Preface

- There are large differences in teacher quality (as measured by value-added), but the vast majority is due to unexplained/unobserved/unmeasured factors.
- Lack of evidence for a relationship between specific teacher characteristics and student outcomes could mean (a) there is no relationship or it could mean (b) that no one has studied it or (c) data aren’t widely/readily available.
- In addition, for many characteristics, the “within” group differences are bigger than the “between” differences– e.g. the differences in effectiveness within the group of all traditionally certified teachers are much larger than the differences between traditionally and alternatively certified teachers.
- In general, teacher effects tend to be bigger for math than reading, so we know more about what makes effective math teachers than we do about what makes effective reading teachers.
- The predictive/explanatory power of multiple indicators grouped together is much stronger than using any one indicator individually. Thus, relying on a single indicator is likely to result in screening out teachers who would be successful (false negatives) or accepting teachers who will not be successful (false positives)– it’s important to consider which of these errors you’re most interested in trying to reduce.
- A lot of these characteristics are related – some positively, some negatively. Ideally, you want teachers who perform well on all the measures but, of course, this isn’t always possible, and tradeoffs are necessary. For example (and to greatly oversimplify), Teach for America corps members tend to have high academic skills but are low on pedagogical training, while teachers from colleges of education tend to be high in pedagogical training but lower in academic skills, on average.
- Obviously, numerous teacher characteristics (experience, practices, etc.) not included in this list are also important, and this document is not to be treated as gospel – I’m sure I missed some important stuff, and research is constantly evolving.

Summaries of Research Evidence

Academic Aptitude

- Operationalized as college selectivity, ACT/SAT/other aptitude test scores, verbal ability, etc.
- Students learn more from teachers with higher aptitude test scores and better verbal ability (Wayne and Youngs, 2003; Rice, 2003; Goldhaber, 1998; Whitehurst, G.J., 2002; Ferguson and Ladd, 1996; Ehrenberg and Brewer, 1995, Ferguson 1991)
  - This is particularly true for at-risk students. (Rice, 2003)
There is little relationship between results of basic skills test and teacher effectiveness. (Rockoff, 2008)
- But BST (Praxis I) results are predictive of content test (Praxis II) results. (Gitomer, 2008)
- Students learn more from teachers from more competitive undergraduate colleges (Rice, 2003; Wayne and Youngs, 2003; Summers, 1977; Ehrenberg, 1994; Winkler, 1975, Clotfelter et al., 2007, Boyd et al., 2008a, Rockoff, 2008.)
  - This is particularly true for at-risk students. (Summers, 1977)
  - There might be a ceiling effect, with competitive better than less competitive, but no difference between competitive and elite. (Clotfelter et al 2007)

Content Knowledge

- Operationalized as content test results, course-taking, degrees, major, etc.
- Students learn more from teachers with more knowledge of the content they are teaching, regardless of measured by course-taking, certification, or degree (Rice, 2003)
  - But also in science. (Chaney 1995; Goldhaber and Brewer 1998; Monk 1994; Wenglinsky 2000.)
  - Less predictive for reading. (Clotfelter et al 2007; Goldhaber, 1997) and elementary school (Ebets, 1984; Rowan, 2002)
  - Some limited evidence suggests that there may be a ceiling effect (e.g. teachers who had taken seven courses were no more effective than teachers who had taken four to six). (Monk, 1994)

Pedagogy

- Operationalized as education major, coursework, or degrees, traditional certification (vs. other routes), pedagogy test results, etc.
- Certification: There is generally more variation within than between various certification pathways
  - Traditional certification: this matters for HS math, but there is little evidence for a relationship with student achievement in lower grades (Rice, 2003)
  - Emergency certification: worse or no difference (Jepsen, 2002, Goldhaber and Brewer 2000, Clotfelter et al 2007)
  - Alternative route: varies widely by route and characteristics of route completers, but, generally, this route produces a small negative effect that vanishes quickly with experience (Clotfelter et al 2007)
    - TFA (and similar academically selective programs): generally, equal to veteran traditionally certified teachers and more effective than other beginning teachers (Decker, 2004; Rockoff, 2008)
• NBPTS: identifies better teachers, but does not make them better (Cavalluzzo, 2004; Goldhaber, 2006; Goldhaber and Anthony, 2004)

• Education coursework & pedagogy tests: Mixed evidence
  o Mixed results may stem from the low academic caliber, on average, of the people that take such coursework (Goldhaber, in NCTQ)
  o Pedagogy coursework seems to matter (up to a point) at all grade levels (Rice, 2003)
  o Pedagogy test performance seems to matter in elementary reading and math (Clotfelter et al 2007)

• There is clear and consistent evidence of no (or negative) effects for masters degrees in education. (Goe, 2007, among many others)

• Licensure Tests:
  o Mixed results, perhaps stemming from the fact that teacher performance data are available only for teachers who pass the tests (Buddin, 2008) and because predicative validity depends on teacher race (Goldhaber, 2009).

Pedagogical Content Knowledge (PCK)

• Defined as a blend of subject matter knowledge and understanding of how that information is best taught to others. (Schulman (1986, 1987)

• Positive relationship between math PCK and student achievement gains in elementary school math – stronger than math knowledge alone (Hill et al., 2005; Rockoff, et al. 2008; Monk,1994)

• Some measures of PC (Ball and Pianta) are currently being investigated further by the Gates Foundation’s MET project.

• Understanding of the science of teaching reading: A lot is known about how children best learn to read (National Reading Panel, 2000), but there’s little direct, empirical evidence linking teachers’ understanding of this to student achievement gains

Performance During Student Teaching

• For the purposes of this document, operationalized as (limited) information on past value-added and teacher evaluation results.

• Year-to-Year and test-to-test correlations are moderate (around 0.3) (Goldhaber, 2010; Sass, 2008).
  o Consistent with performance measures from other occupations, and similar to batting averages or the relationship between SAT and freshman GPA. (Goldhaber, 2010; Glazerman, 2010)

• Principal ratings can effectively distinguish which teachers will have the largest impact on student achievement; principal ratings are more effective predictors than teacher experience and educational attainment (Harris and Sass, 2007; Jacob and Lefgren, 2008; Rockoff, et al. 2008).

• Harris and Sass (2007) find that prior value-added measures and principal evaluations predict current student achievement gains equally well, while Jacob and Lefgren (2008)
find that past value-added estimates generally do a better job than do subjective principal assessments.

Dispositions, Soft Skills, and Personal Characteristics

- Student-teacher match on race: There is some evidence of benefits for both black and white students. (Clotfelter, Ladd, Vigdor, 2007; Dee, 2004; Steele, 1997.)
- Multiphasic Personality Inventory (MMPI) and in the Big Five personality traits: little to no relationship with teacher effectiveness (Getzels and Jackson (1963), Rockoff et al., 2008)
- School-level measures of social capital (collegiality), collective efficacy, and high expectations: There is a positive relationship with school performance, but difficult to discern if this is a cause or a consequence of performance. (Rockoff et al., 2008; Goe, 2007)
- Teach For America (TFA) and the Haberman Educational Foundation have done a lot of work on the relationship between teachers’ soft skills and teacher effectiveness, but a lot of this research is “in house” and unpublished.
  - Reports on the effectiveness of Teach for America teachers suggests that their selection process identifies more effective teachers (Rockoff, et al. 2008; Decker, 2004)
  - There is some evidence that the Haberman Pre-Screener can identify characteristics that are correlated with teacher quality. (Rockoff et al., 2008)
    - There is little systematic evidence on the validity of other commercial teacher selection instruments (Rockoff et al., 2008)
References


Steele, Claude M. A threat in the air—How stereotypes shape intellectual identify and performance, American Psychologist 52(6): 613-629, June 1997.)


