students in Iowa and Wyoming who had teachers with fewer absences performed better on the Iowa Test of Basic Skills (Woods & Montagno, 1997).

Discipline Referrals

There is little to no research on the relationship between discipline problems or referrals and teacher effectiveness. Many studies have looked at the number of office discipline as an outcome measure in reviewing the effectiveness of behavioral interventions (Irvin, Tobin, Sprague, Sugai, & Vincent, 2004; Safran & Oswald, 2003) and in looking at the relationship between discipline referrals and race or other student characteristics (Bryan, Day, Vines, Griffin, & Moore, Thomas, 2012; Skiba et al., 2014; Vincent, Tobin, Hawken, & Frank, 2012). One study found that using the number of office discipline referrals can be an effective means of identify grades, teachers, or classes that need targeted behavioral support intervention (Putnam, Luiselli, Handler, & Jefferson, 2003).

Other studies have looked at the relationship between classroom management and student outcomes. One study found that observed elements of the classroom learning environment,

including classroom management, were positively and significantly related to increased student test scores in math (Kane, Taylor, Tyler, & Wooten, 2011)

3. Methods

Data sources

This research involves a combination of qualitative (e.g., literature reviews, document reviews, interviews, observations) and quantitative (e.g., interview data, probationary teacher review data) data. The results will help inform Huntsville City Schools (HCS), as well as other schools and districts, in selecting the characteristics to measure during the hiring process.

Qualitative Methods

What are Huntsville City Schools' (HCS) current hiring practices?

The background and information about Huntsville City School's (HCS) hiring and probationary teacher review processes came from interviews and observations conducted during two site visits to HCS, in April 2014 and March 2015. The site visits included interviews, collection of relevant documentation, a focus group, and observations.

- Prior to, and during the April 2014 site visit, documentation was gathered that described the teacher hiring and probationary teacher review processes, rubrics used during the processes, and timelines of the processes.
- In April 2014 and March 2015, interviews were conducted with HCS staff in the Talent Management and Assessment and Accountability Departments and focused on the logistics of the teacher hiring and probationary teacher review processes.
- In March 2014, a focus group was conducted with six principals (three high school, two elementary school, one K-8) about their experiences with hiring in HCS. All of the principals had been members of the district hiring committees and three had also been members of the probationary review committees.
- Finally, in March 2014, observations were conducted of staffing meetings between principals and the HCS Talent Management Department to determine staffing vacancies, as well as an observation of five teacher interviews and subsequent deliberations of the interview committee to rate the candidates.

First, timelines and administrative memos created by the HCS Talent Management Department were reviewed and used to create a process map and description of the HCS hiring process. The interview and probationary teacher review rubrics were reviewed to contextualize the meanings of the interview and probationary review scores. An interview protocol for the

April 2014 site visit (see Appendix A) was created to elicit a deeper understanding of the hiring process and answer questions that arose from the documentation review. The interviews with HCS personnel and the principal focus group were recorded and later reviewed to better understand the hiring and probationary review processes, and to elicit any challenges, enablers or other factors that interviewees noted that provided broader context. Notes were taken during the observations of staff meetings and candidate interviews that focused on describing the process and noting any challenges or other contextual factors that arose during the process.

The interviews, focus groups, and observations were also used to confirm the process map and to better inform my understanding of the hiring process. I followed up with HCS personnel to resolve any discrepancies arose between the document review, interviews, and observations. The process map and summary of the hiring and probationary review processes that resulted from these data sources are included in Chapter 4.

Quantitative Methods

The data analyzed in this study include 583 teachers hired using the teacher selection process across two cohorts, during the 2012-2013 and 2013-2014 school years, that underwent the probationary teacher review (337 in Cohort 1, hired in 2012-2013 and 246 in Cohort 2, hired in 2013-2014). The data from the probationary teacher review includes data from the 2012-2013 and 2013-2014 school years. Data on the majority of teachers who applied for a teaching position in HCS and were either not interviewed or interviewed and not hired was not available and therefore those teachers were not included in the analysis.

How well does HCS's existing interview process for rating teacher candidates predict the effectiveness of new hires?

The quantitative portion of this research focuses on analyzing the relationship between the teacher hiring measures and indicators of teacher effectiveness. More specifically, I looked at the relationship between interview scores, probationary review scores, and whether or not Cohort 2 teachers were still teaching in the district at the end of their second year. This study extends the work of others by including measures derived from the actual hiring process, as opposed to measures from newly hired teachers that were collected after they began teaching.

Variables

Dependent Variable

As detailed in Chapter 4, the individual elements of the probationary reviews were determined using a rubric that identified specific thresholds for raw scores that correspond to points on the probationary review (e.g., 4-6 absences corresponds to a teacher receiving 6 out of 9 points on the probationary teacher review). While raw scores on each element of the probationary review (e.g., an SGP of 60) were not available for all teachers, the converted point

values (e.g., 15 out of 15 points on the SGP portion of the probationary review) were. Therefore, the converted point values were used in the analyses.

Moreover, since the thresholds and point values of the rubrics changed between the 2012-2013 and 2013-2014 school years, the point values were converted into a 4-point ordinal scale: 0 points, low, medium, and high. Furthermore, since the point values and relative weightings of each review component changed across years, this made the overall interview scores not directly comparable across years. In order to make the overall review score comparable across years, they were transformed for the analysis so they have a mean of 0 and a standard deviation of 1.

Independent Variable

As detailed in Chapter 4, interview scores consisted of ten questions that were worth four points each, for a total of 40 points. Since not all candidates were asked all questions, some teachers had total possible points of less than 40. Therefore, the interview scores are included in the analysis as the percentage of points received out of possible points.

Teacher Retention

Retention in Year 2 was determined by whether or not a teacher had a Year 2 probationary review score. Teachers may have not had a probationary review score for a variety of reasons (e.g., let go because of low review score in Year 1, dismissed as a disciplinary measure, voluntarily left) and while the data indicates which teachers were dismissed because of low review scores, it does not indicate whether the remaining teachers were dismissed or voluntarily left the district. Therefore, this portion of the analysis focuses on whether or not teachers remained in the district, as opposed to the reasons they left.

Models

Since the overall probationary teacher review score is a main source of decision making for HCS personnel, the first two analyses are ordinary least squares (OLS) regressions looking at how the total teacher hiring interview score predicts the overall probationary review score in year 1, and then in year 2. The schools where teachers are placed would likely affect their review score. For example, teachers at a school with more student behavior issues might have more disciplinary referrals and therefore might get a lower score on their probationary review. While there were not enough teachers hired at more schools across the district to control for the schools teachers were placed at, I instead controlled for whether or not teachers were placed in a Title I school. In Equation (1), Y_{it} is the total probationary teacher review score in year t ($t=1,2$) for teacher i . H_i is the total teacher interview score for teacher i , and T_j is an indicator variable for whether or not teacher i was placed in a Title I school.

$$(1) Y_{it} = \alpha + \beta H_i + \mu T_j + \varepsilon_i$$

Since there is a modest correlation (.3615) between Year 1 and Year 2 probationary review scores which would not eliminate all of the variation in Year 2 scores, I also control for Year 1 probationary review scores when looking at the relationship between Year 2 review scores and interview scores. Next, since not all teachers receive student growth scores and student surveys, and not all teachers receive observations in year 2, I ran a similar analysis, but incorporated dummy variables that interact with the overall interview score. In Equation 2, Y_{it} is the total probationary teacher review score in year t ($t=1,2$) for teacher i . H_i is the total teacher interview score for teacher i , S_{it} indicates whether or not teacher i had a student growth score in year t , Q_{it} indicates whether or not teacher i had a student survey score in year t , O_{i2} indicates whether or not the teacher had an observation in year 2, and T_i indicates whether or not teacher i was placed in a Title I school.

$$(2) Y_{it} = \alpha + \beta H_i + \gamma H_i S_{it} + \theta H_i Q_{it} + \delta H_i O_{i2} + \mu T_j + \varepsilon_i$$

Each analysis above was repeated using a logistic regression where the outcome of interest is whether or not the teacher continued teaching in the school district through the end of Year 2. This set of analyses will only include Cohort 1 teachers since that data is not yet available for Cohort 2 teachers. In Equation 3, Y_i is whether or not teacher i continued teaching in the school district through the end of Year 2, H_i is the interview score for teacher i , T_i indicates whether or not a teacher was placed in a Title I school. Here, β is interpreted as the odds ratio that a teacher is retained given her interview score and holding Title I school status constant.

$$(3) E(Y_i|H_i) = (\exp(\alpha + \beta H_i + \mu T_j)) / (1 + \exp(\alpha + \beta H_i + \mu T_j))$$

In Equation 4, I add a control for teacher i 's Year 1 probationary review score, Y_{i-1} .

$$(4) E(Y_i|H_i) = (\exp(\alpha + \beta H_i + \mu T_j + \Phi Y_{i-1})) / (1 + \exp(\alpha + \beta H_i + \mu T_j + \Phi Y_{i-1}))$$

In Equation 5, I control for missing interview components, where S_{it} indicates whether or not teacher i had a student growth score in year 1, Q_{it} indicates whether or not teacher i had a student survey score in year 1, and O_{i2} indicates whether or not the teacher had an observation in year 1.

$$(5) E(Y_i|H_i) = (\exp(\alpha + \beta H_i + \gamma H_i S_{i1} + \theta H_i Q_{i1} + \delta H_i O_{i1} + \mu T_j)) / (1 + \exp(\alpha + \beta H_i + \gamma H_i S_{i1} + \theta H_i Q_{i1} + \delta H_i O_{i1} + \mu T_j))$$

Finally, each of the analyses was then repeated using an ordered logistic regression (OLR) looking at how the overall teacher interview score predicts the individual components of the probationary review. In Equation 4, Y_i is the score on the probationary review element for teacher i with g levels (i.e., 0, low, middle, high), H_i is the interview score for teacher i , and T_i indicates whether or not a teacher was placed in a Title I school. Here, we estimate the

probability that teacher i receives a score on one element of the probationary review greater than g (e.g., 0 versus low, medium, and high) given teacher i 's interview score. β is the estimate of the odds ratio that a teacher will receive a higher level score on the probationary review element given her interview score and holding Title I status constant.

$$(6) p(Y_i \geq g | H_i) = 1 / (1 + \exp[-(\alpha_g + \beta H_i + \mu T_i)])$$

where $g = 0, 1, 2, 3$

In order to account for the fact that multiple hypotheses were being tested and it's possible some hypotheses could be found by chance, I adjusted the significance level using the conservative Bonferroni method (Bland & Altman, 1995). To account for the 19 hypotheses tested, the .05 significance level was adjusted to .0026 by dividing the significance level of .05 by 19. All significant findings in Chapter 5 are therefore reported at the .0026 level of significance.

Factor Analysis

Following the exploration of the overall interview scores, I focused on the individual components (i.e., interview questions) of the interviews following prior research that suggests composite indicators may be related to teacher effectiveness. This analysis provides insights into whether there are particular interview questions that form underlying constructs and whether those characteristics are related to the overall probationary review or its individual components. Furthermore, the factor analysis might indicate whether there are particular questions that measure similar constructs and could be excluded from future iterations of the interview.

Special education and physical education teachers, as well as counselors, were excluded from the factor analysis since they were not asked the same set of nine questions on their interviews. While the other teachers were asked ten questions, only nine of the questions were common across all teachers and one was specific to grade or subject area. Only the nine common questions were included in the factor analysis.

First, I looked at the correlation between the nine interview questions. Since the interview scores on all of the items were not normally distributed (see Chapter 4), I conducted principal axis factoring to determine the number of factors and selected the number of factors that had eigenvalues greater than 1. Since only one factor emerged, I did not do a rotation. I also looked at the alpha coefficient to explore the internal consistency of the factors and determine if removing any individual variables would increase the alpha coefficient.

After deciding to retain all nine items, I added up the scores for the nine interview items to create the new composite interview score. I then examined the relationship between the new composite and the outcomes of interest (i.e., overall probationary review score, probationary review component scores, Year 2 retention), similar to that specified in Equation 1, but with the overall interview score replaced by the score of the composite created after the factor analysis.

Matrices of costs and advantages of instruments

What are the costs and advantages of alternative instruments as compared to the current process?

The final part of this research identifies the costs and advantages of alternative sources of data that could be used during the hiring process. To the extent that prior research addresses this area, findings were broadly drawn from the literature summarized in Chapter 2. The literature included in Chapter 2, and drawn upon in this analysis, began with searches using Google Scholar, EBSCO, and Education Resources Information Center (ERIC). The review focused on two areas: teacher hiring and teacher effectiveness. Different combinations of the following terms were used: teacher, education, hiring, selection, screening, effectiveness, and practices. The search did not limit sources based on methodologies, sample size, or date. Thus, the literature includes smaller descriptive case studies, as well as larger empirical studies. Abstracts yielded from the search were reviewed and relevant articles were identified. As sources were reviewed and relevant citations were identified, those sources were incorporated into the review process.

Using the literature to the extent possible, I identified the objectivity, simplicity and costs of administering, and proximity to classroom practice of each data source. These results helped populate a matrix that identifies potential costs and advantages of each data source. While there is not systematic evidence on the costs and advantages of each instrument and collecting such data is out of the scope of this research, I looked across the available data sources to broadly classify whether the costs and advantages of each instrument are high, moderate, or low. Such a matrix will help policy makers consider the relative costs and advantages of various hiring instruments and assist in making decisions about which instruments to use.

Additionally, I noted the teacher characteristics that could be measured using each data source, as well as the extent to which those characteristics are related to teacher effectiveness. A second matrix was created that identifies the relationships between the data sources and teacher characteristics. Such a matrix will help a district look at the characteristics that they desire in a teacher and determine how best to measure them.

4. RQ1: Background on Huntsville City Schools

Introduction

This chapter provides background on Huntsville City Schools (HCS) and the context in which the teacher hiring and evaluation processes occur. After providing background, I then describe the teacher hiring and evaluation processes that HCS uses, along with descriptive statistics from the data used in the quantitative analysis in Chapter 5 including interview, evaluation, and retention data. The description of the teacher hiring process in this chapter answers the first research question: What are Huntsville City Schools' (HCS) current hiring practices?

Huntsville City Schools (HCS) Context

Huntsville City Schools (HCS) in Huntsville, Alabama, has approximately 1,700 teachers and 23,374 students across 39 schools. The school district is undertaking many innovative programs, including a digital one-to-one initiative in which all students in the district have a tablet (grades k-2) or laptop (grades 3-12) and use digital curricula. The district is also focused on using data to inform practice at all levels. The directors of the various departments at the district-level (e.g., Talent Management, Assessment and Accountability, Transportation, IT) hold weekly meetings in which they present work their department has done, along with any relevant data, and receive input and guidance from others about future actions. The status of the teacher-hiring process, among other topics, is discussed during these meetings. Similarly, principals meet with the superintendent and district staff to discuss the performance data of their teaching staff on a regular basis.

HCS has large disparities in achievement across its schools and has been under a desegregation order by the U.S. Department of Justice since 1965. One focus group principal noted that because of the Department of Justice order, the district focuses on diversity: "The [hiring and probationary review] committees are diverse and the people we are interviewing are diverse. We consider diversity throughout." Initial conversations with the HCS Talent Management Department indicated that the district does not explicitly take the desegregation order into consideration when hiring teachers and feels that improving their hiring process, selecting the most effective teachers, and placing them into schools as positions arose would address the issue by redistributing effective teachers across all schools, regardless of the racial composition of the teachers or schools. However, the final question of the HCS candidate process asks teachers, "Would you be willing to work in any school you were placed in in HCS?" indicating that the district felt that a teacher's willingness to work in any school in the district is important. While there are not any formal policies in place dictating that teachers who

are not willing to work in any school would not be hired, conversations with district personnel indicated that teachers who did not answer in the affirmative were much less likely to be hired. Conversations with the Talent Management Department also revealed that the final interview question was asked because the district wants teachers who are willing to work for the good of the school district, in any school, and not just in a certain school or subset of schools (e.g., high achieving schools).

Table 4.1, below, summarizes the largest sources of HCS teacher hires for 2012-2013 and 2013-2014, and other data on those hires. There are multiple local and in-state universities with schools of education that provide large numbers of teachers for HCS. In addition, the district hired between six and eleven percent of new teachers each year from Teach For America (TFA) a national, alternative teacher certification program that recruits academically talented teachers for two-year teaching commitments. Along with TFA, the universities that provided the largest number of new hires for HCS between 2012 and 2014 were Athens State; Alabama A & M; Auburn University; University of Alabama; University of Alabama-Huntsville; University of Northern Alabama; and Jacksonville State. Table 4.1 also summarizes the number of teachers hired from each program and the average candidate interview scores, review scores, and percent still in the district at the end of their second year. Both Alabama A & M and Jacksonville State are Historically Black Colleges and Universities (HBCUs). While Athens State is located approximately 30-miles from Huntsville, both the University of Alabama Huntsville and Alabama A & M are located in Huntsville.

Table 4.1 Descriptive statistics by major teacher suppliers

School/ Organization	# Hired 2012-2013 (% of total hires)	# Hired 2013-2014 (% of total hires)	Total		Avg Candidate Interview Score	Avg Year 1 Review Score	% Retained End of Year 2	Avg Year 2 Review Score
Athens State	32 (13%)	34 (15%)	66 (14%)		81.7 (10.6)	77.9 (13.9)	72%	80.1 (14.1)
Alabama A & M	24 (9.8%)	21 (9.1%)	45 (9.4%)		75.2 (14.4)	79.1 (15.7)	75%	77.8 (19.1)
Teach For America	16 (6.5%)	25 (11%)	41 (8.6%)		80.75 (14.2)	75.3 (14.3)	100%	87.4 (7.1)
Auburn University	19 (7.7%)	18 (7.8%)	37 (7.7%)		86.6 (10.5)	78.4 (15.2)	68%	85.6 (12.6)
University of Alabama	18 (7.3%)	15 (6.5%)	33 (6.9%)		82.7 (12.2)	76.5 (20.9)	78%	85.5 (9.1)
University of Alabama – Huntsville	22 (8.9%)	10 (4.3%)	32 (6.7%)		83.7 (13.6)	77.2 (14.4)	77%	83.3 (12.0)
University of Northern Alabama	11 (4.5%)	13 (5.6%)	24 (5.0%)		83.6 (12.1)	79.4 (15.9)	91%	77.5 (6.9)
Jacksonville State	7 (2.8%)	14 (6.1%)	21 (4.4%)		86.5 (9.7)	71.2 (16.8)	71%	85.7 (10.0)

Note: Author's analysis of data provided by Huntsville City Schools

Note: This table only includes the 8 largest teacher suppliers for HCS, which together supplied approximately 60% of teachers for HCS

Huntsville City Schools' Hiring Process

Prior Hiring Process

The new hiring process described below was implemented starting with the 2012-2013 school years. Prior 2012-2013, the process was largely decentralized, candidates would apply to positions at specific schools, and the principal at each school was responsible for structuring the process. Principals who participated in the April 2014 focus group described the previous process as very time consuming, where they were given lists of candidates that they had to sort through. Principals said that there could be hundreds of names on the lists, and they would have to interview teachers within the district that were applying to transfer schools before interviewing other candidates. One principal said that frequently 50 or 60 people had to be interviewed for one position. Principals also noted that the previous hiring system that focused on applying to work at a specific school lead to an inequitable distribution of quality teachers. One principal said, "Before, there was a talent drain. If you [a teacher] got in a school that was not so good, you stay there until you get tenure. As soon as you get tenure, you make a name for yourself and go somewhere else." Another principal shared, "I was at a school where my first year I was struggling to find teachers to fill positions...I had the task of turning the school around without any staffing changes or people being moved out, but I had had some retirees and I spent, I don't know how long, trying to find quality people to come in. Now, I'm getting some of the best candidates in the district, which would have never happened before this [hiring process]."

Principals generally felt that the rigor of the new screening process yielded more effective teachers and that the change in the process was reflected in the number of teachers who gained tenure based on their probationary review scores. As one principal said, "The first year the probationary review team, there were a pretty high number of people that didn't get tenured. Last year it was reduced, and then this year. The first year we did the probationary review committee, we didn't have the hiring process set up. People hired whoever they could find, warm bodies...people got tenure that shouldn't necessarily have gotten tenure that have had to leave the district. There were a lot things being left up to one single person that really should have been more of a team effort."

Background on Hiring Process

This section describes the teacher hiring process in HCS. Figure 4.1 summarizes that process and maps out how candidates go from the initial contact with the school district to being hired by schools.

Figure 4.1 Process map depicting HCS screening process

Note: Author's analysis of data provided by Huntsville City Schools

Determining Personnel Needs in the District

The HCS Talent Management Department has meetings with individual principals in the spring to determine the number of positions and staffing needs for the following school year. Through this process, principals identify specific needs they have (e.g., a band teacher) and the Talent Management staff identify any internal candidates (e.g., “overage” teachers who are tenured and at a school that downsized and have to be placed at another HCS school) that would fit, as well as promising external applicants that were given a letter guaranteeing them employment in the district because (described in more detail below). Principals that participated in the focus group noted that the district offers an incentive to teachers who indicate that they will be retiring early in the year (e.g., October) to help schools identify vacancies early and aid in starting the hiring process earlier. Before the district can hire external candidates, they must place overage teachers into positions.

My observation of one of these meetings in March 2014 indicated that the process of placing overage teachers and promising external candidates was largely informal and did not appear to constrain the HCS Talent Management staff. Going into such a meeting, the HCS Talent Management staff compile a list of any teachers they need to place, including overage staff, as well as promising external candidates. At the time of the observed meeting, there were less than five overage teachers and promising external candidates on the list for all positions. During the meeting, the Talent Management personnel told me that they first try to staff the overage teachers, and then would find positions for promising candidates. As the principals identified specific vacancies at their schools, the Talent Management staff would mention the names of any

overage staff that appeared to fit the vacant position. If the principal was amenable, the overage staff would be placed at the school. In some cases, the principals were hesitant to place the overage staff at their school (e.g., because they had heard from other principals that the teacher struggled) and the vacancy was left open or a promising external candidate was placed in the position. Because the overage teachers were tenured, they were guaranteed a position at a school and HCS personnel noted that any teachers who were not placed during this process would be found a placement eventually.

Personnel Involved in Hiring Process

Three members of HCS's Talent Management Department carry out the initial stages of the selection process. The interview portion of the process is conducted by the Teacher Selection Committee, which is broken up into teams of three: two teachers nominated by their principals as "excellent teachers" and one principal nominated by the regional superintendent⁵ as an "excellent principal." One principal indicated that teachers who are on the interview committees are good teachers and therefore know what to look for when interviewing candidates when he said, "When we nominate a teacher to the committee it's because they're doing these things [asked about in the interview] and exhibiting these behaviors that we think classrooms across the city should look like." The interview committees are trained by Talent Management Department staff on how to conduct the interviews, score candidates on each of the items, and reach agreement in order to provide one interview score per candidate across the entire committee. There are multiple interview committees that serve over the course of a school year and are provided a stipend for their time. While HCS tries to keep the composition of the committees the same over the course of the year, the logistics of scheduling sometimes makes it necessary for different teachers or principals to fill in with a different group.

Application Process

In order to work in HCS, teacher candidates must apply via the statewide Teach Alabama website and submit a statement of interest, resume, evidence of certification, and letters of recommendation. During this time, teachers apply for an open vacancy in the district instead of a specific school. When there are positions that need to be filled quickly, right before the beginning of, or during, the school year, the vacancies are listed for a specific school and teachers apply to, and are chosen by, a specific school instead of HCS at large.

HCS also participates in statewide hiring fairs and recruiting fairs at nearby universities. Teachers who attend the fairs can interview with members of the HCS selection committee and undergo the full interview described below. Teach For America teachers who are part of an alternative certification program and assigned to teach in Alabama attend a statewide hiring fair that includes districts across the state. During hiring for the 2012-2013 school year, the interview

⁵ Huntsville City Schools is divided into 5 districts, each of which is supervised by a regional superintendent.

committee had discussions with Teach For America teachers before hiring them, but those teachers did not undergo the full interview. In 2013-2014 and after, Teach For America teachers underwent the full interview process. Principals can also give the names of candidates they would like to hire to the district, such as student teachers or substitute teachers that taught at their schools, but those candidates still have to go through the entire selection process before being hired at the specific school. Two focus group principals expressed a concern with the new hiring process that they would not be able to find specialized positions (e.g., certified dance instructor, AP teacher), but indicated that the ability to submit names of candidates to go through the process assuaged those concerns.

From the online pool of applicants, three members of the HCS Talent Management Department identify candidates to interview. There were no consistent criteria the three staff members used to select candidates to interview up until hiring for the 2015-2016 school year; each identified different elements of the online application they reviewed to select candidates (e.g., certification area, letters of recommendation). Starting with hiring for the 2015-2016 school year, the Talent Management Department identified specific criteria to select candidates to interview. The criteria includes: has a current Alabama teaching certificate, recently graduated and certification is pending, or qualifies for reciprocity if certified in another state; passes an Alabama and Federal Bureau of Investigation background check; has no “red flags” in the Teach Alabama application (e.g., misdemeanor, felony); there is no gap in employment history that cannot be explained by factors such as maternity or sickness; and there are two or three references out of those submitted that are “satisfactory/strong.”

The actual interviews are scheduled to last 20-25 minutes and are done in person or via Skype for candidates who are not local to Huntsville. One principal suggested that holding the interviews via Skype allowed for a broader range of candidates than if they only conducted interviews in person. Once the candidates are interviewed and given a score, the candidates’ information is placed in a spreadsheet. The structure and logistics of the interview scoring process are detailed in the following section. Candidates who receive a certain score on the interview are notified that they have been placed in the candidate pool and will be contacted by principals to be interviewed. The district personnel noted that a score of 30 out of 40 is the threshold for remaining in the candidate pool, but that the actual number varies based on the grade and subject, as well as the number of applicants and positions that need to be filled. While candidates who are added to the candidate list are not guaranteed a job and must first be hired by a principal, exceptional candidates (e.g., demonstrated effectiveness, hard to staff subjects) can be given a letter that guarantees they will be given a job in the district to entice them to work in HCS.

Description of Interview Process

The interview process in HCS consists of a standardized set of ten questions for each candidate. With the exception of special education teachers, nine of the ten questions are asked

of all candidates and one question is customized by grade level and/or subject area. The first two questions ask candidates briefly to summarize the written responses they submitted before the interview to the first two questions shown in Table 4.2. Table 4.2 lists the questions and “look-fors” for interview questions for all subjects and grades except Physical Education, ROTC, and Family and Consumer Sciences. While raters had previously discussed what to consider when they rated candidates on each question, the explicit “look-fors” were not created and used until hiring for the 2013-2014 school year. During hiring for the 2014-2015 school year, the district changed a few of the ten questions and identified the specific constructs that each question was trying to measure.

Every question is rated on a scale of 1 to 4, where a rating of 1 is “There is no evidence of knowledge or skill set,” a 2 is “There is minimal evidence of knowledge or skill set,” a 3 is “There is evidence of a good level of knowledge or skill set,” and a 4 is “There is evidence of extensive knowledge or skill set.” The committee is scheduled for a two-hour block to interview and rate four candidates. Each of the three hiring committee members individually rates the candidates during the interview and then at the end of the two-hour block, comes to one consensus rating for each candidate.

Observations of the interview process in March 2014 indicated that committee members did not always reach consensus, and when consensus was not reached, the committee went with the majority rating, or in some cases, averaged the two ratings (e.g., a rating of 2 and 3 was recorded as a consensus of 2.5). Principals that participated in the focus group noted that sometimes the interview committee did not have consensus on ratings, but that they were generally able to share discuss the situation and reach a consensus. Additionally, principals said that any inconsistencies were generally small (i.e., between a 3 and a 4 or a 2 and a 3). One principal described the inconsistencies that sometimes occurred with ratings and the fact that raters did not always rely solely on the rubric to rate candidates. He said, “With the interview questions, you have to be consistent, that everyone is getting the same questions so you are hearing consistently back about the same things...I think the rubric helps some, but sometimes when there was a debate, but someone says, ‘well, but they’re trying’ and are not sure if they should pick the candidate. My final question would be, would you want this person to teach your child? You can’t put this on an interview form, but sometimes people would say, ‘well, no.’ The discussions after each interview were pretty powerful because they were picking people for our district.”

For candidates that receive all questions (i.e., those who do not teach physical education, ROTC, or Family and Consumer Sciences), the total interview score is out of a possible 40 points. After the committee comes to a consensus, they also answer the questions: “Would your team recommend this candidate for a teaching position?” and “What are additional comments that your team would like considered regarding this candidate?” It is not clear if and how the district uses the answers to those questions as part of the selection process.

Table 4.2 Interview questions and “look-fors” by subject areas and grade level

Q#	Question	“Look Fors”	Subject
1	Summarize your response to prompt: What does 21 st Century Learning mean to you? How does this learning address the social, behavioral, and academic needs of your students?	Academic, behavioral, and social expectations addressed	All
2	Please explain the data set you analyzed and the instruction you planned to address the needs of the students. <i>How do you plan lessons so that you are meetings the diverse needs of students?</i>	Interprets data and provides examples to meet needs	All
3	How do you measure student learning and/or check for understanding. <i>Can you describe how you use formative assessments in your classroom?</i>	Can differentiate instruction and use between formative assessment	All but SpEd
3	What motivated you to select special education as a career?	Identifies 3-5 reasons that relate to special education children	SpEd
4	Briefly describe your approach to discipline/classroom management. 2012-2013: <i>What classroom structure do you feel is mutually beneficial? Why?</i>	Classroom and lesson structure – Teaching and Management	All but SpEd
4	What are the challenges of facilitating and guiding effective collaborative relationships among teachers?	Team work; Build relationships with teachers; Plan together; Students first	SpEd
5	Varies by subject (e.g., Social Science: How would you introduce current events and related topics of interest to the students through course instruction? What input would be accept of students in this area?)	Identifies specific methods – technology	All but SpEd and Elem
5	What are your beliefs about parent communication? How would you empower parents to actively participate in their child’s learning?	Importance of parent and community involvement	Elem. and SpEd
6	Briefly, describe the characteristics of an effective teacher. How do you teach so that students remember what they have learned?	Provides 3-5 characteristics and explains	All but SpEd
6	What do you consider as you write an Individual Education Plan for your students? How do you meet their individual needs?	Laws/collaboration; Student strengths & weaknesses; Goals/assessment; Involve all stakeholders	SpEd
7	What qualities and talents do you bring to this district? Please explain.	Identifies 3-5 qualities and shows willingness to serve/lead	All
8	If we walked into your classroom, what would we see? How would technology be used in your classroom?	Active engagement – facilitator of learning – technology	All
9	Describe a recent problem that you solved (this can be non-education-related).	Identifies program and solution	All
10	Do you feel ALL children can learn? Are you willing to teach in a school where HCS needs you the most?	Success of students; Strengthen school programs; Encouraged parent involvement	All

Source: Huntsville City Schools

Principal Role in Hiring Process

The final stage of the selection process is carried out by principals, who are given a wide level of discretion as to how they screen and select candidates. This part of the process generally occurs between March and June. Principals that participated in the focus group all felt that having the selection process earlier in the year was advantageous. As one principal said, “You can definitely tell the difference between the candidates in the late interviews and the early

interviews. Everybody is interviewing and picking up those shining stars, that within the first two [interview] questions, you already have that feel [that they are a good teacher].” Principals also said that they were able to fill almost all of their teaching positions before the school year begins, or even before the end of the prior school year, unless there are sudden vacancies.

When principals have an open position, the district sends the information for the four candidates with the highest interview scores that fit the given criteria (e.g., a principal can request a social studies teacher who has experience coaching basketball) to the principal. The principal can then screen the candidates as they see fit and select from those candidates, but must do so within five days. The timing, principals said, did not leave much time for an involved screening process. All principals indicated that they mainly relied on interviewing candidates, but there was variation in who was involved in those interviews. One principal said he would have a department chair sit in on the interview if he was interviewing a teacher for a subject he did not know much about. Another principal first met with the candidate herself and then let the leadership team meet with the candidate. Principals also said that they were able to provide the district with feedback on the candidates after interviewing them. For example, if a principal felt that a candidate was not strong, they could note that when they send the name of the candidate back to the district.

Most principals felt that the district’s screening process generally yielded high quality candidates, allowing them to do a less intensive screening process. Many principals also felt that it helps that the district personnel are familiar with each of the schools, allowing them to recommend specific candidates that fit the schools’ needs. If, in the case that the principal finds all of the candidates unsatisfactory and does not want to hire any, the principal must make a case to the superintendent or assistant superintendent as to why those candidates are not acceptable.

When focus group principals shared what they thought an effective teacher is, many cited specific characteristics that are asked about in the district interviews. All of the principals agreed that it was important for candidates to be student-focused and to want to teach in order to help students. One principal agreed with this sentiment and thought that the last two interview questions (i.e., do you believe all children can learn? are you willing to serve at any school in HCS?) were the most important. As he said, “That to me is an all-telling. While it’s [the answer to both of those questions] a yes, the body language and how you answer that is very, very powerful to me....we have to have employees and individuals serve some difficult positions and if they’re not there for the right reasons, I think that [the answer to those questions] illustrates why they’re here.”

Principals also talked about looking for candidates that provide quality instruction, can use data to inform instruction, and differentiate. One principal explained that a teacher’s ability to move students is important. “In our system, looking at the data right now, we are so beyond measuring our success in all of these kids that are going to score right here [at a high level] because these kids that are going to score right here and going to score right here without you. We’re more concerned about how you’re going to pull these guys up to where they need to be

and that's the beautiful thing about data. You're no longer a good teacher because you teach all the AP classes. The teacher who has the lower classes, you're not so good because your kids aren't smart enough. We're looking at movement and how you're going to move those kids."

Descriptive Statistics on Interview Process

Interview scores are available for 478 candidates who were hired (247 hired for the 2012-2013 school year and 231 hired for the 2013-2014 school year). Because this analysis is focused on the relationship between candidate interview scores and teacher evaluation scores, teachers who did not have interview scores (i.e., teachers who were already in the district and re-hired in a new position, declined the position offered, or had missing interview paperwork) are not included in the sample. Additionally, ROTC, physical education, and Family and Consumer Sciences teachers were excluded from the sample because their interview and evaluation processes are different from the rest of the teachers in the district.

Interview scores tended to be high across both school years. Of the 247 candidates hired for the 2012-2013 school year, the average percentage of total points teachers received was 83.10 (s.d. 13.25). Of the 231 candidates hired for the 2013-2014 school year, the average percentage of total points was 82.10 (s.d. 12.21).

The distribution of scores on the interview questions for candidates hired whose complete interview data is available is in Table 4.3 and Figure 4.2 below. While raters were given the option of rating on a scale of 1-4, there were a few ratings of 2.5 and 3.5. As a more conservative approach, those ratings are coded closer to the middle of the distribution, as a rating of 3 in Figure 4.2 and will be considered a rating of 3 in the analysis.

Table 4.3 Distribution of scores on interview questions

Score	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1	4	5	8	5	3	5	2	2	3	2
2	61	80	85	74	48	48	53	62	64	26
3	206	188	192	184	209	185	187	199	154	226
4	201	197	188	210	213	235	231	209	251	214
Total	472	470	473	473	473	473	473	472	472	469

Source: Huntsville City Schools

Figure 4.2 Distribution of scores on interview questions

Source: Huntsville City Schools

Huntsville City Schools' Teacher Evaluation Process

Background on Teacher Evaluation Process

The first three years teachers are employed in HCS, they are evaluated using a probationary teacher review. The review consists of six components: 1) a principal input form; 2) teacher absences; 3) student discipline referrals; 4) STAR Math and Reading SGP scores; 5) student surveys; and 6) classroom observations.

The probationary teacher review team consists of principals, instructional coaches, Directors of Instruction, and effective teachers. As with the teacher hiring interview committees, effective teachers are identified by their principals to serve on these teams and are provided a stipend to compensate them for their time.

There is a total possible score of 100 on the probationary review, however not all teachers have an SGP score, student survey score, or an observation. Only mathematics and English Language Arts teachers receive an SGP score. Only teachers in third grade and above receive student survey scores. While an observation is conducted for all first and third year teachers, second year teachers are only observed if their principal recommends that they need an observation. Since not all teachers undergo all parts of the probationary teacher review, the total score is calculated by the percent of points a teacher receives out of the total points possible for the review elements that are relevant to that position. For example, a PE teacher would not

receive a score for student SGP scores, so would have fewer maximum possible points than a reading or mathematics teacher.

Teachers who score below 65 on the overall review are not retained by the district. Teachers who score between 65 and 75 are retained with the stipulation that they attend summer learning lab (i.e., targeted professional development). Teachers who score above 75 are deemed “highly effective” and retained by the district. District personnel, including the superintendent, review all of the probationary reviews and in certain cases make exceptions to allow teachers who scored below 65 to be retained.

Many focus group principals felt that the probationary review process was just as important as the interview process. As one principal said, “That’s why the [interview and probationary] review processes go together. Let’s say you interview someone and you think they’re going to show that [be an effective teacher], but then after the first year of that review process... You can tell if a lot of people at the beginning what kind of teacher they’re going to be, but they’re always people who are going to fool you.” The principals went on to explain how the process to receive tenure in Huntsville is no longer as easy as it once was and that getting hire was just the first part of the process. One principal speculated that because candidates had heard about the review process required of teachers after they are hired, that deterred many of them from applying to teach in HCS.

Description of Teacher Evaluation Process

The distribution of how each review components was weighted changed slightly between the 2012-2013 and 2013-2014 school years for three of the components. The weighting for the principal ratings component decreased from 31% to 20%, the weighting for the student survey increased from 15% to 27%, and the weighting for the observation decreased slightly from 21% to 20%. Figures 4.3 and 4.4 show the overall point distribution of the six components of the review for the 2012-2013 and 2013-2014 school year.

Figure 4.3 Point distribution of 2012-2013 probationary teacher review

Source: Huntsville City Schools

Figure 4.4 Point distribution of 2013-2014 probationary teacher review

Source: Huntsville City Schools

The section that follows describes each component and the rubric for determining the point value a teacher receives. While the specific rubrics and criteria were generally the same across both years, the 2013-2014 rubrics added an option for teachers to receive no points on a component and made it more difficult for a teacher to get the second lowest point value. For example, if a teacher got an SGP score of 18 in 2012-2013, the teacher would have received 5 out of 15 points. A teacher who got an SGP score of 18 in 2013-2014 would receive 0 out of 15 points. The specific changes in the rubric for each component are discussed below.

Principal Input Form

Teachers receive points on their probationary review based on how their principals rated them across eight indicators. Each indicator is given a “yes” or “no” rating along with written justification. The eight indicators are:

1. Completes job requirements according to established timelines
2. Exhibits professionalism with peers, administrators, parents/guardians
3. Promotes cooperation with parents/guardians/school/community
4. Adheres to local, state, and federal procedures/policies/regulations/laws
5. Use classroom management strategies effectively
6. Instructional practices
7. List activities teacher coaches mentors
8. Additional information: specialized training/honors received/specialized skills

The principal input form was worth a total of 31 possible points in the overall review score for 2012-2013 and 20 points for 2013-2014. In both years, how many areas teachers were recommended in and the written justifications from the principal were taken into consideration. Table 4.4 below shows the criteria for assigning values for principal ratings across the two years. The possible response options are categorical, and teachers can only receive one of the three or four ratings listed on the rubric.

Table 4.4 Rubric for assigning points for principal ratings in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
0 points	NA	NA	0	The principal does not recommend the teacher and there is documentation in the teacher's HR file to support the principal's recommendation.
Low	7	Recommended in two or more areas. Documentation has been placed in the HR file. The principal does not recommend the teacher.	6	The principal does not recommend the teacher
Medium	21	Recommended in all, but one of the areas, and the justification comments indicate that the teacher meets expectations in all but one area. The principal recommends the teacher.	12	The principal recommends the teacher, but doesn't answer yes in all areas. Justification comments indicate the teacher is developing.
High	31	Recommended in all areas, and the justification comments indicate that the teacher meets and exceed expectations.	20	The principal recommends the teacher and answers yes in all areas. Justification comments indicate the teacher meets and exceeds expectations.

Source: Huntsville City Schools

Attendance

Teachers receive points on their probationary review based on the number of absences they have during the school year. Teacher attendance was worth a total of 9 points for both school years.

While the criteria for receiving a specific score was otherwise the same across both years, in 2013-2014, teachers with more than 10 days absent received 0 points instead of 3. The criteria for assigning points for absences in the probationary review is in Table 4.5 below.

Table 4.5 Rubric for assigning points for absences in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
0 points	NA	NA	0	More than 10 days absent per year
Low	3	7 or more days absent per year	3	7-10 days absent per year
Medium	6	4-6 days absent per year	6	4-6 days absent per year
High	9	0-3 days absent per year	9	0-3 days absent per year

Source: Huntsville City Schools

Student Discipline Referrals

Teachers receive points for the number of student Class I (i.e., minor) discipline referrals a teacher made during the school year. The fewer referrals a teacher made, the more points they receive.

Across both years, discipline referrals was worth a total of 9 points. While all other rating criteria were the same across both years, teachers who had more than 10 Class I Discipline

referrals received 0 points instead of 3 in 2013-2014. The criteria for assigning point values for discipline referrals in the probationary review is in Table 4.6 below.

Table 4.6 Rubric for assigning points for discipline referrals in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
0 points	NA	NA	0	More than 10 Class I Discipline Referrals
Low	3	6 or more Class I Discipline Referrals	3	6-10 Class I Discipline Referrals
Medium	6	1-5 Class I Discipline Referrals	6	1-5 Class I Discipline Referrals
High	9	No Class I Discipline Referrals	9	No Class I Discipline Referrals

Source: Huntsville City Schools

Source: Huntsville City Schools

STAR Math and Reading SGP Scores

Math and reading teachers receive points for their student test scores based on the median student growth percentile (SGP) score on the STAR math and reading tests. If teachers had both reading and math scores, the two scores were averaged. Student growth data was worth a total of 15 points in both years.

As discussed in Chapter 2, the SGP scores indicate the percent of comparable students nationally who took the same STAR test (e.g., based on past performance) who scored equal to or below the student on the assessment. While the criteria for rating SGP scores was otherwise the same across both years, teachers who received an SGP below 20 were given 0 points instead of 5 points in 2013-2014. The criteria for assigning point values for SGP scores in the probationary review is in Table 4.7 below.

Table 4.7 Rubric for assigning points for SGP scores in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
0 points	NA	NA	0	SGP is below 20
Low	5	SGP is below 35	5	SGP is 20-34
Medium	10	SGP is between 35 and 50	10	SGP is 35-50
High	15	SGP is above 50	15	SGP is above 50

Source: Huntsville City Schools

Student Survey

Teachers in third grade and above receive points on their probationary review for their student survey scores. The student survey is administered online to all students once a year and includes 18 items pulled from a variety of sources. Each item asks students to rate their teachers' behaviors in class (e.g., expects students to work hard in class, makes lessons interesting, greets students when entering the classroom) on a 3-point scale (i.e., never, sometimes, always). All students complete a survey for each of their teachers and the total scores are averaged for the teacher.

Student surveys were worth a total of 15 points in 2012-2013 and 27 points in 2013-2014. While the rating criteria in 2012-2013 differentiated between low, average, and high satisfaction, the 2013-2014 criteria specifies the specific percentage that is considered low, average, and high satisfaction. In this case, the percentage represents the overall average percent of students responding “always” to each survey question. The criteria for assigning point values in each year are in Table 4.8 below.

Table 4.8 Rubric for assigning points for student survey scores in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
0 points	NA	NA	0	Survey score is below 25
Low	5	Surveys indicate low satisfaction	9	Survey score ranges between 25 and 49 indicating low student satisfaction
Medium	10	Surveys indicate average satisfaction	18	Survey score ranges between 50 and 74 indicating average student satisfaction
High	15	Surveys indicate high satisfaction	27	Survey score ranges between 75 and 100 indicating high student satisfaction

Source: Huntsville City Schools

Classroom Observations

Teachers receive points from a classroom observation conducted during the school year on their probationary review. While teachers in their first and third year automatically receive an observation, teachers in their second year only receive an observation if their principal recommends it. The observations are conducted by members of the probationary review committee. The unannounced observations are conducted once a year and last 15-20 minutes. There are fifteen indicators across the areas of classroom environment, materials, teacher instruction, student actions, and classroom climate. The committee rates the teacher on each indicator on a scale of 1 to 3. The overall score is then determined by the total points a teacher earned out of the total possible points (i.e., 45).

While the 2012-2013 rubric differentiates between teachers that did not meet minimum expectations, met minimum expectations, and exceeded minimum expectations, the 2013-2014 identifies the specific point values for each level. The classroom observation score was worth a total of 21 points in 2012-2013 and 20 points in 2013-2014.

Table 4.9 Rubric for assigning points for observations in the teacher review

	2012-2013		2013-2014	
	Points	Criteria	Points	Criteria
Low	7	The observation indicated minimum expectations were not met.	7	The Observation Score is between 1.00-1.89, indicating minimal evidence of effective instructional practices was noted during the walk-through. The score indicates 33% to 66% of the possible points were earned.
Medium	14	The observation indicated minimum expectations were met. Potential for improvement exists.	14	The Observation Score is between 1.90-2.39, indicating some evidence of effective instructional practices was noted during the walk-through. Potential for improvement exists. The score indicates 64% to 79% of the possible points were earned.
High	21	The observation indicated minimum expectations were met, and the teacher exceeded these expectations.	20	The Observation Score is between 2.40-3.0, indicating outstanding implementation of effective instructional practices was noted. The score indicates 80% to 100% of the possible points were earned.

Source: Huntsville City Schools

Descriptive Statistics on Teacher Evaluation Process

Year 1 probationary teacher review data are available for 466 teachers (236 for Cohort 1, hired for the 2012-2013 school year, and 230 Cohort 2, hired for the 2013-2014 school year). Year 2 probationary review is available for 174 Cohort 1 teachers. The teachers who do not have probationary teacher review scores did not undergo the probationary teacher review because they resigned in the middle of the school year, were hired late in the school year, or the data is missing (e.g., missing paperwork). Figure 4.5 shows the distribution of teachers' ratings from their principals, by cohort and study year.

Of the 236 Cohort 1 teachers who underwent the Year 1 probationary teacher review in 2012-2013, 39 teachers were not retained by the district (i.e., scored below 65), 33 were recommended to be retained with learning lab (i.e., scored between 65 and 75), and 164 were recommended to be retained (i.e., scored above 75). The average percentage score for the 2012-2013 probationary teacher review was 76.79 (sd 16.50).

Of the 174 Cohort 1 teachers who underwent the Year 2 probationary teacher review in 2013-2014, 15 teacher were not retained by the district (i.e., scored below 65), 17 were recommended to be retained with learning lab (i.e., scored between 65 and 75), and 142 were recommended to be retained (i.e., scored above 75). The average percentage score for the Year 2 probationary teacher review was 83.58 (sd 12.59).

Of the 230 Cohort 2 teachers who underwent the Year 1 probationary teacher review in 2013-2014, 29 teachers were not retained by the district (i.e., scored below 65), 32 were recommended to be retained with learning lab (i.e., scored between 65 and 75), and 169 were recommended to be retained (i.e., scored above 75). The average percentage score for the Year 1

probationary teacher review for Cohort 2 teachers was 81.79 (sd 14.64). Figure 4.5 shows the distribution of recommendations across Cohorts for the 2012-2013 and 2013-2014 school years.

Figure 4.5 Distribution of retention recommendations on probationary teacher review

Note: Author's analysis of data provided by Huntsville City Schools

Below are the descriptive statistics for the individual components of the probationary teacher review. The raw scores are presented for the observations and SGP scores, but only the converted point values using the rubrics described above were available for the other components and are reported.

Principal Input Form

The vast majority of teachers received the highest possible point value (i.e., 31 out of 31 in 2012-2013, 20 out of 20 in 2013-2014) for their principal rating score, meaning that principals recommended their teachers in all areas. The distribution of principal rating scores on the probationary review are below in Table 4.10 and Figure 4.6 below.

Table 4.10 Frequencies of principal rating scores on the probationary teacher review

Rating	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
	Frequency	%	Frequency	%	Frequency	%
"No Points" (i.e., 0/20)	NA	NA	1	0.6%	3	1.3%
"Low" (i.e., 7/31, 6/20)	13	5.7%	10	5.7%	17	7.4%
"Mid" (i.e., (21/31, 12/20)	33	14.4%	14	8.0%	26	11.3%
"High" (i.e., 31/31, 20/20)	183	79.9%	149	85.6%	185	80.1%
Total	229		174		231	

Note: Author's analysis of data provided by Huntsville City Schools

Figure 4.6 Distribution of principal rating scores on probationary teacher review

Note: Author's analysis of data provided by Huntsville City Schools

Attendance

The majority of teachers received the highest possible point value (i.e., 9 out of 9) for their attendance score, meaning that teachers were absent three or fewer days throughout the school year. Generally speaking, Cohort 2 years tended to have higher attendance scores and fewer absences than Cohort 1 teachers. The distribution of attendance scores on the probationary review are below in Table 4.11 and Figure 4.7 below.

Table 4.11 Frequencies of attendance scores on the probationary teacher review

Rating	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
	Frequency	%	Frequency	%	Frequency	%
"No Points" (i.e., 0/9)	NA	NA	3	1.8%	6	2.6%
"Low" (i.e., 3/9)	26	11.6%	14	8.2%	8	3.5%
"Mid" (i.e., 6/9)	51	22.8%	53	31.0%	31	13.7%
"High" (i.e., 9/9)	147	65.6%	101	59.1%	182	80.2%
Total	224		171		227	

Note: Author's analysis of data provided by Huntsville City Schools

Figure 4.7 Distribution of attendance scores on probationary teacher review

Note: Author's analysis of data provided by Huntsville City Schools

Discipline Referrals

About half of the teachers in each cohort and year received the highest possible point value (i.e., 9 out of 9) for their discipline score, meaning that they made no Class I discipline referrals. Generally speaking, Cohort 1 teachers in their Year 1 reviews made more discipline referrals and consequently received fewer points for discipline than Cohort 2 teachers or Cohort 1 teachers in their second year review. The distribution of discipline scores on the probationary review are below in Table 4.12 and Figure 4.8 below.

Table 4.12 Frequencies of discipline scores on the probationary teacher review

Rating	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
	Frequency	%	Frequency	%	Frequency	%
"No Points" (i.e., 0/9)	NA	NA	2	1.1%	12	5.2%
"Low" (i.e., 3/9)	49	22.8%	14	8.0%	26	11.3%
"Mid" (i.e., 6/9)	73	34.0%	60	34.5%	66	28.6%
"High" (i.e., 9/9)	93	43.3%	98	56.3%	127	55.0%
Total	215		174		231	

Note: Author's analysis of data provided by Huntsville City Schools

Figure 4.8 Distribution of discipline scores on probationary teacher review

Note: Author's analysis of data provided by Huntsville City Schools

STAR Math and Reading SGP Scores

On average, teachers across both cohorts and years scored received SGP scores of 50 or better, indicating that their students performed better than expected based on the performance of similar students nationally. Table 4.13 below summarizes the SGP scores in reading and math.

Table 4.13 Descriptive statistics on available SGP scores

	Math			Reading		
	<i>n</i>	<i>Avg (sd)</i>	<i>Range</i>	<i>n</i>	<i>Avg (sd)</i>	<i>Range</i>
Cohort 1 Year 1	41	54.83 (12.29)	12, 81	44	60.86 (12.75)	36, 83
Cohort 1 Year 2	58	57.28 (13.40)	29, 90	75	65.29 (17.23)	20, 97
Cohort 2 Year 1	102	54.52 (13.59)	21, 85	114	57.83 (19.84)	13, 93

Note: Author's analysis of data provided by Huntsville City Schools

The SGP scores were not available for all teachers who had one (e.g., missing paperwork), so the point value teachers received for their SGP scores on the probationary review is presented in

Table 4.14 and Figure 4.9 below. As the raw SGP scores indicate, the majority of teachers received the maximum points for the SGP score, meaning their SGP score was 50 or above. While there was larger variation in the raw SGP scores, the three or four rating categories created for the probationary review limited that variation.

Table 4.14 Frequencies of points on the probationary teacher review from SGP scores

Rating	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
	Frequency	%	Frequency	%	Frequency	%
"No Points" (i.e., 0/15)	0	0.0%	0	0.0%	1	0.7%
"Low" (i.e., 5/15)	7	7.4%	5	5.7%	10	7.4%
"Mid" (i.e., 10/15)	31	33.0%	15	17.0%	37	27.4%
"High" (i.e., 15/15)	56	59.6%	66	75.0%	87	64.4%
Total	94		86		135	

Note: Author's analysis of data provided by Huntsville City Schools

Figure 4.9 Distribution of points on probationary teacher review from SGP scores

Note: Author's analysis of data provided by Huntsville City Schools

Student Survey

On average, teachers in Cohort 2 received more points than Cohort 1 teachers on the student survey portion of the probationary review. While 80% of Cohort 2 teachers received the

maximum point value on their probationary review, only 17% of Cohort 1 teachers received the maximum points in Year 1 and 59% received the maximum points in Year 2. In Year 1, Cohort 1 teachers were rated using a more subjective rubric with indicators of “low,” “average,” and “high” student satisfaction. Such general terms likely led to a wide variety of interpretations and could have led to harsher judgments. In Year 2, then, the rubric included specific numbers that applied to each rating (e.g., 25-49% is low satisfaction). Thus, it is possible that the dramatic increase in review scores for Cohort 1 teachers was due to changes in the rubric. The distribution of student survey scores on the probationary review are below in Table 4.15 and Figure 4.10 below.

Table 4.15 Frequencies of student survey scores on the probationary teacher review

Rating	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
	Frequency	%	Frequency	%	Frequency	%
“No Points” (i.e., 0/27)	NA	NA	2	1.8%	0	2.6%
“Low” (i.e., 5/15, 9/27)	37	16.5%	1	8.2%	11	3.5%
“Mid” (i.e., 10/15, 18/27)	70	31.3%	26	31.0%	49	13.7%
“High” (i.e., 15/15, 27/27)	38	17.0%	84	59.1%	98	80.2%
Total	145		113		158	

Note: Author’s analysis of data provided by Huntsville City Schools

Figure 4.10 Distribution of student survey scores on probationary teacher review

Note: Author’s analysis of data provided by Huntsville City Schools

Classroom Observations

Compared to the other probationary review components, the classroom observations had the most varied ratings, given the categories that the district created. For both cohorts and years, approximately 30-40% of teachers received the “low” rating (i.e., 7 out of 21, 7 out of 20) indicating that they did not meet minimum expectations set forth on the observation protocol. Generally, teachers in the 2013-2014 school year (i.e., Cohort 1 Year 2, Cohort 2 Year 1) performed better than teachers in 2012-2013 (i.e., Cohort 1 Year 1). The distribution of observation scores on the probationary review are below in Table 4.16 and Figure 4.11 below.

Table 4.16 Frequencies of observation points on the probationary teacher review

	Cohort 1 Year 1		Cohort 1 Year 2		Cohort 2 Year 1	
Rating	Frequency	%	Frequency	%	Frequency	%
“No Points” (i.e., 0/20)	NA	NA	0	0%	0	0%
“Low” (i.e., 7/21, 7/20)	85	39.9%	63	38.4%	77	29.4%
“Mid” (i.e., 14/21, 14/20)	103	48.4%	43	26.2%	70	20.1%
“High” (i.e., 21/21, 20/20)	25	11.7%	58	35.4%	67	27.1%
Total	213		164		214	

Note: Author’s analysis of data provided by Huntsville City Schools

Figure 4.11 Distribution of observation scores on probationary teacher review

Note: Author’s analysis of data provided by Huntsville City Schools

Teacher Retention

An additional outcome of interest is whether or not teachers continue working through the end of their second year at the district. Teachers fail to complete their second year in the district for three reasons: 1) they scored poorly on their teacher probationary review and were not recommended to be retained in their second year; 2) they were asked to leave the school district for an issue not related to performance (e.g., extremely poor attendance, other disciplinary actions); or 3) they left the district voluntarily. While the data indicates if teachers were no longer employed at the district because of poor performance, it does not identify teachers who were asked to leave for another reason or left the district voluntarily.

Retention data is only available on the 2012-2013 Cohort (i.e., Cohort 1) of teachers. Of the 247 teachers in Cohort 1, 174 were retained until the end of their second year teaching (i.e., have a probationary review score for their second year) and 73 were not retained until the end of their second year teaching (i.e., do not have a probationary review score for their second year). Of the 73 who were not retained, 39 were let go because of poor performance on their first year probationary review. Data is not available on why the other 34 teachers were not retained; this could be because the teacher voluntarily left the district, or was asked to leave for a reason other than a low probationary review score (e.g., absenteeism, misconduct).

Conclusion

Throughout the teacher hiring and probationary review processes, there are changes that could be made to improve the quality of data HCS gets on teachers and consequently, help to refine its selection process and identify more effective teachers. Over the two years included in this research, HCS has already changed certain elements of these processes that are a move in that direction.

For example, the probationary review rubrics in 2013-2014 included specific numeric indicators that made the review process more objective than the previous year where indicators were more subjective, such as “low satisfaction.” However, as many of the descriptive statistics presented in this chapter demonstrate, the way the probationary review rubrics are constructed limit the variation in teachers’ scores. The rubrics have four possible categories teachers can fit into, and the majority of teachers tend to fall in the higher two categories. It is not clear whether the scores are skewed high because of the way raters are trained or the way they interpret and assign the highest rating.

Also related to probationary teacher review data, not all teachers receive all components of the review. Only mathematics and reading teachers receive SGP scores. Not all teachers in their second year receive observations. Only teachers in third grade and above receive student survey scores. While there are reasons why teachers do not receive scores on these elements, such a system limits the data available on many teachers’ reviews and may skew the scores teachers receive. Furthermore, the district has changed the weighting assigned to each review component

over time. In the most recent iterations, the student survey, SGP scores, and observations together consistent of 63% of a teacher's score, yet not all teachers receive each of those components.

Elements of the teacher interview process have also been improved since hiring for the 2012-2013 school year. Over time, larger constructs, specific "look fors," and rubrics were added to help interviewers more subjectively rate candidates. However, there are still areas for improvement. While there are ten very different interview questions, they are all rated using the same rubric of 1 (i.e., There is no evidence of knowledge or skill set) to 4 (i.e., There is evidence of extensive knowledge or skill set). Such a rubric isn't as appropriate for questions such as "Describe a recent problem that you solved (this can be non-education-related)" or "Do you feel ALL children can learn? Are you willing to teach in a school where HCS needs you the most?"

Interviews in March 2014 with the HCS Talent Management Department indicated that the staff members responsible for selecting applicants to interview used very different criteria. Thus, there were inconsistencies and it's possible that a candidate overlooked by one staff member would have been selected for an interview by another staff member. Interviews in April 2015 revealed that the department created a flow chart with the criteria to determine if an applicant should be interviewed or not, thus creating a more systematic process.

Finally, there are limitations to the data HCS gathers for teachers. While there is information on the candidates that HCS hires, it does not systematically track candidates who are not hired, in terms of characteristics and interview scores.

5. RQ2: Quantitative Analyses

Introduction

This chapter seeks to answer the following research question: What existing instruments or measures hold promise for improving HCS's teacher hiring process? In order to answer that question, I present graphical representations and (linear or logistic) regressions of the relationship between candidate interview ratings and different measures of teacher effectiveness and persistence, including overall probationary review scores and teacher retention in Year 2 of teaching. Then, I present findings that explore the relationship between the overall interview score and individual components of the probationary teacher review (i.e., principal ratings, observation scores, absences, disciplinary actions, student survey, and SGP scores). Finally, I present findings from the factor analysis of candidate interview questions and describe a modified interview score that was informed by this analysis, and discuss the relationship between this measure and the teacher effectiveness outcomes.

Overall Interview Score and Overall Probationary Review Score

Interview Score and Year 1 Review

The Year 1 probationary review score includes the 2012-2013 review score for Cohort 1 teachers and the 2013-2014 score for Cohort 2 teachers. The specific components of the interview and probationary review processes and scores, along with descriptive statistics for each, are detailed in Chapter 4. Figure 5.1 plots teachers' interview scores on the x-axis against their first-year probationary review scores on the y-axis (standardized to have a mean of 0 and a standard deviation of 1 within year). Cohort 1 is represented with blue diamonds, and Cohort 2 with red circles. For both cohorts, we see a denser concentration of points on the top right side of the graph, indicating (as noted in Chapter 4) that the scores are concentrated on the higher ends of each scale. Through visual inspection, I find no clear indication that higher interview scores are related to higher—or lower—first-year review scores for either cohort.

Figure 5.1 Standardized Year 1 probationary review scores and interview scores by cohort

However, it is possible that adjusting for certain features of the probationary review scores, such as whether or not they include measures of student learning or teachers' classroom practices, would reveal a relationship. To examine the relationships while adjusting for these features of the review scores, we turn to the regression coefficients in Table 5.1. In Table 5.1, I present coefficients from OLS regressions of the first-year probationary review scores on candidate interview scores, with controls for teachers' cohort (columns 1 and 2) and for the included components of the probationary review score (column 2). The details of the statistical models that are shown below are described in Chapter 3. As noted in Chapter 3, the adjusted p-value is .0026 and only significant relationships at that level are noted.

Table 5.1 Coefficients (and standard errors) from of Year 1 probationary review scores on candidate interview scores

VARIABLES	Standardized Year 1 Review Score	Control for Missing Review Components
Candidate Interview Score	0.0043 (0.0038)	0.0037 (0.0039)
Title I school = 1	-0.143 (0.0988)	-0.148 (0.0974)
SGP is Part of Year 1 Interview Score		-0.0016 (0.0011)
Observation is Part of Year 1 Interview Score		0.0056* (0.0014)
Student Survey is Part of Year 1 Interview Score		-0.0043* (0.0012)
Cohort 1=1	0.0631 (0.0962)	0.215 (0.103)
Constant	-0.363 (0.320)	-0.434 (0.311)
Observations	463	463
R-squared	0.008	0.073

*p<.0026

Table 5.1 shows there is no statistically significant relationship between the overall candidate interview score and the overall probationary review score in Year 1. When controlling for the presence of particular review score components (i.e., the inclusion of observation, student survey, and SGP scores in the composite score) by incorporating dummy variables that interact with the overall interview score, I find no statistically significant relationship between interview score and review score. On average, Cohort 1 teachers scored a quarter of a standard deviation higher in their Year 1 review than Cohort 2 teachers, but that difference was not statistically significant, given the adjusted p-value. Additionally, having a Year 1 observation score is significant and associated with an overall review score that is .0056 of a standard deviation higher than for those without that score. Having a Year 1 student survey score is significant at the .0026 level and associated with an overall review score that is .0043 of a standard deviation lower. In other words, teachers who receive as observation score as part of their probationary review tend to do better on the overall review, whereas teachers who have the student survey as part of their probationary review tend to do worse on the overall probationary review.

Interview Score and Year 2 Review

Teachers who score below 65% on their probationary review are generally not retained by HCS the following year. Thus, we would expect the lower performing teachers to be removed from the sample in Year 2. As noted in Chapter 3, there is a modest correlation between Year 1

and Year 2 scores (.3165) among Cohort 1 teachers who have a Year 1 and Year 2 review score. Figure 5.2 shows the relationship between candidate interview scores, on the x-axis, and Year 2 probationary review scores, on the y-axis. The pattern of points on the graph shows a slight positive relationship between the Year 2 review score and candidate interview scores, indicating there could be a positive correlation if there were no ceiling effects on the review and interview scores.

Figure 5.2 Interview score percentage and Year 2 probationary review scores

In order to further explore this relationship, I ran a regression of Year 2 probationary review scores on candidate interview scores. In Table 5.2, I present coefficients from the OLS regressions of the second-year probationary review scores on candidate interview scores, with controls for first-year probationary review scores (columns 1 and 2) and for the included components of the probationary review score (column 2). Table 5.2 shows there are no statistically significant relationships between the overall interview score and the overall probationary review score in Year 2 when controlling for the presence of particular review score components (i.e., the inclusion of observation, SGP, and student survey scores in the composite score) (column 2). The relationships between the interview score and whether or not a teacher had an observation score and Year 2 review scores were marginally significant, with p-values of

.078 and .018, respectively. Despite the lack of significance, for every one percent increase in interview score, we would expect an increase in Year 2 probationary review of .0093 standard deviations. Additionally, having an observation as part of the Year 2 review is associated with a decrease in Year 2 review scores by .004 standard deviations.

Table 5.2 Coefficients (and standard errors) from OLS regressions of Year 2 probationary review scores on candidate interview scores

VARIABLES	Standardized Year 2 Review	Control for Missing Review Components
Candidate Interview Score	0.0079 (0.0053)	0.0093 (0.0053)
Title I school = 1	-0.0014 (0.145)	-0.0499 (0.145)
SGP is Part of Year 2 Review Score =1		-0.0003 (0.0017)
Observation is Part of Year 2 Review Score =1		-0.0040 (0.0017)
Student Survey is Part of Year 2 Review Score =1		-0.0024 (0.0017)
Constant	-0.581 (0.455)	-0.406 (0.453)
Observations	170	170
R-squared	0.014	0.059

*p<.0026

Overall Interview Score and Retention in Year 2

Retention in Year 2 (i.e., whether or not a teacher had a Year 2 probationary review score) is only available for Cohort 1 teachers. Figure 5.3 shows the relationship between Cohort 1 candidates' interview scores (x-axis) and Year 1 probationary review scores (y-axis), coded by whether or not they were retained. Teachers who were retained are noted as blue diamonds and teachers who were not retained are coded as red circles. The horizontal line at .65 indicates the threshold for being retained by the district; the district does not retain teachers who score below .65, unless there is approval from the Superintendent. The three teachers who were retained despite scoring below .65 on their Year 1 review received .89 or higher on their Year 2 reviews. Thus, in the case of these three teachers, Year 1 probationary review scores were not predictive of their Year 2 review scores. Aside from the fact that the majority of teachers who scored below .65 were not retained, there appears to be no clear association between teachers' interview scores, Year 1 review scores, and whether or not they were retained in Year 2. Of the teachers who scored above .65 on the Year 1 probationary review but were not retained, it is not clear why those teachers left the district; presumably many of these were voluntary departures.

Figure 5.3 Interview score percentage and Year 1 probationary review scores by retention

Table 5.3 shows the results (reported as odds ratios) of a logistic regression showing the relationship between overall candidate interview score and Year 2 retention, controlling for standardized Year 1 review scores (columns 2 and 3) and missing review components (column 3). There is no statistically significant relationship between overall interview score and being retained in Year 2. This is also true when controlling for Year 1 review score, as well as missing interview components. Not surprisingly, Year 1 review scores are significantly associated with a higher likelihood of being retained in Year 2. More specifically, a one-unit increase in a teacher's Year 1 Review Score is associated with 4.3 and 4.6 higher odds of being retained.

Table 5.3 Odds ratios (and standard errors) from logistic regressions of Year 2 retention on candidate interview scores

VARIABLES	Retained in Year 2	Control for Year 1 Review Score	Control for Missing Review Components
Interview Percent	0.999 (0.0108)	1.001 (0.0134)	.997 (0.0139)
Title I School==1	0.543** (0.155)	0.707 (0.246)	.737 (0.260)
SGP is part of Year 1 review = 1			1.002 (0.0042)
Observation is part of Year 1 review = 1			
Student Survey is part of Year 1 review = 1			1.004 (.0046)
Year 1 Probationary Review Score = 1		4.289* (0.899)	4.544* (1.003)
Constant	3.227 (2.973)	2.923 (3.331)	2.977 (3.416)
Observations	242	234	234

*p<.0026

Overall Interview Score and Probationary Review Components

As noted in Chapter 4, there is not much variability on teachers' scores on the probationary review components. In fact, each component has only three or four scoring options. It is important to consider the lack of variability when reviewing the findings related to the relationship between individual probationary review components and candidate interview scores below.

Principal Ratings

Table 5.4 shows the results of an ordinal logistic regression looking at the relationship between the level of rating (i.e., 0, low, middle, high) a teacher received from their principal on their review and their overall interview score, controlling for whether or not a teacher was in a Title I school. The results from the Year 1 review are presented in column 1 and the results from Year 2 are presented in column 2. The odds ratios reported in Table 5.4 show the relative ratio of odds between receiving a high rating and each other level of rating for the principal. There is no statistically significant relationship between candidate interview scores and principal ratings on the probationary review. While none of the relationships are significant, the odds ratios indicates a one-percentage increase in a candidate's interview score is associated with .986 higher odds of receiving a low versus mid-level or mid- versus high level principal rating in Year 2. This is essentially the same as an increased interview score not impacting the odds of receiving a higher review score.

Table 5.4 Odds ratios (and standard errors) from ordinal logistic regressions of principal ratings on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	0.989 (0.0088)	0.986 (0.0150)
Title 1 School=1	1.375 (0.322)	1.362 (0.602)
Observations	456	170

*p<.0026

Observation Scores

Table 5.5 shows the relationship between the level of rating (i.e., 0, low, middle, high) a teacher received from their principal on their review and their overall interview score, controlling whether or not a teacher was in a Title I school. There is no statistically significant relationship between the two variables in Year 1 or Year 2. However, the relationship between teaching in a Title I school in Year 1 and observation scores was marginally significant ($p<.078$) and associated with an odds ratio of .721. In other words, teachers from Title I schools are less likely (i.e., have .721 times lower odds) to receive a higher observation score (e.g., high versus middle and low, middle and high versus low) than teachers from non-Title I schools.

Table 5.5 Odds ratios (and standard errors) from ordinal logistic regressions of observation scores on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	1.000 (0.0073)	0.989 (0.0114)
Title 1 School=1	0.721 (0.134)	.689 (0.214)
Observations	422	160

*p<.0026

Attendance

Table 5.6 shows the relationship between the score (i.e., 0, low, mid-range, or high) a teacher received on their probationary review for the number of absences they had in the year and their overall interview score, controlling for whether or not a teacher was in a Title I school. There is no statistically significant relationship between attendance scores and interview scores in Year 1 or Year 2. However, the relationship between teaching in a Title I school in Year 1 and

attendance scores was marginally significant ($p < .029$) and associated with an odds ratio of .628. In other words, teachers from Title I schools are less likely (i.e., have .628 times lower odds) to receive a higher attendance score (e.g., high versus middle and low, middle and high versus low) than teachers from non-Title I schools.

Table 5.6 Odds ratios (and standard errors) from ordinal logistic regressions of attendance scores on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	0.992 (0.0084)	0.995 (0.0117)
Title 1 School=1	.628 (0.134)	.735 (0.235)
Observations	447	167

* $p < .0026$

Disciplinary Actions

Table 5.7 shows the relationship between the score (i.e., 0, low, mid-range, or high) a teacher received on their probationary review for the number of discipline referrals they had in the year and their overall interview score, controlling for whether or not a teacher taught in a Title I school. There is no statistically significant relationship between discipline scores and interview scores in Year 1 or Year 2. However, the relationship between teaching in a Title I school in Year 2 and disciplinary action scores was marginally significant ($p < .060$) and associated with an odds ratio of .552. In other words, teachers from Title I schools are about half as likely (i.e., have .552 times lower odds) to receive a higher attendance score (e.g., high versus middle and low, middle and high versus low) in Year 2 than teachers from non-Title I schools.

Table 5.7 Odds ratios (and standard errors) from ordinal logistic regressions of disciplinary scores on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	.994 (0.0103)	1.010 (0.0114)
Title 1 School=1	.956 (0.253)	.552 (0.174)
Observations	229	170

* $p < .0026$

Student Survey

Table 5.8 shows the relationship between the score (i.e., 0, low, mid-range, or high) a teacher received on their probationary review for their student surveys and their overall interview score, controlling for whether or not a teacher was in a Title I school. There is no statistically significant relationship between survey score ratings and interview scores in Year 1 or Year 2.

Table 5.8 Odds ratios (and standard errors) from ordinal logistic regressions of student survey scores on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	1.006 (0.0086)	1.005 (0.0171)
Title 1 School=1	0.844 (0.193)	.924 (0.445)
Observations	301	109

*p<.0026

SGP Scores

Table 5.9 shows the relationship between shows the relationship between the score (i.e., 0, low, mid-range, or high) a teacher received on their probationary review for their SGP scores and their overall interview score, controlling for whether or not a teacher was in a Title I school. There is no statistically significant relationship between the two variables in Year 1 or Year 2. However, the relationship between teaching in a Title I school and SGP scores was marginally significant in Year 1 ($p<.053$) and Year 2 ($p<.086$) and associated with odds ratios of .1.008 and .9817, respectively. In both years, teachers in Title I schools are only slightly more likely (in Year 1) or less likely (in Year 2) to receive a higher SGP score (e.g., high versus middle and low, middle and high versus low) than teachers from non-Title I schools.

Table 5.9 Odds ratios (and standard errors) from ordinal logistic regressions of SGP Scores on candidate interview scores, by teaching year

VARIABLES	Year 1	Year 2
Interview Percent	1.008 (0.0109)	0.982 (0.0190)
Title 1 School=1	1.701 (0.467)	2.465 (1.295)
Observations	226	85

*p<.0026

Teacher Interview Factors and Teacher Effectiveness Outcomes

Factor Analysis

I also conducted a factor analysis to determine if the individual interview questions form underlying constructs that may be related to the overall probationary review or its individual components. Table 5.10 below shows the correlations between the nine survey items that were common across 439 of the teacher interviews. Special education and physical education teachers, as well as counselors, were excluded from this analysis since they were not asked the same set of nine questions on their interviews. The correlations for all of the questions are moderate, ranging from .303 to .608.

Table 5.10 Correlation matrix between interview questions

	High expectations (Q1)	Analyzing data (Q2)	Assessment /Differentiation (Q3)	Classroom management (Q4)	Effective teacher characteristics (Q6)	Qualities/Talents (Q7)	Use of technology (Q8)	Problem solving (Q9)	Belief all students can learn (Q10)
High expectations (Q1)	1.00								
Analyzing data (Q2)	0.468	1.00							
Assessment/ Differentiation (Q3)	0.496	0.423	1.00						
Classroom management (Q4)	0.473	0.422	0.488	1.00					
Effective teacher characteristics (Q6)	0.431	0.303	0.371	0.454	1.00				
Qualities/Talents (Q7)	0.427	0.326	0.368	0.468	0.478	1.00			
Use of technology (Q8)	0.368	0.377	0.414	0.472	0.366	0.427	1.00		
Problem solving (Q9)	0.380	0.385	0.321	0.380	0.266	0.428	0.345	1.00	
Belief all students can learn (Q10)	0.608	0.530	0.546	0.593	0.477	0.542	0.529	0.49	1.00

The factor analysis on the nine items, indicates that there is one underlying factor with an eigenvalue greater than 1. While the first factor had an eigenvalue greater than 4, the next factor had an eigenvalue of .33. Therefore, I created a single scale by adding the values of all nine questions.

As seen in Table 5.11, the overall alpha coefficient the nine items is .8734. The internal consistency for all variables on the first factor is high (i.e., above .80) and removing any individual variables does not increase the alpha coefficient. Therefore, all nine items can be retained.

Table 5.11 Overall and item-level alpha coefficients

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
Q1	445	+	0.7358	0.6512	.2220821	0.8564
Q2	444	+	0.6745	0.5654	.2251526	0.8642
Q3	446	+	0.7058	0.6033	.2206598	0.8607
Q4	446	+	0.7548	0.6685	.2159242	0.8542
Q6	446	+	0.6626	0.5605	.2301568	0.8644
Q7	446	+	0.7077	0.6192	.2261572	0.8592
Q8	445	+	0.6785	0.5802	.2283609	0.8626
Q9	445	+	0.6409	0.5298	.2314275	0.8674
Q10	442	+	0.8309	0.7813	.2202974	0.8483
Test scale					.2244682	0.8734

Relationship between Revised Interview Score and Outcomes

The analyses at the beginning of this chapter that included the overall interview score as the independent variable are repeated below using the nine-item scale constructed based on the factor analysis results. Below, the independent variable is the unweighted composite interview score that includes all nine questions from the factor analysis, excluding question 5 that was not common across all teachers. This analysis also excludes the special education and physical education teachers and counselors. Because the factor analysis yielded a revised interview score that was very similar to the overall interview score, the findings from the additional analysis yielded similar results as the analyses conducted with the overall interview score detailed above and are therefore not presented again below.

Conclusion

Overall, there were few significant relationships between the candidate interview scores and the outcomes of interest. In some cases, there were significant relationships between whether candidates did or did not have certain review components (i.e., student survey, SGP, observation scores) and overall review scores.

It is possible that without ceiling effects and with finer review score categories (i.e., beyond low, medium, and high), marginally insignificant relationships, such as that between Year 2 review scores and interview scores, would be significant.

Furthermore, data is not available for teachers who were not hired. While teachers may not have been hired for a variety of reasons (e.g., low score on interview, no position available, candidate decided not to take job), many candidates who were not hired were likely those with low interview scores. Thus, the sample of candidates who were hired likely have higher

interview scores, causing a restricted range of interview scores. It is possible, then, that including data from candidates who were not hired who performed poorly on the interview, may yield significant relationships.

Recommendations for improving the interview process, along with the probationary review process, will be addressed in the final chapter. Such recommendations can help improve the interview and review processes, and consequently lead to significant findings using data from future cohorts. Given that the HCS interview scores shows no significant relationship to teacher outcomes, I will look at data sources used in other districts to help understand what data sources may identify more effective teachers. Those data sources, along with their costs and advantages, are reviewed in the next chapter.

6. RQ3: Alternatives to HCS' Current Hiring Process

Introduction

As discussed in Chapter 2, the majority of studies on teacher selection and hiring focus on describing the teacher selection and hiring processes in various districts. At the same time, there is little to no focus on the advantages and disadvantages of selection tools or which candidate characteristics are related to later teacher effectiveness. While such information could help districts make decisions about how they select teachers, there is little guidance from the literature. To that end, this chapter pulls from the literature on teacher hiring and teacher effectiveness, where available, and creates a broad synthesis to inform the discussion.

In order to inform which selection tools may be best for a district to use, I focus on three criteria for each: objectivity, simplicity and costs of administering, and proximity to classroom practice. Objectivity refers to the extent to which the information provided in the selection tool is objective versus subjective. Rubrics could be developed for any of the more subjective measures to make ratings more standardized and less objective, but that involves an additional cost in terms of developing the rubrics and training personnel to norm their ratings. While rubrics could be applied to many of the data sources, the discussion that follows considers the objectivity of the data source without the use of a rubric. Simplicity refers to the amount of effort it takes for the district to gather the information, particularly personnel and time costs. Proximity to classroom practice refers to the extent to which a selection tool measures characteristics that are closely tied to what a teacher would do in the classroom. In addition to looking at these three criteria, I also identify candidate characteristics that can be measured by each selection tool and the extent to which those characteristics are related to teacher effectiveness.

This chapter focuses on multiple elements districts should consider when identifying the tools and characteristics they might want to use when selecting teachers. It is intended to support districts' decision-making processes, not to identify the one best tool or characteristic that all districts should use. Before a district selects any instruments, it is important to identify the teacher characteristics they wish to prioritize, as well as the depth of information they want to obtain for each. For example, if a district is interested in academic qualifications, it could look at GPA, degree received, grades in specific classes, the difficulty of courses taken, or the ranking of college attended. While all of this data provides insights into a candidate's academic qualifications, there are varying levels of depth of information to each. Furthermore, districts should consider the literature on teacher effectiveness, which is presented throughout this chapter. It is important to note that while much of this literature provides mixed results in terms of which teacher characteristics are related to effectiveness, the majority of these studies focus on teacher impact on student test scores. Thus, it's important to consider what outcomes a district

wants. These could include student outcomes that have not been traditionally measured in the teacher effectiveness literature that are nonetheless important, such as students' interpersonal skills, motivation, or civic mindedness.

In addition to identifying characteristics of interest, part of the process is to determine which tools are most appropriate at which points in the selection process. Those tools that require less time can be used to screen initial candidates and ensure that they meet basic requirements, while the more time and personnel intensive tools can then be used with smaller numbers of candidates after the initial round of review. Similarly, information that is easier to gather, but more removed from classroom practice, can be gathered during initial stages (e.g., application, resume), and as candidates advance in the selection process, instruments that are more costly, but are closer to actual classroom practice, can be used (e.g., sample lesson). Again, the discussion that follows is intended to provide things to consider when structuring a selection process, which may be structured in numerous ways.

Teacher Screening Tools and Associated Teacher Characteristics

This section covers the various teacher screening and selection tools a district might use. While the findings from this chapter could inform any district or school, the focus of this research is on Huntsville City Schools (HCS). Therefore, I first provide a discussion of the selection tools that HCS is currently using or has access to and then turn to other tools that they might be able to use. Additionally, I note throughout any additional contextual information relevant to HCS. Tables 6.1 and 6.2 summarize the costs and advantages of each selection tool, along with the potential teacher characteristics that could be measured with each described throughout this chapter.

Table 6.1 Matrix of costs and advantages for teacher selection tools

	Objectivity	Costs	Proximity to Classroom Practice
Application/Resume	high	low	low
Transcripts	high	low	low
Interview	low	high	moderate
Writing Sample	low	moderate	moderate
References	moderate	moderate	moderate
Teacher Tests	high	low	moderate
Sample Lesson	low	high	high
Portfolio/Teaching Artifacts	low	moderate	high

Table 6.2 Characteristics measured by each selection tool

	Teaching Practice	Academic Qualifications/ Content Knowledge	Interpersonal Skills	General Personal Attributes	Attitudes, Beliefs, and Values	Work with Extracurricular Activities	Fit with Culture
Application		x				x	
Resume		x				x	
References	x	x	x	x	x	x	x
Transcripts		x					
Portfolio	x	x			x	x	x
Tests of Teaching Practice	x	x					
Writing Sample/Prompt	x	x	x	x	x	x	x
Interview	x	x	x	x	x	x	x
Sample Lesson	x	x	x	x			x

Screening Tools Huntsville City Schools (HCS) is Already Using or Has Access to

Application/Resume

Applications and resumes can provide a variety of information with little cost to the district. Since every candidate that applies to teach in Alabama must complete an application and submit a resume through the Teach Alabama website, HCS already has access to that information with no extra cost to the district. Because of the low cost of gathering the information, applications and resumes are good for first round screening of large numbers of applicants.

Applications can provide information about candidate qualifications, both academically and in terms of teacher experience and certification. Assuming candidates are honest, this information has high objectivity. However, the information provided on a resume is generally pretty removed from classroom practice and the depth of information can be limited. While these data sources do provide information on candidate qualifications, it provides basic information about years of experience, teacher certification, GPA, degrees earned, and schools attended.

There are mixed findings in terms of the relationship between years of teaching experience, academic qualifications, and certification and teacher effectiveness. Research has found that years of teaching experience matter in terms of student test scores, but mainly in the first three to five years of teaching. Goe (2007) conducted a research synthesis of studies on various teacher characteristics and student outcomes and found that teacher experience did matter for student outcomes, but only in the first five years of teaching (2007). Similarly, studies of teachers in New York and North Carolina found that the majority of gains in student test scores occurred in the first two years of teaching (Clotfelter, Ladd, & Vigdor, 2010; Kane et al., 2008). Nye,

Konstantopoulos, and Hedges (2004) conducted a 4-year study in which teachers were randomly assigned to students to estimate the effect of teachers on student achievement. They found that teacher experience accounted for differences in student achievement in a few cases (i.e., 2nd grade reading and 3rd grade mathematics). Finally, Harris and Sass modeled the productivity of teachers, in terms of promoting student achievement, in all public schools in Florida between 1999 and 2005. They found that elementary and middle school teachers generally increased in productivity in their first five years teaching and more modestly beyond that, while high school teachers with more experience tended to be less productive (Harris & Sass, 2011).

The research on teacher certification status has yielded mixed results in terms of its relationship to student outcomes. Goe's (2007) research synthesis found that while teachers having degrees and appropriate certification mattered in mathematics at the elementary level and more so at the secondary level, these qualifications did not matter for other subjects. The study of teachers in North Carolina found that teachers with standard teaching licenses performed better than those with other types of certifications (Clotfelter et al., 2010). A study of the data from the Early Childhood Longitudinal Study (ECLS) found that teacher certification had no relationship to student achievement in first grade (Croninger, Rice, Rathbun, & Nishio, 2007), while a study in New York City Public Schools found that while there were small differences in student test scores between teachers with standard, no, and alternative certification (i.e., Teach For America, Teaching Fellows, International teaching programs), there were much larger differences within each certification groups (Kane et al., 2008). Similarly, Goldhaber and Brewer used the National Educational Longitudinal Study of 1988 to look at the relationship between a teacher's certification type and student test scores in 12th grade. They found that teachers who held a standard certification in the subject area taught, compared to those who held private school or no certification, had students who performed better on standardized tests. Yet they also found that teachers with emergency certifications in mathematics and science performed about the same in terms of student test scores as teachers with standard certifications (2000). A later response to the Goldhaber and Brewer article by Darling-Hammond, Berry, and Thoreson suggested that while teachers with emergency certifications may have performed as well as teachers with traditional certifications, both groups had similar qualifications (i.e., education, degree, prior experience) (2001). Thus, while certification wasn't necessarily related to higher student test scores, such findings suggest that there may be other factors related to certification that account for teacher effectiveness.

GPA has been found to have little to no relationship to student test scores. A meta-analysis of 123 studies looked at teacher performance in pre-service programs, as measured by GPA, as well as teacher tests of basic skill, content knowledge, and professional knowledge. They found that GPA had a larger average effect size than any of the teacher tests, but that while GPA had a high correlation with pre-service measures of teacher performance (i.e., college instructors' and mentor teachers' ratings), it had a low correlation with measures of teacher effectiveness while they are teaching, such as student test scores (D'Agostino & Powers, 2009). Similarly, Kane,

Rockoff, and Staiger, in their study of New York City Public Schools, found that GPA had no relationship to student test scores (Kane et al., 2008). Yet, another study of a school district in Kentucky found that undergraduate GPA had a significant relationship to student mathematics achievement in fifth grade (Kukla-Acevedo, 2009).

Research has also found mixed evidence of a relationship between degrees obtained or institutions attended and teacher effectiveness. The 2007 study of the ECLS data found that teachers with an elementary education degree had students with significantly higher reading test scores than teachers with any other type of degree (Croninger et al., 2007). An earlier analysis of data from the “High School and Beyond” study found that the selectivity of a teacher’s university, as ranked by Barron’s ranking system, had a positive effective on high school students’ gain scores (Ehrenberg & Brewer, 1994). Clotfelter, Ladd, and Vigdor’s study of North Carolina teachers and students found that students with teachers who had advanced degrees did not perform any better on math and reading tests. That same study found that the prestige of a teacher’s undergraduate institution, using Barron’s College Admission Selector rankings, did not have any relationship to students’ performance (2010). Finally, a more recent study of elementary and middle school teachers in Florida found that neither having a college major in education, having an advanced degree, nor the selectivity of the teacher’s college were associated with increases in student test scores (Chingos & Peterson, 2011).

Transcripts

Transcripts are generally low cost to gather and provide objective information about academic qualifications in more depth than an application or resume provides. Along with a resume and application, teachers who apply to teach in HCS are already asked to submit transcripts to the Teach Alabama site. In addition to GPA, transcripts can provide information about specific coursework taken, along with the difficulty of and grades received in that coursework. While transcripts do not provide information that is very proximate to classroom practice, specific information about coursework and grades received can be a closer indicator to teaching than GPA alone.

While little relationship has been found between GPA and student test scores, studies have shown that college coursework can have a significant relationship with student outcomes in certain contexts. The 2007 study using ECLS found that, for first grade teachers, there was a significant, positive relationship between those who had more coursework in elementary math and science and higher student test scores in those subjects (Croninger et al., 2007). An analysis of data from the Longitudinal Study of American Youth (LSAY) found that college coursework in the subject areas taught (i.e., mathematics, science) was generally related to higher student test scores in mathematics, but less so in science. More specifically, subject-relevant courses at the high school level were related to higher student test scores, with diminishing returns after the first few courses. Additionally, coursework on pedagogy linked with content was found to have a stronger relationship to student test scores than courses on content alone (Monk, 1994).

Interview

One study that looked at the teacher hiring process in a mid-sized urban school district found that principals relied on in-person interviews the most to make hiring decisions and felt confident about their ability to rate candidates based on interviews (Theel & Tallerico, 2004). Despite their prevalence and the reported comfort in using them, interviews provide information that is more subjective and requires rubrics or training to ensure consistency in rating candidates. Moreover, conducting structured teacher interviews requires high upfront costs to identify the characteristics of interest, develop questions and rubrics to measure those characteristics, and train personnel to ensure inter-rater reliability. While there are commercial teacher interviews that have already been developed, conducting interviews with prospective teachers requires a high level of ongoing costs in terms of personnel involved in scheduling, conducting, and rating interviews. Given the high costs of conducting interviews, it is a better tool to use once an initial round of screening has been completed with a smaller number of candidates.

Despite the costs and subjectivity inherent in interviews, they do provide a large amount of flexibility and can get, to some extent, at teacher characteristics that are more closely related to actual teaching practice. As previously detailed in Chapter 4, HCS is currently using an interview that focuses on various aspects of teaching and classroom practice (i.e., learning environment, integrating technology, innovation, contributions, classroom management, life-long learning, formative assessment and instructional design, parental and community involvement, content knowledge, motivation to improve all learning, and teamwork). As detailed in Chapter 5, there is little relationship between HCS's teacher interview and later measures of teacher effectiveness.

The interview may be the best source to gather information on interpersonal skills; general personal attributes; and attitudes, values, and beliefs. In fact, many of the available commercial teacher interviews address these three areas. These characteristics may be needed for a teachers' success in school, including collaboration with students, parents, and colleagues, in ways that are not directly captured by student performance on standardized tests. Thus, they are worthwhile for districts to consider when making hiring decisions, even when there is limited direct evidence that they improve student achievement. A summary of the literature in those areas is described below.

Research on interviews themselves has not yielded many promising results. Wise et al. (1987) found that prior research on unstructured employment interviews across fields generally had little significant relationship to later job performance. This issue was attributed to the interviewers' lack of ability to ask appropriate questions to determine the match between the candidate and the position, and not necessarily the interview itself. Additionally, while districts have begun to develop structured interview protocols to ensure more systematic selection across interviewers, there is not much evidence to suggest that these interviews yield more effective teachers. A meta-analysis of the Gallup Teacher Perceiver Interview (TPI) indicates that the commercially available teacher interview protocol is related to certain measures of later teacher

effectiveness (e.g., administrator ratings, absenteeism, student ratings), but not others (e.g., observations from individuals outside the school) (Metzger & Wu, 2008; Young & Delli, 2002).

Studies suggest teachers' interpersonal skills and other personal attributes, measured by tests or other non-interview indicators, are related to teacher effectiveness, including student outcomes and evaluated teacher performance. Rockoff et al. (2011) administered a survey to new math teachers in New York City to explore the relationship between non-traditional predictors of teacher effectiveness and a variety of student and teacher outcomes. The predictors included in the survey were: general cognitive ability, content knowledge, personality traits, personal beliefs regarding self-efficacy, certification type, teacher certification exam scores, and selectivity of undergraduate institution. The student and teacher outcomes included: student achievement on standardized math tests, subjective teacher performance ratings, teacher absences, and teacher retention within a school and district. They found that few of the predictors had statistically significant relationships with any student and teacher outcomes, and none of the individual predictors were able to identify large differences in teacher effectiveness. However, when the predictors were combined into cognitive and non-cognitive factors, they found that both had a small statistically significant relationship to student and teacher outcomes (Rockoff et al., 2011). A meta-analysis of 43 studies found a small but significant relationship between teacher self-efficacy and personality and teacher effectiveness (i.e., evaluated teaching performance, student achievement), with a larger relationship between teacher self-efficacy and evaluated teaching performance (Klassen & Tze, 2014).

Many studies, most of which were conducted during the 1980s and 90s, also found that interpersonal and communication skills (i.e., verbal and non-verbal immediacy) have significant relationships with student affect and cognitive learning (Andersen, Norton, & Nussbaum, 1981; Christophel, 1990; den Brok, Brekelmans, & Wubbels, 2004; Rodríguez, Plax, & Kearney, 1996). While most of these studies were conducted in college classes or in countries other than the United States, they suggest that teachers with strong interpersonal skills may positively influence student learning both directly and indirectly through student affect.

Robertson-Kraft and Duckworth explored grit (i.e., "the tendency to sustain perseverance and passion for challenging long-term goals") in a sample of novice teachers teaching in low-income schools through an alternative certification program. Grit was measured by reviewing resumes for continued involvement and advancement in college activities and work experience. They found that grit predicted teachers' staying in the classroom throughout the entire year, as well as teacher effectiveness (i.e., supervisor ratings of teacher effectiveness based on their students' growth on standardized tests) whereas other traditional indicators of teacher quality (e.g., SAT scores, college GPA) did not (Robertson-Kraft & Duckworth). While the sample was specific to novice teachers, with high prior achievement who were teaching in low-income schools, such a study supports the notion that intrapersonal skills may be related to teacher effectiveness.

On the other hand, studies show mixed results in regards to the relationship between attitudes, values, and beliefs and teacher effectiveness. Goe's research synthesis found that

studies focused on teacher characteristics had mixed results. In some cases, there were significant relationships between attitudes, values, and beliefs, while in others there was no significant relationship between the same constructs and student outcomes (2007).

Writing Samples or Prompts

While asking job candidates to submit writing samples or respond to writing prompts does not require extensive administration time for the district, reviewing the responses can be time-consuming for district personnel. As with interviews, reviews of writing samples or prompts are subjective. Such assessments also require some upfront costs to develop the prompts, establish rating criteria, and to train district personnel accordingly. HCS currently has candidates write short responses to two questions prior to the interview process. The writing prompts are described in more detail in Chapter 4. The responses are discussed during the interview, and it is not clear whether or not the responses are used outside of the interview. In other words, it appears that HCS uses the prompts to prepare candidates for the interview rather than as separate artifacts to be reviewed.

A writing prompt could be written to ask a candidate about any teacher characteristic, but the candidate's writing skills can skew the level of detail or accuracy of information about that characteristic. For example, a poorly written response to a prompt about a candidate's experience dealing with a difficult situation may obscure their actual ability to resolve conflict. Alternatively, a writing prompt may provide valid information about a teacher's writing ability, but the relationship of writing ability to teachers' classroom effectiveness is unclear based on the extant literature. While much of what writing prompts could measure is farther removed from classroom practice, they could also measure a teacher's instructional practices by asking for responses to hypothetical classroom scenarios. There is some evidence to suggest that such prompts could be related to teacher effectiveness in certain situations. One study found that structured vignettes that measure teacher use of reform-oriented classroom practices in math and science generally had a weak negative or no relationship with student achievement in the short-term (i.e., 1 year). However, the study found some significant, positive relationships over a longer time period (i.e., 3 years) for open-ended measures of student achievement in math and science, whereas the multiple-choice measures of student achievement still yielded negative or no relationship (Le, Lockwood, Stecher, Hamilton, & Martinez, 2009). Another study found that a classroom video analysis assessment on teaching fractions was related to other measures of teaching knowledge and student learning related to fractions instruction. In the study, elementary and middle school mathematics teachers analyzed and wrote responses to 13 short video clips on fraction instruction. Responses were evaluated using rubrics and those scores were significantly related to items related to fractions taken from the Mathematical Knowledge of Teaching item bank. Thus, in this case, written responses provided a means of measuring teacher content knowledge beyond tests (discussed below)(Kersting, Givvin, Thompson, Santagata, & Stigler, 2012).

References

References can also provide information on any teacher characteristic with little cost to the district (other than the time required for personnel to review them), but the extent to which staff who review the references provide unbiased and accurate information may vary. References can be made less subjective by providing questions that ask for specific information and ask for the reference to rate the candidate numerically on various characteristics, but the information gathered may still be less reliable than gathering it through other sources because the candidate may have incomplete or inaccurate information on the candidate they are rating. The Teach Alabama site asks candidates to submit up to three references, two of which must be professional references, and one of which must be from the candidate's most recent supervisor or principal. The reference form asks the reference to rate the candidate on 18 characteristics (e.g., general appearance, appropriate dress, grooming; exercises professional judgment in the workplace, knowledge of subject matter) on a scale of unknown, unsatisfactory, satisfactory, and strong.

Tests of Teaching Ability and Content Knowledge

Tests of teaching ability and content knowledge provide standardized, objective measures at little cost to the district. The tests are close proximity to classroom practice in that they measure content knowledge or practices teachers would use while in the classroom. More specifically, to receive a teaching certificate in the state of Alabama, teachers must take the Alabama Educator Certification Testing Program which consists of Praxis II exams (Education Testing Service, 2015).

Research suggests that there is mixed evidence, leaning more heavily towards no relationship, that tests of teacher content knowledge and teaching ability are related to other measures of teacher effectiveness. As Wise et al. (1987) note, many standardized teacher examinations have face validity, but have not been validated since the sample only consists of teachers who have been hired by districts. The Measures of Effective Teaching (MET) study found that the Content Knowledge for Teaching (CKT) test was not related to improved student outcomes. The study suggested that this finding did not necessarily reflect the lack of a relationship between content knowledge and student outcomes, but rather the need for further development of the test (MET Project, 2013). Other prior research has found mixed evidence of a relationship between teacher effectiveness and more traditional tests of teaching ability (e.g., National Teacher Examination) and teacher performance (Rice, 2003). In one study, Goldhaber (2007) found a small significant relationship between some teacher certification tests and later teacher effectiveness, but noted that the test cutoffs selected some ineffective teachers and screened out some teachers who would have likely been effective teachers. D'Agostino conducted a meta-analysis of studies that looked at the relationship between undergraduate GPA, teacher tests, and measures of teacher effectiveness (e.g., student test scores, supervisor ratings). They categorized the tests into four categories: basic skills, professional knowledge (pedagogical skills), content knowledge, and a weighted composite of the professional knowledge and content

knowledge portions of the National Teacher Exam (NTE WCET). They found that while none of the other types of tests was a significantly better predictor of teacher effectiveness than the NTE WCET, undergraduate GPA was (2009).

Other researchers, on the other hand, have found significant, positive relationships between mathematical content and student outcomes. Hill, Rowan, and Ball found that a measure of teacher mathematical knowledge they developed, in concert with data from ECLS, was significantly related to increased student outcomes for first and third graders (2005). Similarly, Campbell et al. (2014) found that measures of mathematics content knowledge and pedagogical content knowledge were significantly related to positive student outcomes for 4th through 8th grade teachers in Delaware, Maryland, and Pennsylvania.

More recent research indicates that performance assessments of teacher practice used with beginning teachers, such as the Performance Assessment for California Teachers (PACT), are related to later teacher effectiveness (L. Darling-Hammond, Newton, & Wei, 2013). Ultimately, while there is little evidence of the relationship between tests of teacher knowledge and ability on student outcomes, there is evidence that tests that are tailored to a content area, as well as performance assessments, may be stronger predictors of student outcomes than those that are more general.

Instruments HCS Is Not Currently Using

Sample Lessons

Having candidates teach a sample lesson is costly in that it requires district personnel to be available to observe the lesson. If the sample lesson is taught in a classroom with students, that may yield a more realistic picture of a teacher's practice, but is also more disruptive and has more constraints to schedule during the course of the school day. On the other hand, if a sample lesson is taught in a room without students, that is less disruptive, but also less indicative of a teacher's teaching ability in an authentic situation. Candidates could also submit a recorded sample lesson, but a video provides a narrow view of the classroom and may be less valid because the teacher selected the sample to submit. Furthermore, because sample lessons provide subjective data, the district may require upfront costs to develop protocols and train raters. Despite being costly, sample lessons have a high proximity to actual classroom practice and are therefore better-suited for use toward the end of the selection process, once there are smaller numbers of candidates.

Sample lessons predominately provide information about a teacher's teaching ability, but can also yield insights into other teacher characteristics. There is little evidence that specific teaching practices are positively related to student outcomes, but more evidence that observation protocols that measure a number of teacher practices are positively related to student outcomes. Goe found that while many studies found that teacher practices were related to student outcomes, the results

were frequently not practically or statistically significant and the study methodologies were flawed. In some cases, the studies relied on student report or teacher self-report of instructional practice. In many cases, the positive findings were significant only with a small subset of the groups included in the studies (Goe, 2007).

While research has been done on specific teaching practices, determining their relationship to student outcomes is often difficult because they are used in concert with other practices and in a classroom environment with many factors that are difficult to control for. Other studies have found that teacher strategies that promote student interaction and discussion in mathematics (Webb et al., 2014), inquiry-based learning in science (Furtak, Seidel, Iverson, & Briggs, 2012), and use of instructional technologies (A. C. K. Cheung & Slavin, 2012; A. C. Cheung & Slavin, 2013; Hew & Cheung, 2013) all have positive relationships to student outcomes. Additionally, as noted in Chapter 2, much of the research on observation protocols that measure current teacher practices have found positive relationships to student test scores (Ing & Webb, 2012; Kane & Staiger, 2012; Little et al., 2009).

Portfolio/Teaching Artifacts

Teaching portfolios can showcase a wide variety of teacher characteristics, including teaching practice, content knowledge, beliefs and attitudes about teaching, work in activities outside of the classroom, and the candidate's fit with the school culture. While teaching portfolios don't require extensive data collection time, they do require time for review and rating by district personnel. Establishing submission and assessment criteria also requires upfront time. Still, many of the elements that can be included in a portfolio are proximate to classroom practice, including lesson plans, assignments, and student work, so there may be merit to considering these artifacts.

While Chapter 2 discusses the use of portfolios for teacher evaluation of current teachers, here the focus is on portfolios for new teachers. In many cases, these portfolios in the relevant studies were developed as part of teacher education and preparation programs. Research has shown that portfolios created in teacher education programs can serve as a tool to help teachers reflect on their own practice (Borko, Michalec, Timmons, & Siddle, 1997). However, another study found that the type of reflection included in teacher portfolios was predominately recalling and evaluating activities, and much less so analyzing, critically processing, diagnosing, or reflecting on those events (Mansvelder-Longayroux, Beijaard, & Verloop, 2007). What is less clear from existing literature is whether the information hiring committees glean from teaching portfolios helps predict the quality of teachers' on-the-job performance. This is an area in which more research is needed.

Conclusion

Ultimately, there is no one best tool when selecting teachers. As the findings throughout this chapter suggest, there is not one selection tool or set of teacher characteristics that consistently yield the best outcomes. Each comes with limitations and advantages that must be considered in concert with the needs of the district. Yet, the majority of these studies looked at student test scores as the measure of teacher effectiveness. The research on interpersonal skills was largely the exception, where studies included a broader range of outcomes, including student affect and evaluations of teacher performance (e.g., from principals). Thus, while the evidence that many of these characteristics have little relationship to student test scores is limited, broadening the scope of outcomes may yield more positive results. To that end, districts should consider other potential outcomes of interest, in addition to the body of current literature, while creating their hiring processes.

This chapter does provide some insights into how to choose the screening mechanisms and characteristics that may lead to hiring better teachers, and result in lower costs to the districts. First, written documentation that candidates submit and require little time for districts to review can be used to ensure teachers meet minimum requirements. Among the characteristics that can be gathered from written documentation, years of experience and relevant coursework have a greater connection to positive student outcomes. While years of experience is easy to gather from an application or resume and has consistently shown a positive relationship to teacher effectiveness in the first three to five years, that characteristic is generally not relevant to hiring new teachers, many of whom are novice. Coursework that is relevant to the subject taught or pedagogy, on the other hand, have been shown to have a relationship with positive student outcomes and can be gathered through reviewing transcripts.

Certification, GPA, and teacher test scores, while all require little effort on the part of the district to collect, should be considered carefully. While candidates must meet certain minimum certification requirements, few studies found positive relationships between certification type and student outcomes, and multiple studies noted that there was more variation within certification types than across certification types. Additionally, while specific content-related tests and performance assessments are less widely used, but appear to have stronger relationships with student outcomes, many of the widely used general teacher tests do not have strong relationships. Thus, while districts can easily gather this information on candidates, given its limited connection with teacher effectiveness, this information should be taken with a grain of salt and in consideration with other teacher characteristics and screening measures.

Once the pool of candidates is narrowed down, more costly and time-consuming methods that provide additional information about potential effectiveness can then be used. For example, interviews are a good source of information for interpersonal skills, which have been shown to have a relationship to later teacher effectiveness. Again, interpersonal skills, while having some

relationship to improved student test scores, also have positive relationships with other outcomes, such as student affect and teacher evaluation ratings. While observations of sample lessons are costly for the district to administer, there is evidence that they can yield information that is useful to the district. Given their cost, it is important to identify what teaching strategies or other teacher characteristics are most important for the district to measure. Given the evidence that specific observation protocols, as well as specific teaching strategies, are related to positive student outcomes, a district could choose to use a protocol wholesale or include specific strategies in its observations. Otherwise, the district could measure additional strategies that are compatible with its pedagogical philosophy but might not have a strong evidence-base, or some combination of the two. Ultimately, the cost of creating an observation protocol and conducting observations will require closer scrutiny to ensure that the district is using its resources to learn the most about candidates.

As noted in Chapter 2, the match between a teacher and the school or district they will be hired in is very important. Consequently, not all selection processes will be the same; each will depend on district resources and preferences. While this chapter is intended to be a tool to aid in that decision-making, the next chapter will focus on specific recommendations for improving the teacher selection and hiring process in HCS.

7. Conclusion and Recommendations

As was made clear throughout this study, there is not one answer or one right way to structure the process of hiring teachers. There is not one characteristic or measure that accurately predicts which candidates will be effective teachers. Past research on teacher hiring and teacher effectiveness has shown this, and the quantitative data analysis presented did not offer conclusive results. Even if there was more conclusive evidence of what makes effective teachers, there is still variation in what districts and schools are looking for in teachers, and what they consider effective teachers. While there is not one right answer, there are a number of ways that HCS can improve its hiring, evaluation, and data collection processes in order to better understand the teachers it is hiring. This chapter provides recommendations in each of those three areas. While these recommendations are specific to HCS, they are also relevant and broadly applicable to other districts looking to improve their hiring and evaluation processes.

Improvements to HCS Teacher Hiring Process

Tailor rubrics to each interview question to improve the quality of your measures

While the vast majority of interview questions are rated as 3s or 4s, the first four interview questions (i.e., learning environment, data-driven instruction, motivation to improve learning for all, team work) have slightly more varied score distributions. Yet, even these questions are heavily rated as 3s and 4s. Currently, all questions are rated on the same 1-4 scale of the extent of knowledge of an area, which is not appropriate for all questions and appears to make raters avoid using the rubric. Each interview question should have a separate rubric to improve rater consistency and increase the distribution of scores. Using language in the rubrics that is specific to each question, as opposed to vague descriptors (e.g., minimal evidence, evidence, evidence of extensive knowledge) can increase the variability in scores by more clearly differentiating between categories (Jonsson & Svingby, 2007).

For example, one interview question is “Tell me about the first experience in your life when you realized that you had the power of change or power to do something meaningful.” Interviewers are given the following guidance of rating: 1) There is no evidence of knowledge or skill set; 2) There is minimal evidence of knowledge or skill set; 3) There is evidence of a good level of knowledge or skill set; and 4) There is evidence of extensive knowledge or skill set. Based on the 2013-2014 “look-fors” provided, this question is meant to focus on the skill set of “reflection,” but that is not reflected in the rubric. Thus, more appropriate rating guidance might include:

- 1) The candidate does not provide an example, or the example does not exemplify the power of change or the power to do something meaningful;
- 2) The candidate provides an example, but does not reflect on how that example exemplifies the power of change or the power to do something meaningful;
- 3) The candidate provides an example and reflects on how it exemplifies the power of change or the power to do something meaningful to a limited extent;
- 4) The candidate provides an example and reflects on how it exemplifies the power of change or the power to do something meaningful to a great extent.

While the example is provided as an illustration, HCS should develop the rubrics to help clarify what they want to learn from each question, and provide benchmarks to aid in consistency across raters. As noted in Chapter 4, HCS added specific “look-fors” in 2013-2014 and the constructs it was trying to measure for each interview question in 2014-2015. They can add on this work by developing exemplar responses for each question in each rating category (e.g., what does a 4 look like? What does a 1 look like?). For example, a level 2 response might be: *I first learned about the power of change when I was tutoring a group of fourth graders after school. It was pretty powerful.* Developing exemplars will help ensure that the interview questions and rubrics accurately align with the constructs of interest and can differentiate between a weak and strong response. It will also help raters have a better understanding of what to look for in each question and how responses should be rated.

Continue to Experiment: Field and test questions to see how they work

In the coming years, as HCS improves both its hiring and review processes, new definitions of good teaching and important teacher characteristics will emerge. As they emerge, the district can pilot and test new interview questions to measure these characteristics. They can also modify existing questions so that they better reflect characteristics of interest. In both cases, the district should field the questions with candidates and track the data on those questions. The district should look at the distribution of candidates’ scores on each question to determine if they are getting enough variation. Additionally, the district can run analyses similar to the ones mentioned in Chapter 5 to determine if there is a significant relationship between the new questions and later probationary review scores.

Identify questions and characteristics of interest to ensure most important teacher qualities are screened

The evidence is inconclusive regarding which screening measures predict the most effective teachers, and the factor analysis suggests all nine items included in the factor analysis form a

single construct. Consequently, HCS should explore other methods of identifying which characteristics to screen for and which interview questions to include.

As reported in Chapter 5, all items included in the factor analysis correlate with each other and appear to measure the same underlying construct. While this was beyond the scope of this study, further analyses can explore the effect that dropping each of the items has on overall reliability. If all items contribute similar levels of information, then the district should select which items to keep based on what they deem important.

In order to ensure that the most important teacher qualities are being screened for, HCS should identify the characteristics of teachers that they deem non-negotiable (e.g., certification, willingness to work in any school) and screen for those, recognizing that the schools will then select from those teachers, allowing more flexibility and discretion in identifying characteristics that are important to each school. Since little is predictive of more effective teachers, it is better to focus on minimum qualifications and basic fit with the school and district instead of which teachers might be most effective. Then, once candidates are selected and hired, focus on identifying the more effective teachers during the probationary review process. In other words, shift the screening filter and have a higher rate of rejection based on more predictive data. National Board for Professional Teaching Standards (NBTS) and other researchers have suggested that while pre-screening measures do little to differentiate more effective teachers, there are measures that can be used after hiring that do so.

Improvements to HCS Teacher Evaluation Process

Any changes in the probationary teacher review should be made while considering the overall purpose of the evaluation process. If the purpose of the process is to identify and motivate good teachers, then the focus should be on codifying what good teachers look like in practice: Define the demonstrated behaviors and achieved outcomes of key importance. If the purpose of the process is to identify and weed out poor teachers, then the focus should be on defining undesirable behaviors and outcomes so that poor performing teachers can be removed in their first years of teaching. More likely, the evaluation will serve both purposes to some extent, and both should be considered in light of the recommendations below.

Improve scoring for probationary teacher review components to increase likelihood of finding significant relationships with hiring process and differentiating between more and less effective teachers

The majority of teachers fall into the two highest categories for each probationary review component under the current rubrics. Improving the scoring rubrics for review components may increase the likelihood of finding significant relationships without ceiling effects and limited variation. The probationary review rubric for SGP scores is currently designed so teachers who score above 50 receive 15 out of 15 points, teachers who score between 35 and 50 receive 10 out of 15 points, etc. Creating more than four possible categories (e.g., 0, 5, 10, or 15 out of 15) can increase the variability in scores. Raising the cutoffs for each category can also limit the number of teachers who receive the highest scores and change the overall distribution of scores. As previously noted, the purpose of the evaluation should be considered when determining the score cutoffs. Making it more difficult for teachers to receive the highest score will have implications for how teachers perceive the evaluation system and how, if at all, they adjust their teaching practice (L. Darling-Hammond, 2013; Liang, 2013).

The rubric for half of the review components are based on objective indicators (e.g., number of teacher absences, SGP score) and adjusting the scoring thresholds is relatively straightforward. However, the rubrics for the observation, principal rating, and student survey review components are based on more subjective measures (e.g., principal's determination of whether or not teacher exhibits professionalism). Therefore, in addition to changing the thresholds for scoring, the actual observation and principal rating form and student survey should be reviewed. More specifically, the scale for each should be considered (e.g., yes/no, 3-point scale, 1-5 Likert scale, etc.) to balance between the ease of rating and the ability to get variation.

Track and record actual scores on probationary review components to provide additional data for future evaluations and better differentiate between more and less effective teachers

While the rubric scores that teachers receive on the probationary review components (e.g., 15 out of 15 points) are currently tracked systematically, the actual scores of each component (e.g., SGP score of 53) are not. Tracking the actual scores of the review components will provide more information when conducting future analyses. While there was not a relationship between the interview score and rubric scores for individual review components, there could be a relationship between interview scores and the actual review component scores. At the very least, the actual review component scores provide more variation in outcomes since the scores are not reduced to a 3- or 4-point scale.

Additionally, tracking the actual review component scores will allow for testing different cutoff scores for the rubrics. While these cutoffs already changed across school years, tracking

teachers' actual scores will allow the district to conduct additional analyses to see if changing the thresholds will yield different results. Such analyses might look at the distribution of scores; how does raising each cutoff SGP score (e.g., the cutoff for the highest category moves from 50 to 60) change the distribution of teachers who receive each rubric score? An analysis might also look at the relationship between the interview score and the different rubric scores; does changing the cutoff for each category make it so that there is a significant relationship between the new scores and interview scores?

Consider the weighting of the elements of the probationary review process to ensure that the most important teacher outcomes are being measured

Identifying the qualities the district would like to screen for during the hiring process will reveal which characteristics it is looking for in teachers. At the same time, the district should consider those teacher outcomes that are important and ensure that those elements are included in the probationary review process. Are these components (i.e., principal rating, observation, student survey, SGP score, attendance, disciplinary referrals) the most important elements in differentiating between strong and weak teachers? Are there other elements that are not captured?

Moreover, does the weight that each component has in the overall review score reflect the importance of that component? Since the observation and principal rating are worth 20 percent each, and the student survey is worth 27 percent, each of these components has the ability to mean the difference between a teacher passing or failing the probationary review process. The district should therefore decide whether or not any one element is important enough to weigh heavily enough to change the outcome of the review. The district did change the weighting assigned to each review component over the 2012-2013 and 2013-2014 school years (as noted in Chapter 4). As the district considers additional changes, analyses can be run to determine how changing the weightings of individual components affects teachers' overall review scores.

Find alternative measures for teachers who do not receive all review components

As the district considers the weighting of each review component, it is also important to consider that not all teachers receive the student survey, SGP scores, and/or observation scores, even though those components make up 63 percent of a teacher's score. Only mathematics and reading teachers receive SGP scores. Not all teachers in their second year receive observations. Only teachers in third grade and above receive student survey scores. While there are reasons why teachers do not receive scores on these elements, such a system limits the data available on many teachers' reviews and may skew the scores teachers receive. This is especially true in the

case where the inclusion of review components affects teachers' overall scores. As noted in Chapter 5, in some analyses teachers with observation scores did significantly better than teachers without those scores on the overall probationary review, while teachers with student survey scores did significantly worse than teachers without those scores.

The district should therefore find alternative, but comparable measures for teachers who do not receive all review components. For example, teachers below third grade could administer a modified survey that is read aloud to students or that uses a simplified rating system and alternative tests could be used that applies to non-math or ELA teachers.

Use teacher evaluation data to inform and improve the hiring process

As the probationary review process changes, the new data can help further inform the interview process. Running similar analyses to those in Chapter 5 using new data (e.g., actual review component scores, rubric scores with new thresholds, different weightings of review components) may provide more insights into the extent to which the interview predicts future teacher performance. Additionally, as the review process and definitions of good teaching evolve, the district should ensure that the interview and hiring process continues to evolve to reflect those definitions.

Suggestions for Future Data Collection in HCS

Better understand teacher attrition: Improved tracking of teacher retention and conducting good exit interviews can help identify methods to retain top teachers

Consider teacher retention and keeping effective teachers as an outcome to track. While it is important to find candidates who will be effective teachers, it is equally important to keep those teachers over time. This study looked at teachers who received a probationary review at the end of their second year as an indicator of retention. Although it was clear why teachers who received probationary review scores below 65 were not retained, it is not clear why the other teachers did not remain in the district through the end of their second year, or when they left. Conducting exit interviews with teachers who are leaving will help the district understand why they are leaving. While teachers might leave the district for reasons the district cannot control (e.g., relocation due to a significant other's job, having a child), there are also a number of reasons that the district could address (e.g., pay, opportunities for professional development, school-level factors). Having this information will give the district an opportunity to address any systematic issues that are causing teachers, particularly effective ones, to leave.

Track data on all candidates to provide context into hiring process and teacher supply

To the extent possible, track data (e.g., GPA, college, interview score, certification, demographics) on all candidates, including those who were not hired. Systematically tracking a wider range of data will allow for a more diverse set of analyses on the relationship between candidate characteristics and teacher effectiveness. For example, regressions could be run that look at the relationship between GPA or certification status and later teacher effectiveness. While this study focused on teacher interview scores, it is possible that looking at the interview scores in concert with these other characteristics will yield significant relationships with later teacher effectiveness. Looking at the relative performance of teachers from different universities can also help the district understand if certain universities supply better teachers. Additionally, analyzing the data in regards to the subjects and schools at which teachers teach can help the district understand if the hiring process is better at identifying effective teachers in certain grades, subject areas, or schools. While the first two years of data may not include enough candidates to be able to do these more specific analyses, tracking this data over time will allow for more extensive analyses.

Additionally, while the district will not have data on teacher effectiveness for those who were not hired, the data can provide more context to inform the hiring process: it can tell the district the extent to which the candidates they hired are similar to those that were interviewed but not hired and those that were not interviewed or hired. While this will not provide definitive answers about which candidates will be the most effective, it will tell the district if candidates that are being screened out of the process look the same (in terms of GPA, college, interview score, certification, demographics) as those teachers who were hired and are performing well (or poorly). Such information can provide insights into the relationship between candidate characteristics and interview scores. It can also help the district understand if they are systematically eliminating specific types of candidates (e.g., those with specific types of certification, people from certain universities) from the hiring process.

Final Thoughts

This research has shown that the current teacher hiring process in HCS is not successfully identifying the most effective teachers. The recommendations above suggest that reframing the interview process and collecting data on that process will find and retain more effective teachers. I have reframed the process map presented in Figure 4.1 so that it focuses on using data in the context of both the hiring and probationary review processes to ultimately inform and improve those processes and yield better teachers.

Previously, the hiring process ended when teachers were hired by HCS. In the revised version, the probationary review process is a continuation of the hiring process in that teachers

who meet HCS's minimum qualifications are hired by the district and then evaluated through the probationary review to determine their effectiveness. Furthermore, the revised process map indicates that there is a feedback mechanism between the probationary review process and the interview, recruiting, and principal screening processes in which the data on which teachers are effective informs which teachers are recruited, interviewed, and hired. The new steps in the process and feedback mechanisms are indicated in Figure 7.1 by green shaded boxes.

Moreover, the hiring process previously focused on data collected during the interview process. However, there are additional areas of data collection that could be added to the process. First, recruitment (the green box in Figure 7.1) is also a part of the hiring process that should be tracked and informed by data from the review process. Second, the district could also collect data on all applicants and on how principals are screening teachers (indicated in Figure 7.1 by the striped and dotted boxes).

Figure 7.1 Revised process map for HCS hiring process

Researchers, policymakers, and practitioners are looking at a number of policies and practices that have the potential to improve student outcomes, especially in schools where students are struggling academically. Of these various reforms, teachers and teaching practices are consistently identified as areas that can significantly help improve student outcomes. Consequently, making decisions regarding which teachers should be hired is an important step on the road to improving student outcomes. This study provides a window into the hiring process in one school district, Huntsville City Schools, with a diverse student population, a vested interest in identifying effective teachers, and access to rich data on student and teacher performance to help inform those decisions.

While there is not conclusive evidence about what attributes HCS, or any other school district, should screen prospective teachers for, these findings highlight the importance of continuing to collect data on, and monitor the effects of changes in hiring policies and practices. As districts continue to accumulate data on these processes, they will not only get an overall picture of the effectiveness of teachers in their districts, but they can also learn more about the distribution of those teachers across schools. Thus, monitoring the hiring process over time can go beyond improving the overall teacher pool and student outcomes, and begin to target the schools and students that need the most support.

References

- Akiba, Motoko, LeTendre, Gerald K, & Scribner, Jay P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher*, 36(7), 369-387.
- Allen, Michael B. (2005). Eight questions on teacher recruitment and retention: What does the research say? *Education Commission of the States (NJ3)*.
- Andersen, Janis F, Norton, Robert W, & Nussbaum, Jon F. (1981). Three investigations exploring relationships between perceived teacher communication behaviors and student learning. *Communication Education*, 30(4), 377-392.
- Baker, Eva L, Barton, Paul E, Darling-Hammond, Linda, Haertel, Edward, Ladd, Helen F, Linn, Robert L, . . . Shepard, Lorrie A. (2010). Problems with the use of student test scores to evaluate teachers. Epi briefing paper# 278. *Economic Policy Institute*.
- Balter, Dana, & Duncombe, William. (2005). Teacher hiring practices in new york state school districts.”. *Report Prepared For The Education Finance Research Consortium, January*.
- Balter, Dana, & Duncombe, William. (2006). *Staffing classrooms: Do teacher hiring practices affect teacher qualifications?* Paper presented at the Annual meeting of the American Education Finance Association, Denver, Co.
- Bland, J Martin, & Altman, Douglas G. (1995). Multiple significance tests: The bonferroni method. *Bmj*, 310(6973), 170.
- Borko, Hilda, Michalec, Paul, Timmons, Maria, & Siddle, Jean. (1997). Student teaching portfolios: A tool for promoting reflective practice. *Journal of teacher education*, 48(5), 345.
- Bryan, Julia, Day - Vines, Norma L, Griffin, Dana, & Moore - Thomas, Cheryl. (2012). The disproportionality dilemma: Patterns of teacher referrals to school counselors for disruptive behavior. *Journal of Counseling and Development*, 90(2), 177-190.
- Burniske, Jackie, & Meibaum, Debra. (2011). The use of student perceptual data as a measure of teaching effectiveness. *Texas Comprehensive Center*.
- Campbell, Patricia F, Nishio, Masako, Smith, Toni M, Clark, Lawrence M, Conant, Darcy L, Rust, Amber H, . . . Choi, Youyoung. (2014). The relationship between teachers'

- mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, 45(4), 419-459.
- Castellano, Katherine E, & Ho, Andrew D. (2013). A practitioner's guide to growth models. *Council of Chief State School Officers*. Retrieved from http://scholar.harvard.edu/files/andrewho/files/a_practitioners_guide_to_growth_models.pdf. Domaleski & Perie. *Promoting equity*.
- Cheung, Alan C. K., & Slavin, Robert E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3), 198-215. doi:<http://dx.doi.org/10.1016/j.edurev.2012.05.002>
- Cheung, Alan CK, & Slavin, Robert E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational Research Review*, 9, 88-113.
- Chingos, Matthew M, & Peterson, Paul E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30(3), 449-465.
- Christophel, Diane M. (1990). The relationships among teacher immediacy behaviors, student motivation, and learning. *Communication Education*, 39(4), 323-340.
- Clotfelter, Charles T, Ladd, Helen F, & Vigdor, Jacob L. (2007). *Are teacher absences worth worrying about in the us?* Retrieved from
- Clotfelter, Charles T, Ladd, Helen F, & Vigdor, Jacob L. (2010). Teacher credentials and student achievement in high school a cross-subject analysis with student fixed effects. *Journal of Human Resources*, 45(3), 655-681.
- Croninger, Robert G, Rice, Jennifer King, Rathbun, Amy, & Nishio, Masako. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26(3), 312-324.
- D'Agostino, Jerome V, & Powers, Sonya J. (2009). Predicting teacher performance with test scores and grade point average: A meta-analysis. *American Educational Research Journal*, 46(1), 146-182.
- Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., and Vasquez Heilig, J. (2005). Does teacher preparation matter? Evidence about teacher certification, teach for america, and teacher effectiveness. *Education Policy Analysis Archives*, 13(42), 1-51.

- Darling-Hammond, Linda. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. Teachers College Press.
- Darling-Hammond, Linda, Berry, Barnett, & Thoreson, Amy. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23(1), 57-77.
- Darling-Hammond, Linda, Newton, Stephen P, & Wei, Ruth Chung. (2013). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 25(3), 179-204.
- den Brok, Perry, Brekelmans, Mieke, & Wubbels, Theo. (2004). Interpersonal teacher behaviour and student outcomes. *School effectiveness and school improvement*, 15(3-4), 407-442.
- Education Testing Service. (2015). Overview of alabama testing requirements. Retrieved from <http://www.ets.org/praxis/al>
- Ehrenberg, Ronald G., & Brewer, Dominic J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review*, 13(1), 1-17. doi:[http://dx.doi.org/10.1016/0272-7757\(94\)90019-1](http://dx.doi.org/10.1016/0272-7757(94)90019-1)
- Ferguson, Ronald F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, 94(3), 24-28.
- Furtak, Erin Marie, Seidel, Tina, Iverson, Heidi, & Briggs, Derek C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching a meta-analysis. *Review of Educational Research*, 82(3), 300-329.
- Gardner, David P. (1983). *A nation at risk*. Retrieved from Washington, D. C.:
- Glazerman, S., Mayer, D., Decker, P. (2006). Alternative routes to teaching: The impacts of teach for america on student achievement and other outcomes. *Journal of Policy Analysis and Management*, 25(1), 21.
- Goe, Laura. (2007). The link between teacher quality and student outcomes: A research synthesis. *National Comprehensive Center for Teacher Quality*.
- Goldhaber, Dan. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources*, 42(4), 765-794.
- Goldhaber, Dan D, & Brewer, Dominic J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.

- Goldhaber, Dan, Gabele, Brian, & Walch, Joe. (2012). *Does the model matter? Exploring the relationship between different achievement-based teacher assessments*. Retrieved from Seattle, WA:
- Government Accountability Office. (2007). *Education should clarify guidance and address potential compliance issues for schools in corrective action and restructuring status*. Retrieved from
- Greenberg, Julie, McKee, Arthur, & Walsh, Kate. (2014). 2014 teacher prep review: A review of the nation's teacher preparation programs. *National Council on Teacher Quality*.
- Greenwald, Anthony G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182.
- Haberman, Martin. (1995). Selecting star teachers for children and youth in urban poverty. *Phi Delta Kappan*, 76, 777-777.
- Hanushek, Eric A, Peterson, Paul E, & Woessmann, Ludger. (2011). Us math performance in global perspective.
- Hanushek, Eric A, Peterson, Paul E, & Woessmann, Ludger. (2012). Achievement growth: International and us state trends in student performance. PEPG report no.: 12-03. *Program On Education Policy And Governance, Harvard University*.
- Harris, Douglas N., & Sass, Tim R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7-8), 798-812. doi:<http://dx.doi.org/10.1016/j.jpubeco.2010.11.009>
- Hew, Khe Foon, & Cheung, Wing Sum. (2013). Use of web 2.0 technologies in k-12 and higher education: The search for evidence-based practice. *Educational Research Review*, 9(0), 47-64. doi:<http://dx.doi.org/10.1016/j.edurev.2012.08.001>
- Hill, Heather C, Rowan, Brian, & Ball, Deborah Loewenberg. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.
- Ing, Marsha, & Webb, Noreen M. (2012). Characterizing mathematics classroom practice: Impact of observation and coding choices. *Educational Measurement: Issues And Practice*, 31(1), 14-26. doi:10.1111/j.1745-3992.2011.00224.x
- Irvin, Larry K, Tobin, Tary J, Sprague, Jeffrey R, Sugai, George, & Vincent, Claudia G. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status

- and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions*, 6(3), 131-147.
- Jacob, Brian A, & Lefgren, Lars. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal Of Labor Economics*, 26(1), 101-136.
- Jonsson, Anders, & Svingby, Gunilla. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Kane, Thomas J, Rockoff, Jonah E, & Staiger, Douglas O. (2008). What does certification tell us about teacher effectiveness? Evidence from new york city. *Economics of Education Review*, 27(6), 615-631.
- Kane, Thomas J, & Staiger, Douglas O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper. Met project. *Bill & Melinda Gates Foundation*.
- Kane, Thomas J, Taylor, Eric S, Tyler, John H, & Wooten, Amy L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587-613.
- Kersten, Thomas. (2008). *Teacher hiring practices: Illinois principals' perspectives*. Paper presented at the The Educational Forum.
- Kersting, Nicole B, Givvin, Karen B, Thompson, Belinda J, Santagata, Rossella, & Stigler, James W. (2012). Measuring usable knowledge teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568-589.
- Klassen, Robert M, & Tze, Virginia MC. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59-76.
- Koedel, Cory, Parsons, Eric, Podgursky, Michael, & Ehlert, Mark. (2012). Teacher preparation programs and teacher quality: Are there real differences across programs? *National Center For Analysis For Longitudinal Data In Education Research, Working Paper*, 79.
- Kukla-Acevedo, Sharon. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review*, 28(1), 49-57. doi:<http://dx.doi.org/10.1016/j.econedurev.2007.10.007>

- Lackzo-Kerr, I., Berliner, D. . (2002). The effectiveness of "teach for america" and other under-certified teachers. *Education Policy Analysis Archives*. Retrieved from Available at: <<http://epaa.asu.edu/ojs/article/view/316%3E>.
- Ladson-Billings, Gloria. (2006). From the achievement gap to the education debt: Understanding achievement in us schools. *Educational Researcher*, 35(7), 3-12.
- Le, Vi-Nhuan, Lockwood, JR, Stecher, Brian M, Hamilton, Laura S, & Martinez, Jose Felipe. (2009). A longitudinal investigation of the relationship between teachers' self-reports of reform-oriented instruction and mathematics and science achievement. *Educational Evaluation and Policy Analysis*, 31(3), 200-220.
- Liang, Guodong. (2013). Teacher evaluation policies in the united states: Implementation and impact on constructivist instruction. *Teacher reforms around the world: Implementations and outcomes*, 179-206.
- Little, Olivia, Goe, Laura, & Bell, Courtney. (2009). A practical guide to evaluating teacher effectiveness. *National Comprehensive Center for Teacher Quality*.
- Liu, Edward. (2003). *New teachers' experiences of hiring: Preliminary findings from a four-state study*. Paper presented at the A paper prepared for the annual meeting of the American Research Association, Chicago.
- Liu, Edward. (2005). Hiring, job satisfaction, and the fit between new teachers and their schools. *American Educational Research Association. Montreal, Quebec*.
- Liu, Edward, & Johnson, Susan Moore. (2006). New teachers' experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly*, 42(3), 324-360.
- Loeb, Susanna, Kalogrides, Demetra, & Bételle, Tara. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education*, 7(3), 269-304.
- Mansvelder-Longayroux, Désirée D, Beijaard, Douwe, & Verloop, Nico. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and teacher education*, 23(1), 47-62.
- Mason, Richard W, & Schroeder, Mark P. (2010). Principal hiring practices: Toward a reduction of uncertainty. *The Clearing House*, 83(5), 186-193.
- Measuring the mathematical quality of instruction. (2011). *Journal Of Mathematics Teacher Education*, 14(1), 25-47. doi:10.1007/s10857-010-9140-1

- MET Project. (2010). *Asking students about teaching: Student perception surveys and their implementation*. Retrieved from Seattle, WA:
- MET Project. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the met project's three-year study*. Retrieved from Seattle, WA:
- Metzger, Scott Alan, & Wu, Meng-Jia. (2008). Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Educational Research*, 78(4), 921-940.
- Miller, Raegen T, Murnane, Richard J, & Willett, John B. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. *Educational Evaluation and Policy Analysis*, 30(2), 181-200.
- Monk, David H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125-145.
- Myung, Jeannie, Martinez, Krissa, & Nordstrom, Lee. (2013). *A human capital framework for a stronger teacher workforce*. Retrieved from Stanford, CA:
- Naper, Linn Renée. (2010). Teacher hiring practices and educational efficiency. *Economics of Education Review*, 29(4), 658-668.
- National Center for Education Statistics. (2012). *Schools and staffing survey* (Institute of Education Sciences, Trans.). Washington, D.C.
- Nye, Barbara, Konstantopoulos, Spyros, & Hedges, Larry V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237-257.
- Office of Civil Rights. (2012). *Helping to ensure equal access to education: A report to the president and secretary of education*. Retrieved from Washington, DC:
- Peske, Heather, & Haycock, Kati. (2006). *Teaching inequality*. Retrieved from Washington, D.C.:
- Putnam, Robert F, Luiselli, James K, Handler, Marcie W, & Jefferson, Gretchen L. (2003). Evaluating student discipline practices in a public school through behavioral assessment of office referrals. *Behavior modification*, 27(4), 505-523.
- Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for america: An evaluation of teacher differences and student outcomes in houston, texas*. Retrieved from Palo Alto:

- Rice, Jennifer King. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*: ERIC.
- Robertson-Kraft, Claire, & Duckworth, Angela Lee. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record*, 116(3).
- Rockoff, Jonah E, Jacob, Brian A, Kane, Thomas J, & Staiger, Douglas O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, 6(1), 43-74.
- Rodríguez, José I, Plax, Timothy G, & Kearney, Patricia. (1996). Clarifying the relationship between teacher nonverbal immediacy and student cognitive learning: Affective learning as the central causal mediator. *Communication Education*, 45(4), 293-305.
- Rutledge, Stacey A, Harris, Douglas N, Thompson, Cynthia T, & Ingle, W Kyle. (2008). Certify, blink, hire: An examination of the process and tools of teacher screening and selection. *Leadership And Policy In Schools*, 7(3), 237-263.
- Safran, Stephen P, & Oswald, Karen. (2003). Positive behavior supports: Can schools reshape disciplinary practices? *Exceptional children*, 69(3), 361-373.
- Skiba, Russell J, Chung, Choong-Geun, Trachok, Megan, Baker, Timberly L, Sheya, Adam, & Hughes, Robin L. (2014). Parsing disciplinary disproportionality contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, 51(4), 640-670.
- Steele, Jennifer L, Baird, Matthew, Engberg, John, & Hunter, Gerald. (2014). Trends in the distribution of teacher effectiveness in the intensive partnerships for effective teaching.
- Stinebrickner, Todd R. (2001). Compensation policies and teacher decisions. *International Economic Review*, 42(3), 751-780. doi:10.1111/1468-2354.00132
- Theel, Ronald K., & Tallerico, Marilyn. (2004). Using portfolios for teacher hiring: Insights from school principals. *Action in Teacher Education*, 26(1), 26-33. doi:10.1080/01626620.2004.10463310
- Vincent, Claudia G, Tobin, Tary J, Hawken, Leanne S, & Frank, Jennifer L. (2012). Discipline referrals and access to secondary level support in elementary and middle schools: Patterns across african-american, hispanic-american, and white students. *Education and Treatment of Children*, 35(3), 431-458.
- Webb, Noreen M, Franke, Megan L, Ing, Marsha, Wong, Jacqueline, Fernandez, Cecilia H, Shin, Nami, & Turrou, Angela C. (2014). Engaging with others' mathematical ideas:

Interrelationships among student participation, teachers' instructional practices, and learning. *International Journal of Educational Research*, 63, 79-93.

Wise, Arthur E, Darling-Hammond, Linda, Berry, Barnett, & Berliner, David C. (1987). *Effective teacher selection*: Rand Santa Monica, CA.

Woods, Robert C, & Montagno, Ray V. (1997). Determining the negative effect of teacher attendance on student achievement. *Education*, 118(2), 307.

Young, I Phillip, & Delli, Dane A. (2002). The validity of the teacher perceiver interview for predicting performance of classroom teachers. *Educational Administration Quarterly*, 38(5), 586-612.

Appendix A. Interview Protocol for March 2014 Site Visit

Interview Questions for Teacher Hiring Personnel

Research on Staffing Practices in Huntsville City Schools Informed Consent

This study involves working with the Huntsville, City Schools to assess their teacher selection/hiring process and explore the relationship between the hiring process and teacher effectiveness. It will involve reviewing their current hiring system, assessing available secondary data sources that might inform a new system, and looking at the relationships between these sources and teacher effectiveness. Data will include information collected from teachers and students (aggregated at the teacher level) in the school district, as well as interviews with personnel involved in the hiring process. The interviews will last approximately 45 minutes and will cover the various elements of the teacher hiring process.

We will use information gathered in this interview to inform understandings about how the teacher hiring process works. That information will help inform the data analysis process and interpretation of the results of the data analysis. The study will result in a report to Huntsville City Schools with findings from the data analysis and recommendations about how to improve the process, as well as a dissertation that summarizes the research study. We will protect the information you provide in several ways. Aside from de-identified quotations, we will not share your individual responses with anyone else outside the project except as required by law. We will not identify any individuals by name in our study reports. Your responses will be combined with others and reported in the aggregate. If quotations are used in any written reports, they will be included only for illustrative purposes and will not be attributed

to any individual. Given your unique role and knowledge of the school, though, it is possible that quotations or insights you share could reveal your identity. We will try to avoid this to the greatest extent possible. At the end of the study, we will destroy any information that identifies you.

Please feel free to tell us if you would like to share information with us that you would us like to keep “off the record.” We will not include that information in any of our reports.

Your participation in this interview is voluntary. You may choose not to participate, decline to answer any question, or stop the interview at any time.

If you grant us permission, we will record the interview on an audiotape/digital recorder, so we can check our notes for accuracy. At the end of the study we will destroy the tapes/recordings.

If you have any questions about the study, please contact the Principal Investigator, Mollie Rudnick, 1776 Main Street, PO Box 2138, Santa Monica, CA 310.393.0411 ext. 6535 or by email at mrudnick@rand.org.

If you have any questions or concerns about your rights as a research participant, please contact the Human Subjects Protection Committee at RAND, 1776 Main Street, PO Box 2138, Santa Monica, CA 90407, 310-393-0411, ext. 6124, or by email at Carolyn_Tschopik@rand.org.

Recruitment/ Teacher Pool

- Where does the teacher pool come from (e.g., local teacher education programs, alternative certification programs)?
- How do you advertise open positions?

General Teacher Qualities

- What is your definition of a highly effective teacher (e.g., raise student achievement on standardized tests, process skills, improve student affect, etc.)?
- What characteristics are you looking for in the teachers you hire?
 - How do you know if they have those characteristics or not?
 - Why are these characteristics important?
 - Which of the characteristics are most important?
 - How do you know whether or not teachers (and candidates) have those qualities?
- Are there any red flags (characteristics that are undesirable)?

Screening/Hiring Process

- What steps does a prospective teacher go through before they are hired? What are the different parts of the hiring process (e.g., screening, interview, hiring, school assignment, class assignment)?
 - Who is involved in each step (from the district and schools)?

- How do you expect each of those elements to be related to teacher effectiveness (e.g., teachers who possess that are more likely to stay in their jobs longer, have students who get higher test scores)?
 - Have there been any efforts to connect those elements to teacher effectiveness (i.e., predictive analytics)?
- How are each of those elements rated?
 - Are there rubrics to rate the components?
 - How do you establish reliability across raters, candidates, and time (e.g., rubrics, training)?
- How do you determine which teachers are hired?
 - Are teachers given an overall score?
 - Are there subjective elements to that determination?
 - How is each of the screening elements weighted/considered?
 - Do you track how many teachers dropout/are eliminated through the process?
- What is the timing of the process?
 - When do you start posting vacancies for the following year?
 - When do you start interviewing teachers for the following school year?
 - What percentage of your open positions are filled before the beginning of the school year (as opposed to after)?
 - About how many vacancies do you have per year?
 - How does that break down by subject and grade level?
 - Which vacancies are hardest to fill?
 - How many vacancies usually have to be filled in the middle of the year?
 - What is the process for seniority-based hiring and transfer?
 - How long do you have to post a position internally before it can be posted externally?
 - What percentage of vacancies are filled internally versus externally?
 - Do principals have to accept an internal hire before they can hire from the outside?

Policies and Context

- Are there any state or local policies that influence how you hire teachers?
- What lends support to the hiring process and hiring effective teachers?
- What hinders the hiring process and hiring effective teachers?