POLICY BRIEF
Cognitive Ability and Teacher Efficacy¹
January 2018 • Principal Author: Dr. David Steiner

“Having a top-quartile teacher rather than a bottom-quartile teacher four years in a row would be enough to close the black-white test score gap” (Gordon, Kane, and Staiger 2006).

Summary of Recommendations

- Combining the results of several assessments will give a stronger signal about future teaching effectiveness than using any single measure. Relevant assessment categories include (a) cognitive assessments not specific to teaching (e.g., SAT, ACT); (b) teacher-specific assessments (e.g., Praxis, edTPA, PPAT) that capture subject- and grade-specific content-knowledge or general pedagogical skills; and (c) assessments of “non-cognitive” or social skills (e.g., Haberman PreScreener). These different measures may be used to different degrees and with different weights depending on the teaching area the candidate seeks to enter.

¹ Elements of this policy brief were developed with the partnership of Chiefs for Change. The Institute is grateful to Chiefs for Change for permission to use this material. The author would also like to thank Dr. David Blazar for his consultation on its substance.
• Existing research cannot support requiring a specific score on any single assessment as an absolute barrier to entry into the teaching profession. However, very weak or very strong scores – at least a full standard deviation below or above a college-ready score on a content assessment – are likely to be appropriate signals of a teachers’ future effectiveness in some dimensions of teaching practice.

• For middle- and high-school teacher candidates who plan on specializing in specific content areas, it is appropriate to set a floor score on a content-knowledge exam. This is especially true in the case of high-school STEM subjects.

• Well-constructed hiring screens that focus on the skills teachers need in the classroom (e.g., ability to manage a classroom effectively) are recommended.

• The best predictor of teaching skill is teaching in the classroom – thus a full-year residency in which teachers have multiple, supervised opportunities to demonstrate their effectiveness (as indicated by student learning gains) remains the best screening device for future full-time employment.

The Challenge

The strongest education research finding in the last twenty years is that the quality of a teacher is the single greatest in-school determinate of student outcomes (E. Hanushek 1992). Having a high-quality teacher leads not only to better academic progress in school, but also to increased development of social skills, reduced risk of adverse life events (e.g., teenage pregnancy), and greater gains in the labor market (Chetty et al. 2011), (Chetty, Friedman, and Rockoff 2014). Stanford economist Eric Hanushek has quantified the differential economic impact of teacher effectiveness on lifetime earnings, showing that, relative to having an average teacher, having a 90th percentile teacher can improve the lifetime earnings of a class of 30 students by almost $1,000,000 (E. Hanushek 2011):

The importance of high-quality teachers to students’ lifelong success is most pronounced for historically under-served populations. The achievement gains from
a highly effective teacher are strongest for minority and low-income students (Policy Studies Associates 2005).

How then should we select the teachers who have the greatest capacity for effective instruction, both when they enter the classroom and also over the course of their careers?

**What do we know about the components of effective teaching?**

![Image of pie chart showing the distribution of factors influencing teacher effectiveness.]

Elusive Qualities

*The teacher characteristics that we can measure—experience, education level, certification status, and so on—only explain 3 percent of the differences in student achievement that are attributable to their teachers’ influence.*

% of Teachers’ Influence Attributable to Various Characteristics

Measurable Aspects of Teacher Quality: Years of Experience, Education Level, Performance on Vocabulary Tests

- 3%

Intangible Aspects of Teacher Quality such as Enthusiasm and Skill in Conveying Knowledge

- 97%

If we could easily measure the degree to which prospective teachers exhibited the characteristics, skills, and knowledge that are most powerfully correlated with effective results in the classroom, we could naturally create a teacher selection process that mirrored these characteristics.

Despite convincing evidence on the importance of teachers to student outcomes, decades' worth of research has been generally unsuccessful in identifying the specific characteristics that differentiate individual teachers. Goldhaber suggests that only 3% of the variation of teacher effectiveness in terms of their students’ test scores is explained by observable characteristics (see figure to the left) (Dan Goldhaber 2002). Some of these characteristics, such as teaching experience, are not relevant for teacher candidates prior to entering the classroom, leaving as much as 97% of the differences between teachers unexplained.

Understanding that as yet we can only measure a fraction of what makes for an effective teacher still leaves us responsible for making choices about that fraction. Research indicates that some tools currently used to screen teachers for specific characteristics do indeed have policy usefulness.

**Do smarter individuals make better teachers?**

For some time, the general approach has been to ensure that teachers are reasonably smart, as signified by an adequate score on a widely accepted cognitive assessment such as the SAT or ACT. The strongest case for using cognitive ability as a key selection filter has been made by Russ Whitehurst (Whitehurst 2005). Synthesizing across a range of research, he writes:

> The most robust finding in the research literature is the effect of teacher verbal and cognitive ability on student achievement. Every study that has included a valid
measure of teacher verbal or cognitive ability has found that it accounts for more variance in student achievement than any other measured characteristic of teachers.²

Other studies support Whitehurst’s general conclusions. Summers and Wolfe found that student achievement is positively associated with the selectivity of the college a teacher attended, which may be seen as a proxy for cognitive ability or achievement. Ferguson and Ladd (1996) and Kane et al. (T. Kane, Rockoff, and Staiger 2006) find that a teacher’s ACT scores are positively related to teacher quality, as measured by gains in students’ academic performance. These trends also hold up for teacher licensure scores, which rather than focus on cognitive ability, attempt to capture some dimensions of content knowledge. There is supporting research from the United States (Cloftelser, Ladd, and Vigdor 2006), (Dan Goldhaber 2007)³, as well as outside of the U.S. context (see, for example, Hanushek, Piopiunik, and Wiederhold 2014).

But other studies that relate teacher candidates’ academic background and later effectiveness (Harris and Sass 2006) show no correlations, and thus call into question the magnitude and importance of this relationship. Even in studies that do show a correlation, the relationship is weak. For example, examining Helen Ladd’s findings that ACT test scores are predictive of teacher effectiveness, we find that the actual results are far from strong: students assigned to teachers with higher licensure test scores apparently do better in math, but the effect is relatively modest. A 1 standard deviation (SD) increase in teacher test score implies at most a 0.017 SD increase in average student math test scores and a somewhat smaller increase in reading scores.⁴

As a point of comparison, these student achievement gains are smaller than the difference between a teacher’s effectiveness growth from their first to second year in the classroom (Kane, Rockoff, and Staiger 2006). They are far less significant than one-year academic gains attributable to shifting from a poor curriculum to a strong one (Chiefs for Change 2017). Moreover, screening for students with content knowledge within one SD of the most used cut score – college readiness – will result in the exclusion of large numbers of minority candidates. On the ACT exam, a score of 21 is considered “college-ready.” On average, black student test-takers score roughly a full SD (where 1 SD is equivalent to 4.7 scale points) below this, at roughly 17 (SAT and ACT Scores by Race/Ethnicity: 2003, 2006). We return to the challenges presented here, particularly with developing a diverse teacher workforce, below.

An important caveat to these findings is the unique case of high-school math and science teachers. Examining the relationship between subject-specific content-based

---

² Please see Whitehurst conference text for full references.
³ The 2006 version of Goldhaber’s paper is used below.
⁴ Some questions have also been raised about technical issues with this and related studies. Goldhaber et al (2002) write: “The fact that these studies were done at the aggregate (school or school district) level casts some doubt on them. It’s unclear whether higher-scoring teachers lead to higher-scoring students or whether affluent districts, which tend to have higher-achieving students, also tend to hire teachers with higher scores.”
licensure tests and middle- and high-school student outcomes in math and science, Goldhaber, Gratz, and Theobald (2016) find that basic skills tests are moderately predictive of middle- and high-school students' math performance and highly predictive of high-school science performance.

Summarizing the research, Goldhaber concludes that “teacher licensure test performance is a signal of teacher effectiveness,” and the policy implications are clear: "If states are seeking criteria to ensure a basic level of quality, than licensure tests appear to have some student achievement validity." But Goldhaber strongly cautions that the relationship is modest:

The point estimates, however, are generally quite small, [thus] states face significant tradeoffs when they require particular performance levels as a precondition to becoming a teacher: despite the testing, many teachers whom we might wish were not in the teacher workforce based on their contribution toward student achievement are nevertheless eligible based on their test performance. Conversely, many individuals who would be effective teachers are ineligible due to their performance on these licensure tests.

We can conclude that the use of scores from cognitive assessments and licensure tests should only be used as one indicator of a future teacher's potential skills.

**What about other knowledge and skills that teachers bring to the table?**

High-quality teaching is more than just developing students' content knowledge. Effective teachers also ensure that they are motivated, can persevere, and regulate their own behavior (D. Blazar and Kraft 2017) which in turn influences lifelong outcomes (Jackson 2016). Therefore, it may be equally important to screen teachers for a set of “non-cognitive” factors that may enable them to be successful in the classroom.

In one of the earliest papers on this topic, Rockoff, Jacob and Kane (2011) attempted to isolate the predictive validity of cognitive and non-cognitive factors to forecast the effectiveness of teachers in the classroom. Specifically, the team tested the relative importance of the following non-cognitive factors upon teachers' ability to improve students' test scores: teachers' self-efficacy; personality traits such as agreeableness, conscientiousness, emotional stability, extraversion, and openness to experience; and a measure of beliefs and values assessed via two commercially-available tools that purport to predict teacher effectiveness - the Haberman PreScreener and the Gallup TeacherInsight Assessment (Gallup TIA). Also included in their analysis were several cognitive indicators: undergraduate major; graduate education; selectivity of undergraduate institution; college entrance scores (i.e., ACT or SAT); the Liberal Arts and Science Test (LAST), the primary teacher certification exam used in New York State; a direct test of cognitive ability, Raven’s Progressive Matrices Standard Version; and mathematics assessment.
In models that focused grouping together two sets of predictors – a composite measure of teachers’ non-cognitive skill and a measure of their cognitive knowledge – the authors found that both were associated with gains in students’ academic performance at roughly similar magnitudes (roughly 0.025 SD). These estimates suggest that the set of non-cognitive measures provide additional useful information to differentiate teachers.\(^5\)

Additional lines of research that focus on teachers’ classroom behaviors suggest that some of these are more strongly predictive of productivity than either set of characteristics described above. For example, in the Measures of Effective Teaching (MET) project, teaching skills such as behavior management techniques and relationships with students had moderate relationships to student achievement gains (T. J. Kane and Staiger 2012). Blazar found that these sorts of teaching skills were much stronger predictors of gains in students’ math test scores than traditional measures of teachers’ content knowledge and other observable characteristics (e.g., experience) (David Blazar 2015). Blazar and Kraft (2017) found that in-service teachers’ general pedagogical skills (e.g., behavior management or relationships with students) and math-specific skills (e.g., correct presentation of content) were relatively strong predictors of several dimensions of teacher quality, including gains in students’ test scores, engagement in class, and self-reported self-efficacy. Relationships were in the range of 0.1 to 0.2 SD, upwards of ten times the magnitude of relationships described above relating teachers’ content knowledge or non-cognitive skills to student achievement gains.

This research focuses on in-service teachers because the skills being assessed are only fully observable in real classrooms. Because these skills are important predictors of future success, residency training programs, in which future teachers are able to teach while being supervised by mentor teachers, are clearly a promising structure through which to screen for teacher quality.

In recent years, testing companies have sought to develop performance assessments that measure the kinds of skills pinpointed in this research. The best known – the edTPA developed at Stanford University - involves a videotape of the teacher-candidate in the classroom, along with her or his analysis of what was occurring and why. Because the edTPA is a relatively new tool, very little research on its predictive validity exists. To date, research has come to mixed conclusions, with some data suggesting a moderate, positive relationship between performance on the edTPA and later effectiveness (Bastian and Lys 2016), but potentially only in some subject areas (Goldhaber, Cowan, and Theobald, 2017). One concern with this line of research is what exactly teacher candidates’ edTPA scores indicate, given the varied nature of the tasks and skills

\(^5\) A caveat to the Rockoff study is however in order: the range of cognitive abilities within the sample group was more limited than would be those of the general cohort of teachers who would fall under a new state policy. This sample group was comprised of candidates already selected to teach in NYC by the NYC Department of Education. The TFA and Teaching Fellows who make up part of the sample were especially likely to skew an already-selective group towards the higher end of the cognitive scale.
assessed, and the fact that faculty at teacher preparation programs can have very varied influence on the written portion of the assessment.\(^6\)

Some school districts use screening tools that match the observation protocols described above more directly, and that are used for in-service teachers. In the Spokane Public School system in Washington State, screening protocols that assess prospective teachers on an array of classroom practices show promise. Teachers’ observed classroom management skills were strongly predictive of later student achievement gains in both math and reading, as were teacher skills related to general instruction (Goldhaber, Grout, and Huntington-Klein 2017).

These findings suggest that efforts to screen student-teachers on skills involving interactions with students in the classroom are amongst the more promising of potential screens for identifying effective future teachers. But they are not yet nearly as effective as using student learning gains as the measure of teacher effectiveness.\(^7\)

**Beyond Cognitive and Non-Cognitive Attributes**

In recent years, policymakers are paying ever-increasing attention to recruiting an ethnically and racially diverse teacher workforce. This focus is motivated, in part, by an obvious imbalance between the percentages of minority students and that of minority teachers. In 2014, the Center for American Progress noted, "Students of color make up almost half of the public school population. But teachers of color are just 18 percent of the teaching profession" (Boser 2014).

There are several challenges to raising the percentage of minority teachers in the profession. One that has gained considerable attention is the licensure exam, which remains a barrier to entry despite the research above that suggests that the skills assessed on these exams (and the exam scores themselves) are often weakly related to later effectiveness.

Gaps in licensure exam scores between white, black, Latino, and Asian teacher candidates map closely to the test score gap between white, black, Latino, and Asian students. Using data from the College Board, Reeves and Dimitrios (2017) show that black and Latino students score significantly lower on the SAT than do white and Asian students:

---

\(^6\) Faculty are not supposed to actually write any of what a student submits. But preparation for the edTPA assessment can involve close coaching of how to respond to the written prompts. As became clear to me when serving as Commissioner of Education in NYS, which requires the edTPA for teacher certification, the level at which faculty is involved in direct support for a candidate’s submission varies across schools of education.

\(^7\) For a variety of reasons – some political, some technical – the use of value-added measures (VAM), which link teacher performance to their student learning gains remains controversial. But major studies show that the measure has stronger validity than any (currently available) alternative (T. Kane and Cantrell 2010).
For students in the same income bracket, the gaps are often equal to or greater than 1.5 SD. These gaps also have grown over time (Jaschik 2015):

### Combined SAT Score, and Changes Since 2006, by Race/Ethnicity

<table>
<thead>
<tr>
<th>Group</th>
<th>Combined Score 2015</th>
<th>Change Since 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian</td>
<td>1423</td>
<td>-27</td>
</tr>
<tr>
<td>Asian-American</td>
<td>1654</td>
<td>+54</td>
</tr>
<tr>
<td>Black</td>
<td>1277</td>
<td>-14</td>
</tr>
<tr>
<td>Mexican-American</td>
<td>1343</td>
<td>-28</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>1347</td>
<td>-16</td>
</tr>
<tr>
<td>Other Hispanic</td>
<td>1345</td>
<td>-26</td>
</tr>
<tr>
<td>White</td>
<td>1576</td>
<td>-6</td>
</tr>
</tbody>
</table>
Increasing diversity in the teacher workforce is more than an issue of inclusion; it also is an issue of equity for students. Evidence from the random assignment of teachers to students indicates that having a same-race teacher improves student test scores, particularly for black students (Dee, 2011). Further research suggests that when minority teachers are paired with minority students, the results include fewer suspensions and expulsions (Lindsay and Hart 2017), more referrals to gifted classes (Grissom and Redding 2016), and lower dropout rates (Gershenson et al. 2017). Black teachers, research has found, also have higher academic expectations for black students than other teachers (Gershenson, Holt, and Papageorge 2016).

The key question comes down to this: for minority students, what is the relative impact on their learning outcomes of having a teacher of their own race compared with having a teacher with stronger licensure exam scores (assuming one has to choose)? Frankly, we don’t know for sure. To date, only one study has addressed this issue directly. Using North Carolina certification test scores on the Praxis teacher test, Goldhaber and Hansen (2009) show that Praxis results are more predictive of white students’ performance than black students’ performance.\(^8\) Aligning these findings with other research, the authors conclude:

> Black students significantly benefit... from being matched with a Black teacher....[moreover] Black teachers’ students, on average, outperformed White teachers’ students in both reading and math, holding observable student attributes constant...This relationship holds in spite of Black teachers’ poorer average performance on licensure tests... Enforcing strict cutoffs has the potential to both adversely affect minority student outcomes and decrease workforce diversity.

**Can we do better?**

Given the modest effectiveness of using teacher-specific assessments such as the Praxis as predictors of future teacher performance, and the absence of recent, robust research showing that using ACT/SAT would do any better, it is clear that relying on a cut score on a single test as a gateway to the profession is problematic. Including a Praxis score as one of several data points that school districts can look at makes sense; districts can then balance the modest, general predictive power of that score against other information that could push their decision in another direction. Using an assessment of relevant non-cognitive attributes (such as the Haberman test that Rockoff uses) is one possible addition.

\(^8\) We should note one important finding from Goldhaber that can get lost in the summaries. The Praxis II was actually two exams – one in curriculum, instruction and assessment, the other in math and literacy. The former was modestly predictive of white teachers’ impact on student learning, the latter of black teachers’ impact.
The greatest promise appears to come from screening protocols that assess the skills that teachers enact directly in the classroom. Prospective teachers' ability with classroom management techniques and other instructional skills is predictive of their (eventual) students' test-score performance. While research on appropriate screening devices is still rare, the findings thus far suggest that building effective screening protocols is both possible and important.
About

The Johns Hopkins Institute for Education Policy (the Institute) believes that building partnerships across different constituencies is necessary to advance excellence and equity for all of America’s children. The Institute operates on the understanding that education policy must be informed both by real-world conditions and also by excellent research; that it is possible to translate the technical vocabularies of research into a language that is accessible and useful to policy experts, principals, teachers, and parents; and that in our richly diverse nation, education must be driven and sustained by evidence about what works and what does not. http://edpolicy.education.jhu.edu

David Steiner is Executive Director of the Johns Hopkins Institute for Education Policy and Professor of Education at Johns Hopkins University.
Citations


