

Staff

Sarah Kabay
Program Director, Education

“On the Existence and Gravity of Ghosts”: Education Data Challenges—and a New Book!



Ten years ago, I was managing data collection for a randomized evaluation (RCT) in Ugandan primary schools, when a national newspaper reported on a discussion between two senior accounting officers for the country’s Universal Primary Education program. This discussion was “opening a new debate on the existence and gravity of ghosts.”

I was very familiar with ghosts. In school systems around the world, the term “ghost” is used to refer to pupils, teachers, and even schools that exist in records and data systems but not in reality. As many education systems provide funding based upon enrollment, there are strong financial incentives to overstate the number of children enrolled in school. Outright graft does exist, but more often this is simply a strategy to bring marginally more resources to significantly underfunded schools.

Investigating discrepancies in education data—why Malawi’s recorded repetition rate (15.5)

might be just under half the average repetition rates found in two household surveys (25.8 and 28.4 percent), for example, or why official reports on the non-state provision of education are likely to underestimate the phenomenon, can be a helpful exercise. Our understanding of these issues depends on accurate data, but also, investigating why data might be inaccurate can bring to light important underlying dynamics and sensitivities.

In the two years for which we tracked children, 24 percent of our sample transferred between schools (about half of whom transferred into private schools), and 6.7 percent dropped out of school entirely.

As I spent more time in schools, I learned that it can be difficult to track children. On average, a fifth-grade classroom in this sample of schools included 88 pupils (earlier grades are often larger.) At the beginning of each semester, it typically takes a few weeks for children to report to school. Absence rates were often around 20 percent, but might be higher during planting season, or on market days. Even though the education system had a policy of automatic promotion, the vast majority (88 percent) of our sample of students had repeated a grade, and 11 percent had repeated three or more grades. In the two years for which we tracked children, 24 percent of our sample transferred between schools (about half of whom transferred into private schools), and 6.7 percent dropped out of school entirely.

You can imagine how difficult it might be for a teacher to monitor nearly a hundred children in such a dynamic setting, or even how long it might take to simply conduct a roll call. In the wake of the school closures brought on by the coronavirus pandemic, it will be that much harder to determine who is enrolled in what school, which class, and when.

Working on this study sparked my interest in children's movement in, out of, and between schools, and also in the issues that existing data do not seem to capture. For the past several years, I've been studying some of these issues, and I've just published a book, *Access, Quality, and the Global Learning Crisis*.

The heart of the book is comprised of three empirical chapters, each investigating a particular topic: grade repetition, private primary schools, and school fees. I review the literature on each topic and then use data from a sample of Ugandan primary schools and communities to analyze a set of empirical questions. Together, these empirical questions shed light on an important theoretical concern: **is there a connection between access to education and education quality?** And in particular, **is there a tradeoff between efforts to improve access to education and education quality?** (Spoiler: there doesn't have to be! I find evidence for a possible positive association and argue that encouraging dynamics where efforts to improve access to education and education quality are mutually reinforcing can be an exciting direction for future work in education.)

The book addresses a number of different topics and themes, from complexity theory to the Global Learning Crisis, but the starting point was always a data challenge or discrepancy. Each of the three issues I studied, grade repetition, private primary schools, and school fees present data collection challenges and sensitivity. My strategy was to combine multiple forms of data, such as basic descriptive statistics and quantitative analysis, as well as qualitative

methods with multiple stakeholders, insights from newspaper articles, and ongoing debate at local, national, and international levels.

Investigation into sensitive subjects, whether that is inaccurate enrollment registers, children's work in contexts with child labor laws, or the costs associated with attending school in light of free education, takes time and thoughtful, considerate data collection. And maybe a willingness to believe in ghosts.

More information on the book can be found [here](#). You might be able to access it for free online at [Oxford's Online Access](#)—check your institutional login (such as BRAC University) or look into their Developing Country Initiative.

November 02, 2021