Extended abstract:

Value-added and students' long-term outcomes

In recent years, there has been great interest in and investigation of the potential of value-added models (VAM) to identify teachers and schools of high quality. Chetty et.al. (2014a, b) find that VAM models controlling for lagged test scores exhibit little bias when used to forecast teacher quality and that VAM also successfully predict long-term outcomes, like college enrollment and earnings. In particular, when teachers change classes they teach, both test scores and long-term outcomes change as one would expect based on the teachers' estimated value-added. Despite several successful replications, this interpretation of the data is challenged by Rothstein (2017), who argues that that the findings of Chetty et.al in part follow from how they specify their analyses, that VAM have a non-trivial bias, and that there is little robust evidence about long-term effects.

In this paper I study school quality in a very different context: Norwegian compulsory schooling. A limitation doing so is that while exams and teacher grades from the end of compulsory schooling are available since early 2000's, standardized tests throughout compulsory school were introduced only from 2007. Thus, only for relatively recent cohorts is it possible to control for lagged test scores, and for these cohorts there is limited information on longer-term outcomes. Therefore I construct indicators of school quality by controlling for rich data on family background. For recent cohorts I compare these indicators to indicators controlling for lagged test scores. I also evaluate the indicators in a quasi-experimental way by comparing observed and expected outcomes of students who move and students in neighborhoods that change which school they are assigned to. I then proceed to investigate how the school quality indicators are related to students' long-term outcomes, including high school completion, labor force participation and earnings. Finally, I investigate causal effects of school quality on long-term outcomes using the same quasi-experiments as I use to validate the relevance of the indicators for learning outcomes.

I find that value-added estimates controlling only for socio-demographic characteristics gives similar results to indicators controlling for lagged scores. The different indicators have similar dispersion, and the correlation between different indicators for the same year is reasonably high (>.8). Also, lagged indicators controlling only for socio-demographic characteristics are as predictive about value-added indicators as lagged indicators controlling for lagged scores, suggesting little forecast bias. This is reassuring, given that Chetty et.al (2014a) and the following literature find that indicators controlling for lagged scores are informative about teacher quality. Also, Deming (2014) and Angrist et.al. (2015) find that indicators controlling for lagged scores are informative about school quality, studying random variation in school assignment from school lotteries.

In the two different quasi-experiments students' school assignment change, arguably independent of school quality. In the first I study actual movers, students that are first registered at one school and later at another. As students' school assignment is only observed in test records, I can only do this for recent student cohorts. Because information on school quality traditionally has not been publicly available in Norway, students changing schools are unlikely to reflect ambitious families moving to better schools. Also, there is no strong tendency in who change schools, or any large change in average results. For students that change schools early in secondary school I find that estimated school quality of both schools is predictive of exam scores at the end of lower secondary, conditional

on standardized test scores from the start of lower secondary and average student characteristics at the schools. The sum of the associations with the quality of the two schools is very similar to the association between estimated school quality and for students not changing schools. This holds both for indicators controlling for lagged test scores and for indicators using only data on socio-demographics, for current and lagged indicators and the estimated effect of the last school is robust to including fixed effects for the movers' first schools.

In the second quasi-experiment I study neighborhoods changing which school they belong to, i.e. neighborhoods where students overwhelmingly attend one school before a given year and another after that. Also in this case I find that school value-added is indeed informative about exam score at the end of compulsory schooling. When a neighborhood changes local school, exam scores change (I study exam score conditional on neighborhood fixed effects) about as much as would be expected from the change in school quality, given the level of compliance with the local school.

I then move on to study the relationship between estimated value-added and long-term outcomes, i.e. completion of upper secondary school and labor market outcomes. Conditioning on observables, I find that students attending schools with higher estimated quality have better progress through upper secondary, more often complete upper secondary, less often are neither employed nor under education in early adulthood and to some extent earn more (this is just barely so for the first cohort of students, the remaining students are still so young that there is a negative association between individual exam results and earnings). This is the case both when relating the long-term outcomes to school quality estimated from the students' own cohorts and when relating it to school quality estimated from different cohorts. Indeed, the relationship between school quality is remarkably consistent across different measures of school quality, and often also very similar to the corresponding cross-sectional relationships between individual students' results and long-term outcomes.

Finally, I address the causal impact of school quality on long-term outcomes using the two quasi-experimental approaches. Unfortunately, the available data puts severe limitations on this. As long as I only observe students moving recently, I have limited data on long-term outcomes for these. Furthermore, as neighborhoods' changes of schools are relatively rare, and I need to observe students both before and after the change, I'm likewise restricted studying such school changes. Still, despite these limitations, the available data for moving students strongly suggest that estimated school quality has a substantial impact on early outcomes from upper secondary (e.g. completion of the first year) as well as early measures of being outside employment and education. Also, there are indications of changes in long-term outcomes in the expected directions when neighborhoods change local schools.

Taken together, my results further confirm the importance of school quality, not only for improving students' short-term learning outcomes, but also their longer-term outcomes. I.e, good schools teach for life, not only to the tests. Furthermore, my results strongly suggest that it is possible to construct informative indicators of school quality, even lacking data on lagged scores. This is useful, as it means that it is possible to construct such indicators also in school systems and at stages where no previous test scores exist. As the Norwegian context illustrates, this allows construction and evaluation of indicators at times where there is only one achievement measure, and thus may also allow analysis of long-term consequences of school quality at a much earlier time than what would have been

possible if one were to wait for students to complete entry and exit test at different stages of their educational careers. Also, even in educational systems with systematic testing regimes, tests measuring students' performance before entering schools may not exist. This limits the possibility to construct value-added indicators for the entire contribution of the schools; any contribution before the first test will not be measured. This challenge can be circumvented if it is possible to construct indicators controlling only for data on family background, which is truly predetermined relative to the students' educational careers.

Finally, the current study links learning outcomes and long-term outcomes. E.g. when interventions in the schooling system are evaluated, results are usually in the form on an effect on learning outcomes. Indeed, actual long-term outcomes can only be studied a long time after the implementation of the intervention. However, the motivation for interventions is often, at least in part, a belief that they will promote not only short-term learning, but also longer-term outcomes. This study connects learning outcomes to long-term outcomes of interest to policy makers. Furthermore, it does so using fairly general variation in school quality, suggesting that the (implied) effect of learning on long-term outcomes may be generally relevant (as opposed to e.g. very specific interventions, that may impact strongly on either learning or long-term outcomes, depending on their exact design). While it is still too early to draw strong conclusions about earning effects, such knowledge will likely be relevant for informing cost-benefit analysis of future interventions in schools.