

# Principles for Phone-Based Assessments of Learning

**Noam Angrist, Peter Bergman, David K. Evans, Susannah Hares, Matthew C. H. Jukes, and Thato Letsomo**

## Abstract

School closures affecting more than 1.5 billion children are designed to prevent the spread of current public health risks from the COVID-19 pandemic, but they simultaneously introduce education risks as well as new, longer run health risks, via lost education. While some studies measure student involvement in educational activities during the crisis through phone-based surveys, the literature on assessing learning by phone is almost nonexistent, despite the fact that learning loss has major implications for school dropout and rising inequality. In this article, we draw on our pilot testing of phone-based assessments in Botswana, along with the existing literature on oral testing of reading and mathematics, to propose a series of preliminary principles to guide researchers and service providers as they try phone-based learning assessments. We provide guidance to help teams (1) ensure that children are not put at risk, (2) test the reliability and validity of phone-based measures, (3) use simple instructions and practice items to ensure the assessment is focused on the target skill, not general language and test-taking skills, (4) adapt the items from oral assessments that will be most effective in phone-based assessments, (5) keep assessments brief while still gathering meaningful learning data, (6) learn from the speed and confidence of responses, (7) use effective strategies to encourage respondents to pick up the phone, and (8) build rapport with adult caregivers and youth respondents.

**Keywords:** education, assessment, learning, mobile technology, COVID-19

**JEL:** I21, I25, O15

# Principles for Phone-Based Assessments of Learning

Noam Angrist  
Young 1ove and Oxford University  
noam.angrist@bsg.ox.ac.uk

Peter Bergman  
Teachers College, Columbia University  
psb2101@tc.columbia.edu

David K. Evans  
Center for Global Development  
devans@cgdev.org

Susannah Hares  
Center for Global Development  
shares@cgdev.org

Matthew C. H. Jukes  
RTI International  
mjukes@rti.org

Thato Letsomo  
Young 1ove  
tletsomo@young1ove.org

The team thanks Caton Brewster, who is leading data collection for the digital response in Botswana, Efua Bortsie, who provided digital coordination support, Colin Crossley, who co-developed digital content, colleagues from Pratham, who provided content inspiration and technical assistance, and Amina Mendez Acosta, who provided research assistance. Şeyə Abimbólá, Sebastian Fehrler, Harry Fletcher-Wood, and Alexis Le Nestour provided helpful comments on the draft.

The Center for Global Development is grateful for contributions from the Bill & Melinda Gates Foundation, Elma Relief Foundation, UBS Optimus Foundation, and Vitol Foundation in support of this work.

Noam Angrist, Peter Bergman, David K. Evans, Susannah Hares, Matthew C. H. Jukes, Thato Letsomo, 2020. "Principles for Phone-Based Assessments of Learning." CGD Working Paper 534. Washington, DC: Center for Global Development. <https://www.cgdev.org/publication/principles-phone-based-assessments-learning>

**Center for Global Development**  
**2055 L Street NW**  
**Washington, DC 20036**

202.416.4000  
(f) 202.416.4050

**www.cgdev.org**

The Center for Global Development works to reduce global poverty and improve lives through innovative economic research that drives better policy and practice by the world's top decision makers. Use and dissemination of this Working Paper is encouraged; however, reproduced copies may not be used for commercial purposes. Further usage is permitted under the terms of the Creative Commons License.

The views expressed in CGD Working Papers are those of the authors and should not be attributed to the board of directors, funders of the Center for Global Development, or the authors' respective organizations.

## **Contents**

Introduction .....	1
Principles for phone-based assessments.....	1
Conclusion.....	5
References.....	6

## **Introduction**

School closings around the world due to COVID-19—with more than 1.5 billion learners affected—pose the potential to add an education crisis and further health challenges to the pandemic.<sup>1</sup> In the short run, school closings are associated with rises in adolescent pregnancy.<sup>2</sup> School closures also lead to dropout, with adverse impacts on subsequent health behaviors and health status.<sup>3</sup> To keep students engaged and learning, education systems have rolled out a wide variety of distance learning platforms: television programs, radio programs, web-based instruction, phone tutorials from teachers, and others.<sup>4</sup> Existing studies have measured how much children are engaging with educational content.<sup>5,6</sup> But how much are they actually learning? Students commonly fall behind during school closures,<sup>7,8</sup> and that can also increase dropout rates.<sup>9</sup> Children do not lose learning equally: children from high income families gain learning during school closures, while children from low socioeconomic backgrounds lose the equivalent of several months of learning.<sup>10</sup>

Current research projects in low- and middle-income countries, where internet access can be both limited and inconsistent, seek to evaluate student learning by phone to avoid putting enumerators and youth at risk. There is a limited history of phone-based behavioral and learning assessments. Several studies have assessed the validity of phone-based assessments of cognitive function among elderly patients,<sup>11</sup> including one study of literacy assessments in adults.<sup>12</sup> Other studies have enabled community health workers to assess and report child health.<sup>13,14</sup> We are not aware of any published studies on direct learning or health assessments of children that were administered by phone. This article combines past research and experience with oral assessments, with ongoing piloting of phone-based assessments in a middle-income setting (Botswana) to propose a series of preliminary principles for the assessment of learning by phone.

In our piloting effort, Young Love—one of the largest non-government organizations in the country—worked in partnership with the Ministry of Basic Education to collect over 10,000 phone numbers in schools in four out of ten regions in Botswana before schools closed for lockdown. Since schools have been closed, caregivers and students have been contacted to participate in remote learning interventions rolled out as a randomized controlled trial in partnership with Columbia University and the Jameel Poverty Action Lab (J-PAL). In the course of that, pilot assessments of math and literacy have reached 3,000 children.<sup>15</sup>

Assessment by phone is a nascent field of research, and much will be learned in the current crisis and beyond. In addition to ongoing work in Botswana, authors of this paper are associated with efforts in Sierra Leone and Tanzania in partnership with the Center for Global Development and RTI International, and other teams in other countries are also implementing pilots. This article seeks to orient researchers and service providers in the right direction as they create new knowledge about how to assess learning by phone effectively.

## **Principles for phone-based assessments**

While phone-based assessments are little studied, oral assessments of learning are commonly used directly with children and have much to teach about phone-based assessment. Orally

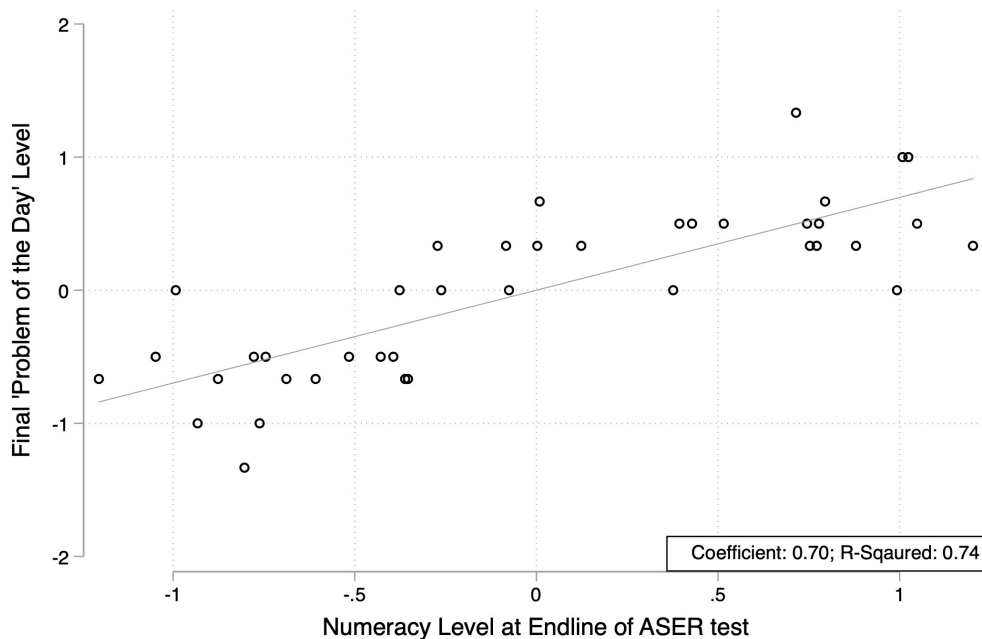
administered tests are effective. Commonly used early-grade assessments of reading and mathematics (such as ASER, Uwezo, and the Early Grade Reading and Math Assessments) are administered orally and have been validated.<sup>16-18</sup> Instructions are presented orally and the response required from the participant is also oral. These assessments are, therefore, suitable for adaptation to phone surveys. By contrast, assessments with written instructions implicitly test children's ability to read, and technology-based assessments (e.g., via the internet) test children's ability to use technology.

In adapting oral assessments for administration by telephone, here are suggestions drawn from experience and literature.

**1. Protect children.** Much has been written about best practice in phone surveys,<sup>19,20</sup> but few phone surveys gather data directly from children.<sup>5</sup> It is vital to adapt face-to-face consent procedures and enumerator training to make sure that children and youth are not put in harm's way in the process. For example, assessors can ensure that parents are aware that tests have no direct consequences for children (i.e., these are low-stakes assessments), so that adults do not discipline children if they overhear low performance. Supervisors can also monitor a sample of calls to make sure enumerators are interacting appropriately with children and youth. Researchers should adapt general principles of research with children and youth for phone-based assessment.<sup>21</sup>

**2. Test the reliability and validity of your measures.** Before rolling out an assessment, it is essential to ensure that it measures the specific skill that you want to measure - rather than, for example, general language skills - and does so reliably.<sup>22,23</sup> Fortunately, this can be done at a fraction of the cost of the overall assessment. The simplest psychometric assessment to carry out is one of internal reliability, which assesses the correlation between performance on different items intended to measure the same construct (e.g., all addition-focused items within a numeracy test). (The statistic commonly used to measure internal reliability is called Cronbach's Alpha.<sup>24</sup>) If the correlation is low, it helps to identify problematic questions. Ideally, phone-based assessments should be validated against established face-to-face assessments. Such a test of concurrent validity is not possible while schools are closed and communities are locked down. An approximation to concurrent validity may be possible where assessments on the same sample using similar tools have recently been conducted and data are available. For example, in Botswana, we compared simple 'problems-of-the-day' (e.g., *what is  $25 \times 4$ ?*), which are more easily adapted to a phone-based assessment, and more comprehensive assessments, such as the ASER assessment.<sup>16</sup> Both were conducted in school between February and March of 2020 with students in grades 3-5. The problems-of-the-day were administered daily for a period of 15 days, recorded in individual student booklets, and compiled by student and class. We correlated the difficulty level of the final problem-of-the-day with the comprehensive assessment taken shortly after, and found a correlation of 69 percent. We further find a high R-squared of 0.74 and average relationship estimated by a multivariate regression of 0.70 when we control for school-level variation (Figure 1). If replications demonstrated this relationship to be stable in the study population, it would represent reasonable concurrent validity.

**Figure 1. The relationship between student answers on “problem of the day” on the last day of class and average learning levels for the whole class after 15 days.**



Estimates were averaged at the class level within a school for a sample of 40 classes. Each individual student answered a ‘problem of the day’ in an individual booklet which was compiled by the class teacher. If students answered problems correctly, they progressed to more difficult items. At the end of 15 days, a more comprehensive multi-item oral assessment (ASER) was administered. In this figure, we compare the final problem of the day level of difficulty with performance on the ASER test.

**3. Keep instructions simple and use practice items to ensure that respondents understand the exercise.** As with other oral assessments, whatever you evaluate by phone bundles receptive language skills with the skill you are attempting to test. By phone, in the absence of visual cues, oral assessments are even more of a test of receptive language skills like vocabulary, listening, and processing skills. Acknowledging and adjusting for this is particularly important in settings where different respondents speak different languages and may comprehend the language of instruction orally either better or worse than they read it. Simple instructions and practice items can ensure that more of the assessment is focused on the target skill.

**4. Some assessments will be more conducive to phone assessment than others.** The elements of oral tests with minimal visual stimuli will be easiest to adapt to phone-based testing. For example, the ‘word problems’ subtest of the Early Grade Math Assessment involves only oral stimuli, whereas the ‘missing number’ subtest has a grid of numbers that may be hard to replicate on a phone display. That said, phone-based assessments can still incorporate text. Assessors could send a text message and ask the respondent to read the message aloud. In Botswana, we have tried sending simple texts for students in grades 3 through 5, such as “Katlego has 32 apples and organizes them by PLACE VALUE. How

many TENS does she have?” We asked the child to read the problem out loud (assessing literacy skills), and then ask the child to solve the problem (assessing mathematics skills). We send the text message immediately before the phone call.

**5. Keep it short.** Home environments, particularly during lockdowns, may be crowded and noisy and phone calls can be frequently interrupted. Brief calls and assessments are more effective than longer calls. General guidance on conducting phone surveys suggests keeping them to thirty minutes.<sup>19</sup> However, assessments with young people should be shorter. The Early Grade Reading Assessments, a text-based but orally administered assessment, typically takes about 15 minutes per child.<sup>25</sup> Pilot calls in Botswana are taking between 10 and 15 minutes.

Although full-length assessments may not be possible to conduct via phone, micro assessments that are high-frequency, simple and cheap can still be informative and easily conducted over the phone. With shorter assessments, if teams want estimates with the same level of statistical precision across the assessed students, the sample size will need to rise.<sup>26</sup> In face-to-face assessments in Botswana earlier this year—as described above—we observed a strong correlation between performance on the single “problem of the day” and performance on the class assessment (Figure 1). Although this was a face-to-face assessment, it demonstrates how simple tests like these, administered by phone, could indicate levels of learning loss or gain and therefore provide useful information for policy makers and schools systems attempting to mitigate the adverse health and educational effects of school closures.

**6. Track the speed of responses.** In existing oral assessments of both reading and mathematics, it is common to track whether the students immediately answer the question correctly or if they answer it correctly after first making a mistake, as this is one indicator of content mastery.<sup>27</sup> As a complement, our piloting in Botswana suggests that quick answers are much more likely to indicate solo work, whereas slow answers may indicate that others are prompting the respondent. However, note that slower responses may also indicate other interference, such as additional household noise or a poor connection inhibiting the respondent from hearing the question clearly. Furthermore, in some cultures being careful and accurate is prized over being quick, and there is evidence that children who are more patient do less well in timed tests.<sup>28</sup>

**7. Experiment with how to get people to pick up the phone.** Piloting in Botswana revealed the combination of a text message followed by a call yielded the highest pick-up rates. This is consistent with evidence elsewhere. Sending a text message to alert respondents to an upcoming call delivered the best responses in India.<sup>29</sup> A program in Liberia sent a text message five minutes before the call and found it helpful to boost answer rates.<sup>13</sup> In Botswana, few people replied to texts alone, and about 70 percent answered calls alone. Thus, a combination of the two may be most effective.

**8. Establish rapport with adult phone owners and youth respondents.** Respondents--both adults and youth--are often nervous as this is a time of global and local crisis. In many low- and middle-income countries, conversations between adults and children are less common, as are interactions with strangers.<sup>30</sup> Questioning oriented to children in some

cultures is predominantly to obtain information that you are lacking, rather than to test the knowledge of another person. Rapport, explanation and examples can all help overcome these barriers. Having an advance call with an adult responsible for the target child can increase accuracy, honesty, and a willingness to participate (beyond the obvious need for consent).

In some cases, initial assessment instructions can be delivered through a caregiver with requests to put the child at ease. It is likely that phone-based tests will be challenging with children in early grades of primary school or in preschool.

## **Conclusion**

Efforts to assess learning by phone are still new and so should not be used for high-stakes decisions around the future of individual students. However, understanding whether distance learning efforts are leading to learning, and identifying which groups of children are being most or least disadvantaged by being out of school, should be a central part of the current response as well as any initiatives to help disadvantaged groups catch up once schools do reopen. In the longer term, if consistently reliable methods and tools to measure learning by phone can be developed, they have the potential to disrupt the way we do measure learning, by enabling both high frequency diagnostics and more cost-effective ways to assess learning outcomes. Ensuring that children are learning, even when out of school, is crucial to their education, but also to their health outcomes and the quality of their whole lives.



## References

1. UNESCO. COVID-19 Educational Disruption and Response [Internet]. 2020 [cited 2020 May 27]. Available from: <https://en.unesco.org/covid19/educationresponse>
2. Bandiera O, Buehren N, Goldstein MP, Rasul I, Smurra A. The Economic Lives of Young Women in the Time of Ebola : Lessons from an Empowerment Program [Internet]. The World Bank; 2019 Feb [cited 2020 May 4] p. 1–80. Report No.: WPS8760. Available from: <http://documents.worldbank.org/curated/en/452451551361923106/The-Economic-Lives-of-Young-Women-in-the-Time-of-Ebola-Lessons-from-an-Empowerment-Program>
3. Cutler DM, Lleras-Muney A. Education and Health: Evaluating Theories and Evidence [Internet]. National Bureau of Economic Research; 2006 Jul [cited 2020 May 27]. Report No.: 12352. Available from: <http://www.nber.org/papers/w12352>
4. Center for Global Development. CGD - COVID education policy tracking [Internet]. [cited 2020 May 27]. Available from: <http://www.cgdev.org/covid-education-policy>.
5. Asanov I, Flores F, Mckenzie DJ, Mensmann M, Schulte M. Remote-learning, Time-Use, and Mental Health of Ecuadorian High-School Students during the COVID-19 Quarantine [Internet]. The World Bank; 2020 May [cited 2020 May 27] p. 1–25. Report No.: WPS9252. Available from: <http://documents.worldbank.org/curated/en/328261589899308503/Remote-learning-Time-Use-and-Mental-Health-of-Ecuadorian-High-School-Studentsduring-the-COVID-19-Quarantine>
6. Le Nestour A, Moscoviz L. Five Findings from a New Phone Survey in Senegal [Internet]. Center For Global Development blog. 2020 [cited 2020 May 27]. Available from: <https://www.cgdev.org/blog/five-findings-new-phone-survey-senegal>
7. Quinn DM, Polikoff M. Summer learning loss: What is it, and what can we do about it? [Internet]. Brookings blog. 2017 [cited 2020 May 27]. Available from: <https://www.brookings.edu/research/summer-learning-loss-what-is-it-and-what-can-we-do-about-it/>
8. Slade TS, Piper B, Kaunda Z, King S, Ibrahim H. Is ‘summer’ reading loss universal? Using ongoing literacy assessment in Malawi to estimate the loss from grade-transition breaks: Research in Comparative and International Education [Internet]. 2017 Dec 26 [cited 2020 May 27]; Available from: <https://journals.sagepub.com/doi/10.1177/1745499917740657>
9. Zuilkowski SS, Jukes MCH, Dubeck MM. “I failed, no matter how hard I tried”: A mixed-methods study of the role of achievement in primary school dropout in rural Kenya. International Journal of Educational Development. 2016 Sep 1;50:100–7.
10. Busso M, Munoz JC. Pandemic and Inequality: How Much Human Capital Is Lost When Schools Close? [Internet]. Ideas Matter blog. Inter-American Development Bank. 2020 [cited 2020 May 27]. Available from: <https://blogs.iadb.org/ideas-matter/en/pandemic-and-inequality-how-much-human-capital-is-lost-when-schools-close/>

11. Rapp SR, Legault C, Espeland MA, Resnick SM, Hogan PE, Coker LH, et al. Validation of a Cognitive Assessment Battery Administered over the Telephone. *Journal of the American Geriatrics Society*. 2012;60(9):1616–23.
12. Sticht TG, Hofstetter CR, Hofstetter CH. Assessing Adult Literacy by Telephone. *Journal of Literacy Research*. 1996 Dec 1;28(4):525–59.
13. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: Systematic review and meta-analysis. *J Glob Health [Internet]*. 2016 [cited 2020 May 27];6(1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4643860/>
14. Hazel E, Amouzou A, Park L, Banda B, Chimuna T, Guenther T, et al. Real-time assessments of the strength of program implementation for community case management of childhood illness: Validation of a mobile phone-based method in Malawi. *The American journal of tropical medicine and hygiene*. 2015 Mar 1;92(3):660–5.
15. Teachers College, Columbia University Institutional Review Board. Young love’s activities of learning interventions and assessment (20-299 protocol).
16. Vagh SB. Validating the ASER Testing Tools: Comparisons with Reading Fluency Measures and the Read India Measures [Internet]. ASER Centre; 2012. Available from: [https://img.asercentre.org/docs/Aser%20survey/Tools%20validating\\_the\\_aser\\_testing\\_tools\\_oct\\_2012\\_2.pdf](https://img.asercentre.org/docs/Aser%20survey/Tools%20validating_the_aser_testing_tools_oct_2012_2.pdf)
17. Uwezo Kenya. Are Our Children Learning? Annual Learning Assessment Report [Internet]. Twaweza; 2012. Available from: <https://www.twaweza.org/uploads/files/UwezoKE-ALAREport2012.pdf>
18. Dubeck MM, Gove A. The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*. 2015 Jan 1;40:315–22.
19. Kopper S, Sautmann A. Best practices for conducting phone surveys | The Abdul Latif Jameel Poverty Action Lab [Internet]. Abdul Latif Jameel Poverty Action Lab (J-PAL) blog. [cited 2020 May 27]. Available from: <https://www.povertyactionlab.org/blog/3-20-20/best-practices-conducting-phone-survey>
20. Hughes S, Velyvis K. Tips to Quickly Switch from Face-to-Face to Home-Based Telephone Interviewing [Internet]. *Mathematica*. 2020 [cited 2020 May 27]. Available from: <https://www.mathematica.org/commentary/tips-to-quickly-switch-from-face-to-face-to-home-based-telephone-interviewing>
21. Alderson P, Morrow V. *The Ethics of Research with Children and Young People: A Practical Handbook - SAGE Research Methods* [Internet]. 2011 [cited 2020 May 27]. Available from: <http://methods.sagepub.com/book/ethics-of-research-with-children-and-young-people>
22. Brennan RL. *Educational Measurement, Fourth Edition*. Rowman & Littlefield Publishers; 2006.
23. Coaley K. *An Introduction to Psychological Assessment and Psychometrics*. Second edition. SAGE Publishing; 2014.
24. Bland JM, Altman DG. Statistics notes: Cronbach’s alpha. *BMJ*. 1997 Feb 22;314(7080):572.

25. USAID. Early Grade Reading Assessment [Internet]. 2011. Available from: [https://www.globalreadingnetwork.net/sites/default/files/eddata/EGRA\\_FAQs\\_25Oct11.pdf](https://www.globalreadingnetwork.net/sites/default/files/eddata/EGRA_FAQs_25Oct11.pdf)
26. Beckstead JW. On measurements and their quality: Paper 2: Random measurement error and the power of statistical tests. *International Journal of Nursing Studies*. 2013 Oct 1;50(10):1416–22.
27. Banerjee AV, Bhattacharjee S, Chattopadhyay R, Ganimian A. The Untapped Math Skills of Working Children in India: Evidence, Possible Explanations, and Implications [Internet]. New York University; 2017. Available from: <https://static1.squarespace.com/static/5990cfd52994ca797742fae9/t/59a896aee6f2e11b76983238/1504220847338/Banerjee+et+al.+2017+-+2017-08-17.pdf>
28. Jukes MCH, Zuilkowski SS, Grigorenko EL. Do Schooling and Urban Residence Develop Cognitive Skills at the Expense of Social Responsibility? A Study of Adolescents in the Gambia, West Africa: *Journal of Cross-Cultural Psychology* [Internet]. 2017 Dec 19 [cited 2020 May 27]; Available from: <https://journals.sagepub.com/doi/10.1177/0022022117741989>
29. Kasy M, Sautmann A. Adaptive Treatment Assignment in Experiments for Policy Choice [Internet]. 2019 [cited 2020 May 27]. Available from: Available at SSRN: <https://ssrn.com/abstract=3434834>
30. Lancy DF. *The Anthropology of Childhood: Cherubs, Chattel, Changelings* [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2014 [cited 2020 May 27]. Available from: <https://www.cambridge.org/core/books/anthropology-of-childhood/B34D307F81527FC3C91AE9D0B02D48D7>