EDUCATIONnext

A Flexner Report on Teacher Preparation

By Thomas J. Kane | April 10, 2014

Abraham Flexner transformed American medical training with his 1910 report, "Medical Education in the United States and Canada." His chief recommendations—higher admission standards, two years of laboratory training, two years of clinical training in a hospital setting—left an imprint which is still visible a century later.

However, the landscape of medical education at the beginning of the last century was very different from the state of teacher preparation today. Many university-based medical schools were already combining laboratory training with clinical training in an affiliated hospital. The American Medical Association had been championing such a model before 1910. By personally visiting all the medical schools in the country, Flexner documented conditions in schools which were *not* using that model (for instance, identifying those still teaching students homeopathic medicine and flagging those with grossly inadequate laboratory facilities.) The Flexner report did indeed transform medical education—but not by persuading schools to change. Rather, the report created pressure on state licensing agencies to *close* the medical schools which were teaching outdated theories or providing inadequate facilities. In 1910, there were 155 M.D. granting institutions, with more than 25,000 students. By 1935, there were 66 schools (and about half as many medical students as before).

In teacher preparation, there are no model programs (at least none that are broadly recognized). A modern day Flexner report on teacher preparation would first need to provide the evidence for a new model.

Three types of changes are needed. The first is higher admission standards. When it comes to improving teacher education, raising admission barriers is tempting, because it's politically expedient and not complex to implement. However, raising admission standards can be costly in other ways, by eliminating from the pipeline anyone who could be discovered as effective later. If we had mechanisms to identify effective teachers in the initial years of teaching (or during training), it would be counter-productive to screen out during admissions. Therefore, we need to be confident that substantially higher admission standards will yield a substantially more effective pool of candidates.

There is some evidence to support higher standards, but it's not overwhelming. A recent random assignment study of the impact of middle school and high school teachers recruited by Teach for

America and The New Teacher Project provides reason for caution. Although the Teach for America corps members were statistically significantly more effective than comparison teachers, the difference was modest (.07 student-level standard deviations in math.) The strategy of highly selective recruitment, followed by minimal training, may work for Teach for America, but it's unlikely to work for a whole state or district. If TFA can produce only a .07 student-level standard deviation gain with their exhaustive efforts to recruit 5,000 of the most academically gifted candidates in the country, we should expect much less than that from the next 245,000 candidates in line. The differences in academic credentials between TFA corps members and other teachers were gigantic: 81 percent of TFA teachers had graduated from a selective college or university, compared with 23 percent of the comparison teachers. And, the TFA corps members scored .92 standard deviations higher on a test of mathematical content knowledge than the comparison teachers. However, the most readily measured academic qualifications (undergraduate institution and math knowledge) explained none of the difference between TFA teachers and the comparison. To the extent that TFA did identify promising young teachers, it seems to have been due to the more subtle factors they look for in their intensive interviewing process. But you can't write subtlety into state admission requirements.

Therefore, a second change would combine higher admission standards with better training. Even if many teacher preparation programs are not providing it effectively, we *know* that clinical training matters. How do we know this? Hundreds of studies using value-added methods have found that teachers improve in effectiveness during their first few years of teaching. Indeed, the magnitude of that growth has been strikingly consistent across a number of sites and research methodologies: the average teacher's effectiveness improves between .05 and .08 student-level standard deviations between their first and third years of teaching.

The question is whether we can produce similar improvements in effectiveness without putting teachers into the crucible of their own classrooms. Many hope so, but we actually don't know. (Much of the evidence on teacher residencies, for instance, suggests that despite a year shadowing a mentor teacher, former residents perform like other novices when they are eventually given their own classrooms.) However, the large number of teachers being hired as "alternatively certified" should offer us an opportunity to find out. Right now, most districts let candidates for alternative certification choose their own programs. That makes it hard to distinguish between the value of training and the types of candidates choosing different programs. Suppose a group of districts were to form a "study group on teacher training." Rather than let candidates choose, they could assign (ideally, randomly) candidates to different forms of teacher training—and learn from the results. Priority should be given to new models which emphasize practical training in classroom management and other skills. For instance, prospective teachers could be sent to the Relay Graduate School of Education or the Sposato Graduate School of Education for an intensive summer of training. At TeachingWorks at the University of Michigan, educators have been developing courses focused on certain practices they believe to be "high-leverage."

Better pre-service training would be worthwhile, if only because we could avoid imposing the cost of breaking in rookie teachers on our most needy students (in the form of diminished

achievement). However, eliminating the rookie slump with better pre-service training will *not* have a large impact on aggregate achievement on its own. It's simple math. About eight percent of all teachers are in their first year of teaching. If we were to eliminate the rookie slump entirely, we'd expect only a .0040 to .0064 standard deviation improvement overall (.08 * .05 to .08 standard deviations).

As a result, the new model program is likely to require a third change: greater selectivity after initial selection and before placement. We know that teachers have very heterogeneous effects on students. We also know that it's hard to identify effective teachers based on academic qualifications. However, we don't know how much could be learned about a teacher's potential *during their training* (before they are placed in a job). Perhaps training institutions could be more selective once they can directly assess a candidate's teaching (rather than at admission and before a candidate is assigned a classroom.) To investigate, we could ask trainers to predict which teachers are likely to succeed and then study the predictive validity of those assessments.

Along similar lines, some states are adopting a "performance assessment" which pre-service teachers are required to pass in order to become licensed. To be useful, such assessments must be tested for their predictive validity. If they're not sufficiently predictive, the tests would be just another costly barrier to entry, making matters worse, not better.

Finally, some states, such as Tennessee and Louisiana, have begun tracking the value-added scores and teacher evaluations of teachers entering the profession from different institutions. That's a huge step forward and other states should follow suit. However, unlike medical education, we can't expect to improve teacher training dramatically through regulatory action alone. In a recent study by Goldhaber and Liddle, the most highly rated programs in Washington state produced teachers with value-added of .046 student standard deviations higher than the average teacher in math. The worst programs were producing teachers whose students lost .020 standard deviations relative to the average teacher. (Similar work has been done in New York City, Missouri, Louisiana, Florida and North Carolina. Only in New York City were the differences sizeable.) These are not large differences. We won't raise average effectiveness much by closing programs when even the worst performers are not that different from average.

Because medical schools varied so greatly in quality in 1910, the Flexner report could have a huge impact by closing the worst ones. Yet because teacher preparation is almost universally weak, our challenge is quite different today. A modern-day Flexner report should focus on finding a more effective model of teacher training—combining all three mechanisms of initial selection, high quality training and post-admission selection based on future promise. Once such a model is developed, the state regulatory process could take it from there (like the state medical boards in Flexner's day), and close those programs which fail to implement the new model effectively.