IMAGINE

THE FUTURE OF LEARNING

Unlocking Talent in Malawi: Analysis Plan

Our primary research questions for the efficacy study are:

1. What are the impacts over standard instruction on literacy and numeracy outcomes of using onebillion's onecourse software in Chichewa for 40 minutes per day for 8 months?
2. What impact does attendance in the intervention have on learning outcomes?
3. How far do children progress toward the ultimate learning goals of reading fluency with comprehension and comparable numeracy skills (i.e., arithmetic fluency with number sense)?
4. How are subgroup characteristics such as school, gender, and age category associated with learning outcomes?

We will collect data from three main sources: (1) baseline and endline assessments, which will provide the outcome measures for evaluating the impact of the intervention on literacy and numeracy learning; (2) application usage data, which will enable analysis of the impact of attending the intervention on learning outcomes; and (3) monthly monitoring visits, which will help to interpret treatment impacts by describing the quality and fidelity of the intervention.

Before conducting the outcome analyses, we will confirm baseline equivalence of the final analytic sample. Following What Works Clearinghouse guidelines, we will calculate standardized mean differences between the literacy treatment group and the control group, and between the numeracy treatment group and the control group, on measurable baseline characteristics including gender, age category, and baseline achievement in literacy and math. Baseline equivalence will be satisfied without the addition of covariates for differences ≤.05 and with the addition of covariates for differences >.05 and ≤.25. Standardized mean differences >.25 will not meet equivalence standards.

We will conduct our analysis in two main parts: (1) impact analysis of the intervention on learning outcomes for the overall sample and (2) exploratory analysis of the association of subgroup characteristics with learning outcomes. We will produce two sets of impact estimates: Intent-to-Treat (ITT) estimates representing the impact of being assigned to the intervention, relative to being assigned to the control group; and Treatment-on-the-Treated (TOT) estimates representing the impact of attending the intervention at least 50% of the days that the learning center was open, relative to attending the intervention fewer or no days. Attending at least

50% of the offered days is considered minimum compliance with the treatment. For the TOT analysis, we will use treatment status as the "instrumental variable" to predict attendance, and will then use predicted attendance (met threshold, did not meet threshold) as the treatment indicator.

For both impact analyses, we will follow standard practice for estimating impacts from a non-clustered, blocked individual random assignment design evaluation. Because we purposively selected the two schools in the study, and randomly assigned learners independently within each school, the schools represent independent samples. We will conduct the impact analysis as a multi-site randomized trial, averaging separately derived site-level impacts and assuming fixed site effects. To produce the average treatment effect for each outcome measure, we will estimate the treatment impact and associated effect size separately for the two schools and then average the estimates. We will calculate standard errors for the averages by pooling the site-level standard errors.

To obtain more precise school-level impact estimates, we will adjust for baseline student characteristics in an ordinary least squares regression model that uses the gain score for each outcome measure as the continuous dependent variable. We use gain scores instead of endline outcomes as the dependent variables to avoid attenuation bias due to measurement error in the baseline measures. In the regression model we will include fixed effects for the gender-age category strata as well as baseline covariates for gender, age category (6-7 years vs. 8-10 years), the relevant outcome measure, and the opposite-subject composite outcome measure. We will not make adjustments for data nonresponse (which we expect to be low).

For each site-level impact estimate, we will compute the associated effect size, which reflects the magnitude of the impact relative to the variation in the outcome measure in the sample (the treatment and control groups combined). Site-level effect sizes will be calculated using Cohen's d. When reporting impact findings, we will use the following convention based on the What Works Clearinghouse (Procedures Handbook, Version 4.0, pages 21-24) and on effect-size benchmarks proposed by Kraft (2018) for causal studies: we will use "statistically significant positive effect" if the treatment effect is both positive and statistically significant; "substantively important positive effect" if the treatment effect is not statistically significant but is positive and equal to or larger than 0.25 standard deviations; "suggesting positive effects" if the treatment effect is not statistically significant but is between 0.15 and 0.25 standard deviations; and "indeterminate effect" if the treatment effect is not statistically significant and is between -0.15 and 0.15 standard deviations. We do not anticipate treatment effects that fall below -0.15 standard deviations.