

## **Using back-end user data to observe curriculum implementation fidelity and explain treatment effects: Experimental evidence from South Africa**

**Stephen Taylor (South African Department of Basic Education)**

Global efforts to identify and roll out educational reforms based on evidence of what actually improves learning are now recognizing that the fidelity of programme implementation matters critically. More research is trying to explain observed treatment effects on the basis of implementation fidelity. Similarly, there is a growing concern that efforts to scale up programmes that worked in small experimental settings may suffer from weaker implementation when administered on a larger scale or by governments rather than NGOs. Implementation matters; so we need more research on how to conceptualise and measure implementation.

One class of interventions where implementation questions are particularly relevant is structured pedagogy programmes. These programmes provide lesson plans, learning materials and professional support to teachers, all aimed at improving curriculum implementation by teachers, and thus improve child learning. But observing the intermediate outcome of curriculum implementation is difficult. Sometimes work done by pupils in exercise books is used as a proxy for curriculum coverage; sometimes researchers observe a sample of lessons to assess fidelity to the intended pedagogy. These have their limitations, such as changes in behaviour in response to being observed.

In this paper, we exploit back-end user data from an application containing daily lesson plans provided to teachers on an electronic tablet as part of a structured pedagogy programme in South Africa. The programme was run by the South African government and the lesson plans were aligned to the official curriculum for grades 1 to 3. The data reveals how often, when and for how long teachers accessed lesson plans across 2022 and 2023. This provides a unique opportunity to gain new insights into the extent of curriculum coverage and how programme implementation varies across teachers and contexts.

There are several noteworthy findings. Firstly, curriculum implementation becomes weaker later in the week (Monday's lessons are more likely to be accessed than Friday's), later in each term (Week 1 and 2 lessons are more widely accessed than Week 9 and 10 lessons), and later in the Year (Term 1 and 2 lessons are more widely accessed than Term 4 lessons). This is in line with global evidence of over-ambitious and packed curricula, which teachers struggle to complete, especially in contexts where significant amounts of teaching time are lost due to competing activities. This pattern also implies that the technological aptitude of teachers was not a significant barrier to use, something that was confirmed by qualitative case studies.

Further insights emerge because of an experimental research design: Some schools were randomly assigned to receive on-site coaching; others were randomly assigned to receive internal curriculum implementation support from their school's Head of Department, who received training and support over the period ("HoD-led internal coaching"); a control group received no coaching (whether internal or external). But all schools, including control, received the electronic tablets with the lesson plans as well as centralized training on a quarterly basis.

The back-end app usage data shows that teachers who received on-site coaching from external specialists accessed significantly more lesson plans compared to teachers in the control group and compared to teachers in the "HoD-led internal coaching" group. This finding adds to the

literature on teacher coaching, suggesting that at least one key contribution made by coaches is simply to prompt better implementation of the learning programme being administered. Furthermore, it is found that on-site coaching had a significant impact on student learning, whereas “HoD-led internal coaching” did not. The back-end app-usage data thus provides important insights into the intermediate outcomes around curriculum implementation.

Linking the back-end app usage data to teacher questionnaire data, we then compare the patterns in the back-end data to more traditional ways of measuring implementation and curriculum coverage, such as teacher self-reports and observations of students’ workbooks. We also explore how teacher characteristics predict curriculum coverage according to the app-usage data.

Lastly, we grapple with a methodological problem plaguing many attempts to use implementation data to explain variation in treatment effects: Many experimental studies measure implementation fidelity in the treatment group only, but cannot validly identify a subset of the control group who would have implemented with high fidelity had they received the treatment. In our dataset, we observe teachers in both treatment and control all implementing the same set of lesson plans provided on an electronic tablet. Using this, we explore ways of identifying sub-groups of teachers across treatment arms who are more likely to resist or embrace such a structured learning programme. This allows us to use implementation data to explain treatment effects under more plausible identification assumptions.

Looking to the future, this paper points to promising opportunities for using new technologies, and the associated zero-cost back-end data, to better understand the role of implementation in explaining impact evaluation results.