Teacher intelligence: What is it and why do we care?

Andrew J McEachin
Provost Fellow
University of Southern California

Dominic J Brewer
Associate Dean for Research & Faculty Affairs
Clifford H. & Betty C. Allen Professor in Urban Leadership
Professor of Education, Economics & Policy
University of Southern California
INTRODUCTION

Over the past several decades, few topics have been discussed and researched more than the decline or stagnation of the United States K-12 public education system. The Coleman Report (Coleman, Campbell, & Hobson, 1966) was one of the first attempts to parse out the causes and consequences of achievement disparities between the rich and poor and Caucasians and members of racial and ethnic groups. Although the methods used in the report have been widely debated, a key finding was that teachers’ aptitude, or verbal intelligence, was highly correlated with student achievement. In the report, teachers were given a short assessment measuring their verbal abilities at the end of a survey. The teachers’ score on the assessment was highly correlated with the academic performance of their students (Ehrenberg & Brewer, 1995). Since then, researchers, policy-makers, and district officials have tried to measure teacher intelligence using a variety of proxies: SATs, college selectivity, licensure exams, and undergraduate major (e.g., Angrist & Guryan, 2008; Clotfelter, Ladd, & Vigdor, 2007; Goldhaber & Hansen, 2010). In this chapter we first broadly define intelligence, and discuss the different ways it has been measured and used. We then focus on how measures and proxies of teacher intelligence have been used in education and their relationship with student achievement. We close with comments on the use of measures of teacher intelligence in staffing and policy decisions.

What is Intelligence?

In a perfect world, one assessment could derive an error-free measure of intelligence which would be correlated with outcomes of interest—e.g., labor productivity, credit worthiness, voting patterns. In our imperfect world, however, intelligence cannot be properly evaluated by a single measure, nor can it be measured without error. We instead have to settle for proxies of intelligence. Furthermore, before one can meaningfully measure any latent traits (e.g., intelligence), one must first develop a meaningful construct. In other words, one must clearly define what is meant by “intelligence.”

Similar to a traditional K-12 public schools system where the difficulty of tasks and knowledge acquired build over time, intelligence is a hierarchical construct (see Shavelson, & Huang (2003) for more detail). At the lowest level is contextualized domain specific knowledge (e.g., 10th-grade Physics) and at the highest level is a decontextualized general ability (often referred to as “Intelligence Quotient” (IQ) or “G”). Even though measures of intelligence are neither perfect nor deterministic (Shavelson, & Huang, 2003), they can serve as useful predictors of future success. Broad abilities, the categories between domain specific knowledge and “intelligence”, such as verbal, quantitative and spatial reasoning, are commonly measured proxies for intelligence. For example, admission offices in colleges and universities around the world heavily rely on measures of broad abilities (e.g., SAT, GRE, LSAT).

Domain specific measures of intelligence are less generalizable than broad abilities measures. Replacing a board exam with an MCAT score would be a foolhardy policy. Whereas the latter may predict the future success in medical schools by assessing one’s verbal, quantitative, and spatial reasoning, it does not measure the specific knowledge required to be a doctor. Depending on the construct of interest, measures of intelligence are used in two ways: as screens and signals. Often measures of intelligence are used as screens, mechanisms that keep unqualified individuals from entering professions or other uniform communities. Doctors, lawyers, plumbers, and many other professionals must first attain a passing score on a licensing exam. One is only interested in whether the applicant attained a passing score, not the actual magnitude of the score. Domain specific measures of intelligence are the most commonly used screens. Conversely, the magnitude of an intelligence measure can be used as a signal for future success. Both domain specific (Praxis) and broad abilities (general GRE, MCAT) are often used as predictors for success in post-secondary education and job performance.

RESEARCH FINDINGS
Research has demonstrated that teachers are the single most important school-input in the education production process (Coleman, Campbell, & Hobson, 1966; Hanushek, Kain, O'brien, & Rivkin, 2005) For example, students with a teacher in the 85th percentile of teacher quality outperform students with a teacher at the median of teacher quality by .22 standard deviations (Hanushek, Kain, O'brien, & Rivkin, 2005). But how would one go about measuring teacher quality? In many circumstances, educational stakeholders use proxies of teacher intelligence as screens and signals of teacher quality. The research on the relationship between teacher ability, or characteristics, and student achievement often falls into five groups: the rating of teachers’ undergraduate institutions, course taking and degrees, certification status, test scores, and other (Wayne & Young, 2003). In this review, we focus on proxies of teacher intelligence: achievement and licensure tests (e.g., SAT, Praxis, CBEST). Starting with the Coleman report and ending with more recent literature on teacher licensure exams, we briefly review the teacher intelligence literature as it pertains to student achievement.

Teacher Testing

Researchers have used both decontextualized general measures of ability and domain specific knowledge as proxies for teacher intelligence. The first strand of literature dating back to the 1960s relied more on general measures of teachers’ verbal intelligence (Coleman et al, 1966; Ehrenberg, & Brewer, 1995), while the more recent research focuses on the teacher licensure exams which are measures of domain specific knowledge. Although both may be referred to measures of intelligence, they measure different skill sets and may therefore be differentially related to student achievement.

The Coleman report released in the 1960s started a wave of research attempting to connect measures of teacher intelligence to student achievement (Coleman et al, 1966). Early research found a positive link between teacher ability, as measured by an achievement test, and student achievement (Ehrenberg and Brewer, 1995). A specific association was found between teachers’ verbal ability, as measured by a short assessment at the end of the Coleman survey, and student achievement (Ehrenberg, & Brewer, 1995). College admission tests and selectivity have also been used as proxies for teacher intelligence or ability. Schools with teachers from more selective universities, a proxy for intelligence, had higher gains in student achievement than schools with lower average selectivity (Ehrenberg & Brewer, 1994).

A more recent wave of literature exploits the rise in the number of states requiring teacher-licensure exams. Since the passage of No Child Left Behind, states require teachers to meet certain requirements before applying for a teaching credential (Angrist, & Guryan, 2008; Goldhaber, 2007; Goldhaber, & Hansen, 2010). Recent research in the United States has begun to use student level longitudinal data to examine the relationship between teacher characteristics and student achievement (Clotfelter, Ladd, & Vigdor, 2007; Goldhaber, 2007). A one standard deviation increase in teachers’ Praxis scores yields only a 1-2% of a standard deviation increase in student achievement (Clotfelter et al, 2007). The results do not appear to be specific to just the United States. A similar positive relationship exists between teacher licensure exams and higher average student achievement in Mexico (Santibanez, 2007).

Furthermore, the one can assess the quality of the Praxis both in terms of its ability to act as a signal future teacher quality and as a screen for entering the teaching profession (Goldhaber & Hansen, 2010). The authors take advantage of a unique dataset that includes observations for teachers who initially failed the Praxis but were still allowed to teach, assuming they passed the Praxis within a calendar year. On average, the Praxis serves as a weak signal of teacher quality. There is a no relationship between passing the Praxis and students’ English scores and a small, positive relationship with math scores. Although the Praxis was not established to serve as a signal, the authors note that individuals involved in the teacher hiring process may use the scores when deciding between candidates. On average, the Praxis did not serve as a signal of teacher quality. Research also demonstrates that black teachers had a positive statistically significant effect on black students, regardless of their Praxis scores (Goldhaber & Hansen, 2010). Therefore, the interpretation of teachers’ Praxis score in North Carolina cannot be removed from other characteristics. Researchers have also explored the use of California licensure exams (CSET, CBEST, and RICA) as a signal for future teacher
performance. The scores on the CA licensure exams had no significant relationship with levels or gains in student ELA or Math achievement (Buddin & Zamarro, 2009). There is also no significant difference in student achievement gains between teachers who initially failed the California Licensure exams compare to teachers who passed the first time.

Developing research suggests that principals may devalue measures of teacher ability or intelligence during the interview process because they assume that credentialed teachers have met certain intellectual requirements during the credentialing process (Harris, Rutledge, Ingle, & Thompson, 2010). A conflict in the hiring process may well exist if the professional screens used by many states and districts are unrelated with student achievement. Furthermore, unless a more unified approach is taken, the relationship between teacher-licensure exams and student achievement will be heterogeneous across states.

To make matters more muddled, research suggests that the presence of teacher licensure exams has failed to raise the quality of the teaching profession. Using state-level variation in teacher licensure requirements, researchers find that the presence of teacher-licensure exams does not raise the quality of individuals entering the teaching profession (Angrist & Guryan, 2008). The cost of studying for, and the presence of measure error in, the exams can actually keep high-quality teachers from entering the profession. Furthermore, as noted above, licensure exams may also keep minority teachers kept out of the profession since they are more likely to fail the exam (Goldhaber, 2007). Other researchers have found that the quality of teachers, as measured by high school class ranking, college admission tests, degree type has declined over the past few decades, with the most notable changes at the tails of the quality distribution (Corcoran, Evans, & Schwab, 2004).

**SUMMARY AND RECOMMENDATIONS**

In this review we focused on the relationship between teacher ability, as a predictor of teacher quality, and student achievement. As noted by the above discussion, observable teacher characteristics, “Although not a complete measure of teacher quality, teacher ability is clearly a quantifiable predictor of teacher effectiveness in the classroom” (Pelayo & Brewer, 2010, p. 180). The question now becomes: “How do we validly and reliably measure teacher ability as a predictor of future teacher quality? (Pelayo & Brewer, 2010, p. 180)” The prior literature paints a confusing picture for policy-makers and education stakeholders. The relationship between teachers’ grades, test scores, university selectivity, and other observable factors are loosely related to student achievement. Yet, teachers have a large impact in student learning (Hanushek, Kain, O’Brien, & Rivkin, 2005). Policy-makers and educational stakeholders would benefit from a more objective, unified measure of teacher ability.
References


