How much do Brazilian college students learn? Value-added estimations with ENADE, 2005-2011

The expansion of standardized tests and accountability systems in many countries reflect a global trend of assessing education quality. Although most of these efforts are focused on K-12, higher education is increasingly becoming the target of studies and public policies for education quality assessment. In this paper, I analyze students' learning gains and institutional factors that affect their test scores in a national context in which a comprehensive accountability system was established, but receives little attention, Brazil.

From 1994 to 2012, the number of students enrolled in higher education increased from 1.6 to 7.3 million, reaching a gross enrollment rate of 36%. Higher education institutions (HEI) that are privately administered increased their enrollment even faster until 2007, and currently have almost 75% of the market share, up from 65% two decades ago. Such rapid expansion led the government to be concerned about the quality of education, to which it responded by creating a nation-wide accountability system. The system started in 1998 with the National Exam of Courses, and changed in 2004 to the current version in which students take the National Exam of Student Performance (*Exame Nacional de Desempenho dos Estudantes* - ENADE).

This paper has as its overarching objectives to examine the evaluation of higher education programs in Brazil in order to understand the processes of this complex sector and to provide new insights for public policy. We explore three questions: A) How much the students are learning? B) Do more selective programs have higher value-added than those less selective? And C) Are quality of faculty and institutional resources correlated to higher scores?

I use data publicly available from ENADE test scores and the overall accountability system, both at the program and student-level from 2005, 2008, and 2011. The exam is accompanied by a survey that contains questions about students' socio-economic status (SES) and opinion about their programs. I analyze the five fields evaluated that had the most students enrolled in those years: Education, Language Arts, Biology, Electrical Engineering, and Computer Science. My estimation strategy in all three research questions is to compare value-added estimations using two approaches, intercohort (cross-sectional) and intracohort (longitudinal).

A) How much the students are learning?

The test scores of students in the final year are, on average, higher than those of first-year students. This means that students are improving their general and, particularly, their specific skills and knowledge by studying in college. Regarding equity, the variation of the value-added estimations is often reduced in the general test and increased in the specific exam. The distribution of general skills becomes more equal among programs as students reach the last year of study, with more program averages close to the overall

mean. On the other hand, the distribution of specific skills gets less equal over the course of higher education, because more students are farther from the mean.

The average value-added gains vary by year, field and approach used. In addition to the differences by field which are significant, the variation by year is smaller in the intercohort than on the intracohort measures. The intercohort approach provides a distribution of program average test scores significantly different from those obtained by the intracohort, particularly for the 2005-2008 students. The intercohort data are more stable than the intracohort, with similar results within each field between 2005 and 2008.

B) Do more selective programs have higher value-added than those less selective?

For all fields and measures, the relationship between value-added in the general exam and selectivity of programs is statistically the same: for each standard deviation of difference in the initial test score, the value-added decreases about 0.75 standard deviations. This means that the relation that students have with general knowledge, as measured by the test, does not vary by the field that they choose to study in college.

On the other hand, in the specific exam, the behavior of this relationship varies both by field and by measure. The bottom 50% of programs from all fields have a negative relationship between value-added and selectivity, but that does not happen in the top half of programs. For those more selective programs, each field has a particular relationship: some also have a negative relation, such as we reported in Education, and others have a somewhat positive relation between value-added and selectivity, such as Computer Science. These behaviors are found in both inter- and intra-cohort measures, and in all periods of time considered. Thus, the specific knowledge content learned by students enrolled in the less selective programs is similar to their general knowledge learning gains, but in more selective programs specific knowledge gains are greater than general knowledge gains.

C) Are quality of faculty and institutional resources correlated to higher scores?

The first year scores are the most influential in the final year test scores, meaning that one's achievement at the end of college is still very dependent on initial knowledge. Socio-economic status controls are often significant, race more frequently than mother's education. The ratio of initial to final year students and private institutions are negatively associated with test scores. Faculty information is the second most influent group of variables, with the percent of faculty with PhD degrees standing out in almost every case.

The quality of infrastructure and organization vary considerably in their influence by field, having positive effects in most cases, and negative effects in some, such as the general exam of Language Arts. When comparing the results from intercohort to those using the intracohort measure, there is no apparent pattern of differentiation. This means that overall, the results from both approaches are somehow equivalent when considering the variables with the highest influence on test scores. At the same time there are some

local variations, such as lower coefficients for initial test scores and higher for quality of infrastructure or organization in intracohort measures, when compared to intercohort results.

These findings show that the approaches for measuring value-added as well as the estimates in each field produce different behavior under the similar situations. An example is the coefficient of the percentage of full-time professors that is significant only in some cases. It may also demonstrate that the variables considered have limited explanatory power on the variation of final year test scores.

Conclusions

In summary, students do have significant learning gains, with the least selective programs having higher average gains. Despite that, for several of the fields analyzed, the final achievement levels of students are highly unequal. The differences between the selected fields occur mainly in the value-added distribution on the specific exam. The results of my study could have implications for helping students select programs and HEI, as well as for helping companies to select candidates for jobs based on their degree, because the way students accumulate human capital depends on the field. This study can also help the government more efficiently select programs that need intervention to improve quality, by having specific criteria depending on the field.