
Psychometric considerations for remote phone-based measurement

Diego Luna Bazaldua
Education Specialist



Learning Assessment Platform



WORLD BANK GROUP
Education

Team members and acknowledgements



Julia Liberman



Victoria Levin



Aishwarya Khurana



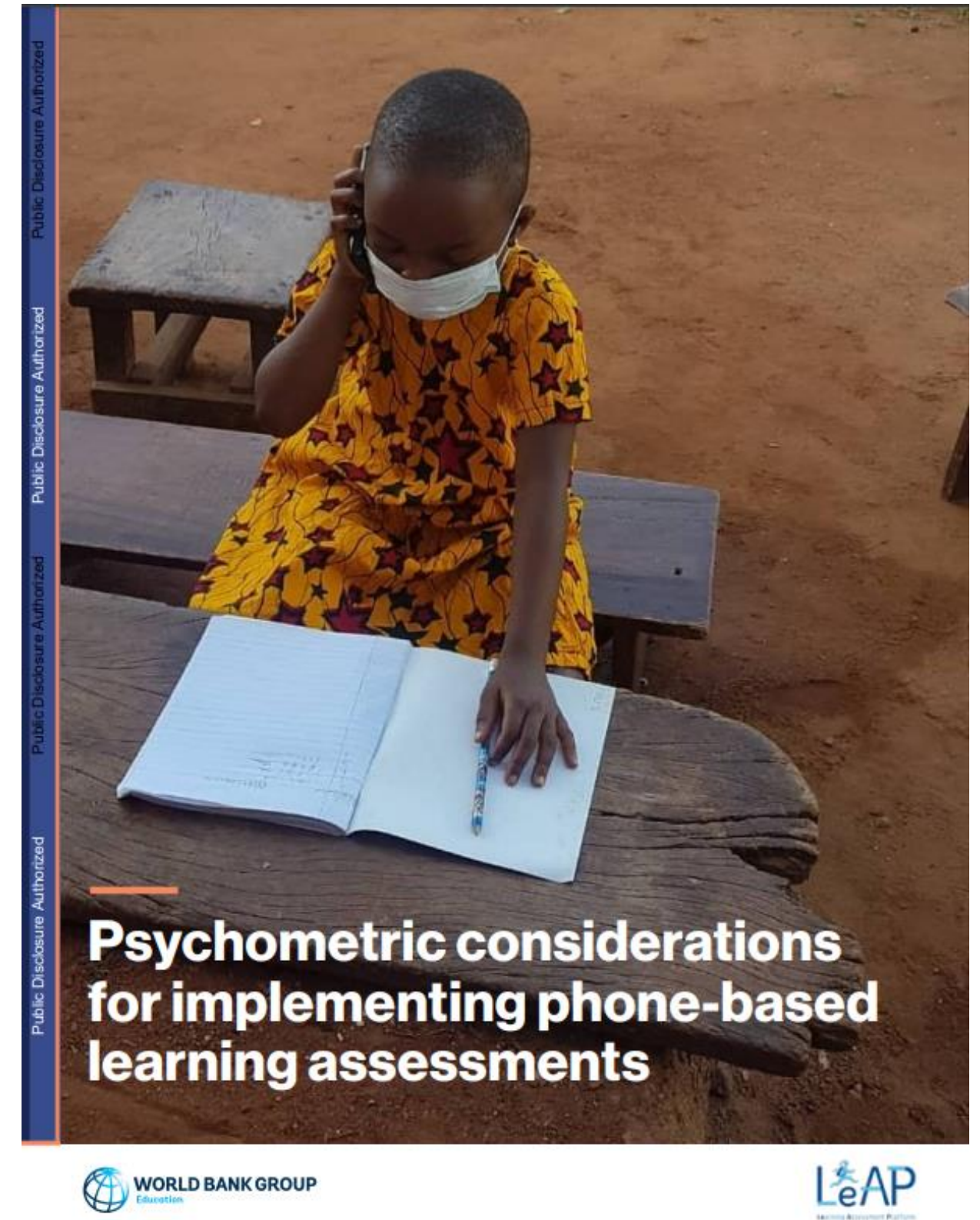
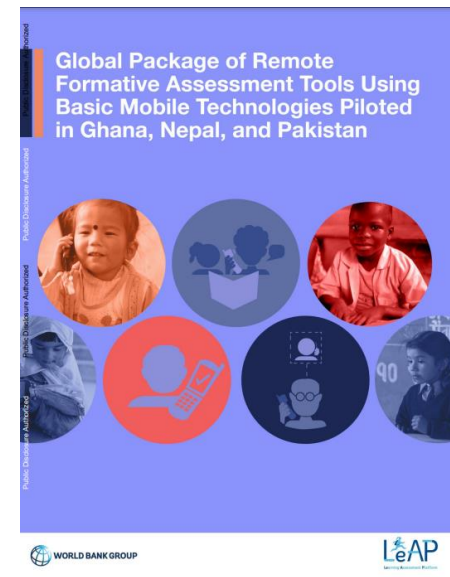
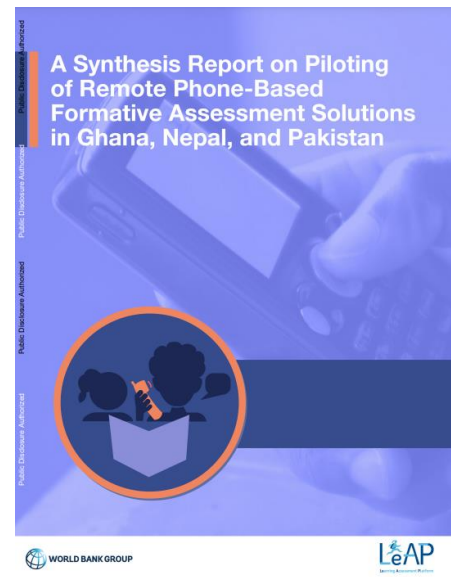
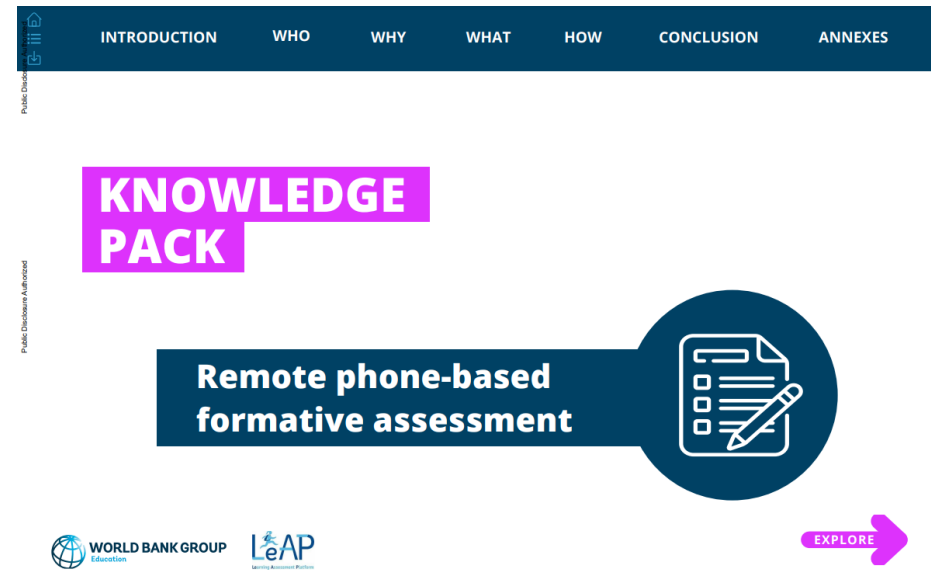
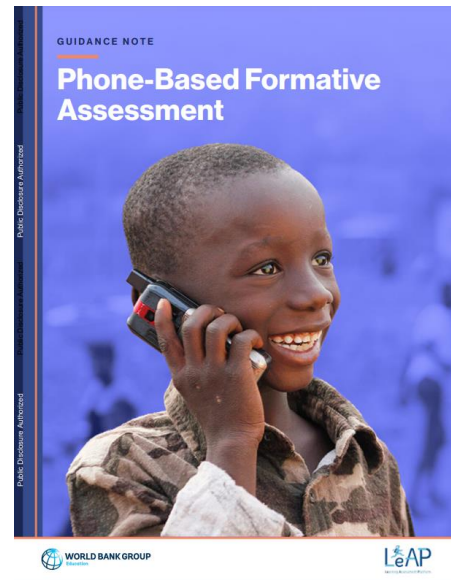
Lucy Kruske



Diego Luna-Bazaldua

This project and the global public goods developed under the Continuous and Accelerated Learning in response to COVID-19 ASA were possible with the support of the **Global Partnership for Education (GPE)** grant.

Global Public Goods



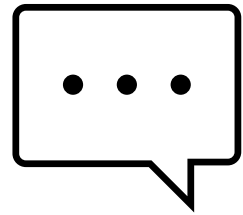
How can we know whether and what students are learning?

ASSESSMENT OF STUDENT LEARNING

The process of gathering and evaluating information on what students know, understand, and can do in order to make an informed decision about the next steps in the educational process.

- Fundamental feedback mechanism in education
- Within the context of COVID-19 disruptions, assessment needs to take place “outside the classroom box”
- Encompasses formative assessment which is critical in guiding students along their learning trajectory

Approaches to real-time formative assessment using basic (mobile) phones and their advantages

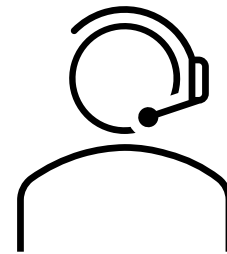


Text / SMS

SMS is cheaper than IVR in survey applications, making it the most cost-effective tool.

Users can respond at their convenience.

It works for any type of phone.



Live Phone Calls (LPC) to teachers

LPC allows a more personal approach to formative assessment as teachers interact directly with students.

Users can monitor students at a distance.

It can support students with visual disabilities and students living household with illiterate adults.



Interactive Voice Response (IVR) materials

IVR can support students with visual disabilities and students living household with illiterate adults.

Content can be delivered in multiple languages.

It ensures that students are paying attention and completing their assessments at once.

Validity-related considerations of PBAs

Validity refers to degree to which accumulate evidence support a specific interpretation and use of test scores.

1. What are the **assessment's objectives** and intended uses?
2. Is the assessment content relevant and representative of the knowledge domain or skill that is to be measured? Is there any learning content that will be omitted from the assessment if it cannot be assessed by phone?
3. Is the use of phones making it harder for students to understand and answer the items? Are adaptations of paper-based assessments necessary to reduce the **complexity** of their administration over the phone? In addition: What cognitive processes are students using to answer the questions over the phone?
4. Are the item scores showing internal coherence with each other?
5. Are the assessment scores correlated with any other external variables?





Reliability-related considerations of PBAs

Reliability refers to the degree to which test scores for a group of test takers are accurate and free of random errors of measurement.

1. Are phone-based assessment scores accurate?

- Internal consistency. Are item scores consistent among themselves when administered over the phone?
- Temporal consistency. Are phone-based assessment scores accurate over time?
- Inter-rater reliability. Are assessment scores produced by different enumerators administering the assessment over the phone consistent with each other?

Psychometric validity in PBA

Uses of learning assessment	Intended purpose	Important limitations in PBA validity
Impact Evaluation 	Analyze the impact of interventions or policies on key outcomes by comparing groups.	<ul style="list-style-type: none"> - Reliability of phone-based data collection might be lower than paper-based assessments due to logistics of assessment administration. This has implications for the validity of any IE results. - Lengthier assessments may decrease test-taker engagement and increase missing data.
System-level monitoring 	Provide feedback on overall health of system at particular grade or age level(s).	<ul style="list-style-type: none"> - System-level representative samples might not be achievable due to penetration rates of landlines and mobile phones, particularly among marginalized communities.
High-stakes decisions 	Select or certify students as they move from one level of education system to the next.	<ul style="list-style-type: none"> - The LeAP team does not recommend the use of PBAs for any high-stakes decisions on students or other stakeholders given challenges with test proctoring, standardized test administration, and potential malpractice.
Formative 	Provide immediate feedback to inform classroom instruction.	<ul style="list-style-type: none"> - Some items and tasks cannot be administered over the phone, or their administration presents logistical challenges, which limits any inferences on what students know and can do.

Limitations in PBA

- Many limitations require implementors to take action from the planning and development stages of the PBA. Statistical and psychometric analyses will not fix what by design is not optimal.

PBA stage	Important limitations in PBA validity	How to address the limitation?
1. Planning	<ul style="list-style-type: none">- The LeAP team does not recommend the use of PBAs for any high-stakes decisions on students or other stakeholders given challenges with test proctoring, standardized test administration, and potential malpractice.	<ul style="list-style-type: none">- This issue cannot be solved. By design, it is unlikely that you can prevent cheating at home with the use of PBAs.
2. Development	<ul style="list-style-type: none">- Some items and tasks cannot be administered over the phone, or their administration presents logistical challenges, which limits any inferences on what students know and can do.	<ul style="list-style-type: none">- Pilot PBA administration, improve clarity in the items and tasks based on test-taker feedback.
	<ul style="list-style-type: none">- Reliability of phone-based data collection might be lower than paper-based assessments due to logistics of assessment administration. This has implications for the validity of any IE results.	<ul style="list-style-type: none">- Conduct training of enumerators and compare their performance and accuracy with respect to master test administrator (i.e., inter-rater reliability). Provide feedback to enumerators based on their performance in mock scoring tasks.



Limitations in PBA

- Many limitations require implementors to take action from the planning and development stages of the PBA. Statistical and psychometric analyses will not fix what by design is not optimal.

PBA stage	Important limitations in PBA validity	How to address the limitation?
3. Implementation	<ul style="list-style-type: none"> - Lengthier assessments may decrease test-taker engagement and increase missing data. 	<ul style="list-style-type: none"> - After a PBA pilot, try to reduce test length by including items that contribute the most to the test reliability. - Plan PBA administration over several shorter phone calls.
4. Analysis	<ul style="list-style-type: none"> - System-level representative samples might not be achievable due to penetration rates of landlines and mobile phones, particularly among marginalized communities. 	<ul style="list-style-type: none"> - This issue cannot be solved. Rely on propensity scores or other methodologies to understand to what populations you could generalize your results.
	<ul style="list-style-type: none"> - Some items and tasks cannot be administered over the phone, or their administration presents logistical challenges, which limits any inferences on what students know and can do. 	<ul style="list-style-type: none"> - Complement the administration of the PBA with portfolio analysis, semi-structured interviews, or other sources of information to improve your understanding of their performance.

Thank you

XXXXXXXXXXXXXXXXXXXX

- XXXXXXXXXXXXXXXXXXXX
- XXXXXXXXXXXXXXXXXXXX