Practical Lessons 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 new short- and long-term health risks through lost education. Measuring these effects in real-time is critical to inform effective public health responses, and remote phone-based approaches are one of the only viable options with extreme social distancing in place. However, both the health and education literature are sparse on guidance for phone-based assessments. 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 practical lessons to guide researchers and service providers as they try phone-based learning assessments. We provide preliminary evidence that phone-based assessments can accurately capture basic numeracy skills. 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) use effective strategies to encourage respondents to pick up the phone, (7) build rapport with adult caregivers and youth respondents, (8) choose the most cost-effective medium, and (9) account for potential bias in samples.

Center for Global Development
www.cgdev.org

Working Paper 534
July 2020*

Practical Lessons 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

Contents

Summary box

- Assessing children and youth remotely is essential to mitigating the adverse short and long-term public health and education impacts of the COVID-19 pandemic, as well as future school closures due to health and other crises.

- There is existing literature on best-practice strategies to carry out phone-based surveys to adults, on oral face-to-face testing of learning among children and youth, and on using technology to help community health workers identify ill or at-risk children. However, there is little evidence on assessing learning among children and youth over the phone.

- Pilot experience with phone-based testing among our team, together with experience with oral assessments and phone-based surveys, provide preliminary guidance to orient those who would assess learning for out-of-school children when face-to-face assessments pose a public health risk.

Introduction

School closures around the world due to COVID-19—with more than 1.5 billion learners affected—pose the potential to add a second public health challenge to the pandemic.[1] In the short run, school closures are associated with rises in adolescent pregnancy.[2] School closures also lead to dropout, with adverse impacts on subsequent health behaviors and health status.[3] 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.[4] Existing studies have measured how much children are engaging with educational content.[5,6] But how much are they actually learning? Students commonly fall behind during school closures,[7,8] and that can also increase dropout rates.[9] 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.[10]

Ongoing research projects in low- and middle-income countries, where internet access can be both limited and inconsistent, seek to evaluate student learning by phone during the COVID-19 school closures to avoid putting assessors 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,[11] including one study of literacy assessments in adults.[12] Other studies have enabled community health workers to assess and report child health.[13,14] 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.

Assessment by phone is a nascent field of research, and much will be learned in the current crisis and beyond.[15,16] 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. Not all learning can be assessed by phone; understanding which domains of learning can be assessed with validity and improving the quality of these assessments may open the door to more cost-effective measurement of student learning even after schools re-open. This article seeks to integrate principles from existing literature on face-to-face assessments with findings from a pilot study of phone-based assessments in Botswana to propose an initial set of guidance on which future research can build. These lessons also have applications to assessing child health by phone.

Pilot phone-based and complementary assessments in Botswana

In our piloting effort, Young 1ove—one of the largest non-government organizations in Botswana—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 Abdul Latif Jameel Poverty Action Lab (J-PAL).[17]

In this paper, we draw on two assessments from our work in Botswana, a phone-based assessment and a face-to-face assessment (from before schools shut down). The phone-based assessments were administered to 2,250 students who were in grades 3-5 before schools shut down. They were conducted by over 70 former teacher aides (which we will subsequently refer to as "assessors") who call households directly. Training for all assessors was conducted using voice notes and sharing of written material via WhatsApp. During the survey, assessors call a parent and request to speak to the child at the household. They request that the parent provide the child with privacy and make clear the questions are "low-stakes" in that they have no reward and consequences associated, in order to facilitate honest responses. Numeracy questions are then read out loud in order of difficulty of operation: addition, subtraction, multiplication, and division. Word problems are texted and the child is asked to read them out loud. The questions are each timed with a maximum of two minutes to ensure uniformity across assessments. Finally, the child is asked to explain their work to ensure they understand and provide another measure of independent response.

We complement our data from these phone assessments with data from face-to-face assessments, also collected in Botswana, but before schools were shut down for the pandemic. This included two assessments, both conducted in schools between February and March of 2020 with 1,080 students in grades 3-5. In the first, classroom teachers evaluated a "problem of the day." Specifically, one problem would be assigned to the whole class and

students sit on the floor at arms-length from their classmates to ensure they respond individually. The teacher reads a problem out loud or puts it on the board. Students proceed to write the problem and their response in an individual booklet. When they are finished, they raise their hand and the teacher collects the booklet. After class, the teacher flips through all booklets and marks whether the problem was correct in a scoring sheet, as well as the type of problem using a defined scheme (e.g., addition, subtraction; 1 or 2-digit; with or without borrowing, etc.) as well as a subjective rate for the level of difficulty of the problem. 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. To the same sample of students, we administered the more comprehensive Annual Status of Education (ASER) test of numeracy.[16]

Figure 1. Percentage of students reaching each level of question difficulty (no operations, addition, subtraction, multiplication, division) in the phone-based sample and in the face-to-face ASER test of the same content.
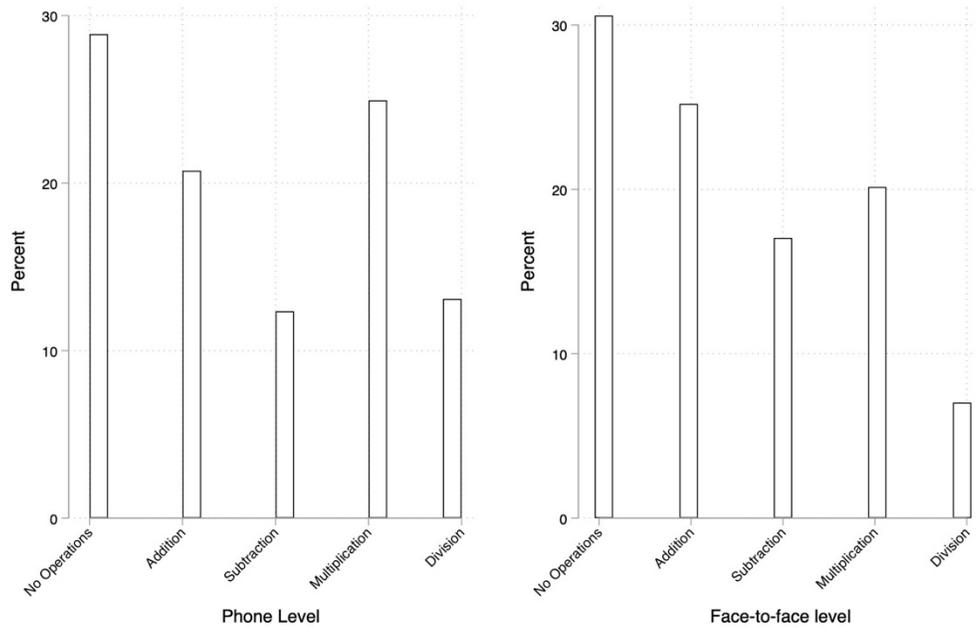
These graphs are for the same regions and largely the same set of schools and grades, but they are not matched to the exact same cohort of students. They reveal a similar distribution of learning levels using phone and face-to-face assessments at the population level in similar geographies and ages, and they increase our confidence in phone assessment. However, this is not yet a formal validity assessment. We plan to implement that in future phone-based assessments.

These data inform our preliminary practical lessons for future phone-based assessments. Comparing a sample of phone-based assessments with the more comprehensive ASER test demonstrates the promise of phone-based assessments for assessing basic skills. A sample of students from the phone-based assessment of numeracy reflect a similar skill level to the sample of students from the more comprehensive ASER assessment, administered face-to-

face before schools shut down (Figure 1).[18] Moreover, we observe that responses cover the entire distribution of skills from not being able to do any operations to being able to do division, suggesting that even with simple operations, we can capture an array of student ability.

Practical lessons 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.[19–22] Instructions are presented orally and the response required from the participant is also oral. Some aspects of these assessments are, therefore, suitable for adaptation to phone surveys.

Conducting valid assessments by phone also presents challenges. We aimed to address some of these challenges in adapting oral assessments for administration by telephone, with suggestions drawn from experience and literature. We developed these lessons through an iterative process, in which team members shared their ongoing experiences with phone-based assessments and previous experience with and literature on oral assessments of learning to identify suggestions that we would recommend to any team beginning the process of phone-based assessments.

1. Protect children. Much has been written about best practice in phone surveys,[23,24] but few phone surveys gather data directly from children.[5] 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 assessors are interacting appropriately with children and youth. One way to accomplish this is to record a random sample of calls and have those automatically sent to supervisors. Furthermore, for assessments with young children, since the phones usually belong to parents who receive the call, there is almost always another person besides the child in the household who is aware of the assessment to provide a layer of accountability and oversight. Researchers should adapt general principles of research with children and youth for phone-based assessment.[25]

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.[26,27] Fortunately, this can be done at a fraction of the cost of the overall assessment. The simplest psychometric assessments examine the internal structure of the phone-based assessment, to examine the internal reliability (Cronbach's Alpha[28]), factor structure or item analysis, for example using item response theory models. Such analyses help support the reliability of the tool and can 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.

In Botswana, we approached current validity in a two-step process. The first step was to simplify an existing oral assessment in preparation for administration by phone. In this first stage, we assessed whether the simpler version of the face-to-face administered assessment was a valid proxy for the more comprehensive test. Specifically, we compared the "problem-of-the-day" assessment, which is easily adapted to a phone-based assessment, with the more comprehensive ASER assessment.[19] 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 2). If replications demonstrated this relationship to be stable in the study population, it would represent reasonable concurrent validity, a first step towards establishing overall construct validity for the test.[29] The second stage of validation will be to test the concurrent validity of the phone-based assessments against face-to-face assessments.

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. Data from our first wave of phone-based results reveals that over 75 percent of students understood and answered all the problems, and 24 percent understood all the problems but could only answer some. (Whether or not the student "understands" the problem is a subjective judgment made by the assessor, and so is less objective than the rate of correct answers.) The alignment of skill levels across the phone-based assessment and the face-to-face assessment (Figure 1) suggests that with simple problems, like the arithmetic operations we are using, phone-based assessments can accurately capture student ability.

Figure 2. 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.
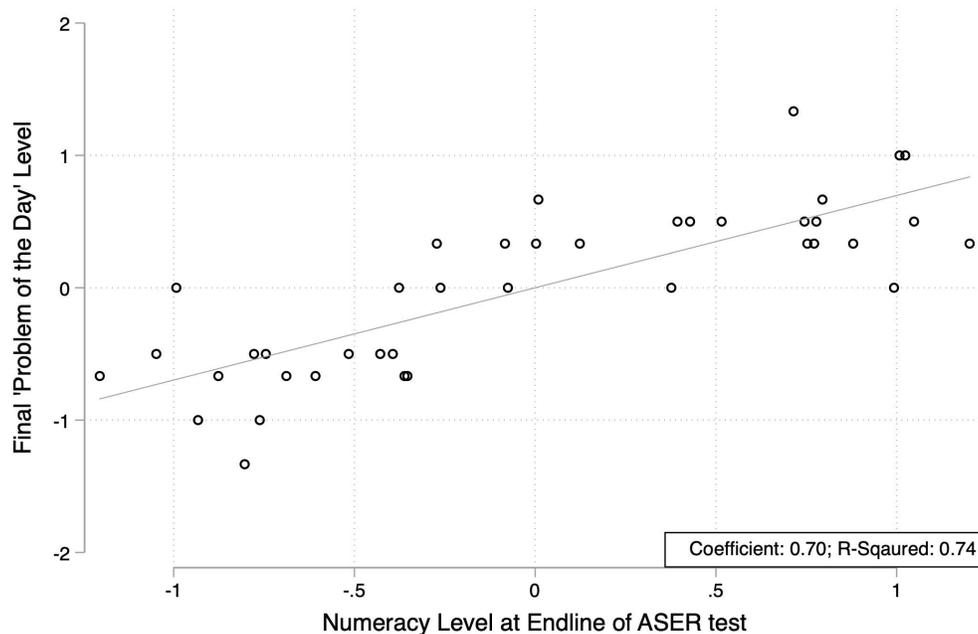
Estimates were averaged at the class level within a school for a sample of 40 classes. Each individual student answered a 'problem of the day' in an individual booklet which was compiled by the class teacher. If students answered problems correctly, they progressed to more difficult items. At the end 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.

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.[23] 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 when administered in full.[30] Some evaluations have used a

shorter version of EGRA, focusing on only 3 subtests.[31,32] Calls for phone-based assessments in Botswana are taking between 15 and 20 minutes. About 50 percent of those calls are logistical (scheduling, organizing the set-up of the house, and building rapport), and the other 50 percent are dedicated to the assessment. Obviously, the best data on this would derive from a series of tests of assessments of different lengths: in the absence of that, our experience may inform other teams in designing their assessments.

Although lengthy assessments may not be possible to conduct via phone, short 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.[33] 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 2). 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. 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.[34] A program in Liberia sent a text message five minutes before the call and found it helpful to boost answer rates.[13] 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.

7. Establish rapport with adult phone owners and youth respondents. Respondents—both adults and youth—may be nervous, particularly during this 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.[35] 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.

8. Choose the most cost-effective approach. The full cost of phone calls to over 2,250 households was about US$10,000, including airtime, personnel time, questionnaire design, and piloting. This equates to about US$4.40 per child. To put this cost in context, the international assessment Progress in International Reading Literacy Study (PIRLS), which included Botswana in 2011, has standard fees for country participation. In Botswana, the costs are around US$250,000 and about 4,000 students participated, yielding a cost of about US$62.5 per child. This is likely a lower bound, as country participation fees likely do not capture all costs. An average of school-based testing (for students in school) and home-

based tracking (for students out of school), combined with classroom observations, as part of a randomized controlled trial in Liberia, cost US$150 per child.[36,37]

We use direct phone calls by assessors since access to phones is nearly universal and are a common denominator that is widely applicable across contexts. Another, potentially lower cost approach is the use of interactive voice response (IVR) calls, but they may have context-specific capability depending on the provider landscape. In Botswana IVR infrastructure was not readily available. Lessons from direct calls reveal that about 50 percent of calling time is spent on logistics, including scheduling with parents and students, re-scheduling, setting up at the household for the assessment, and creating a conducive environment. This might imply that methods such as IVR, which might be cheaper and more scalable, might also have lower take-up rate since they might be harder to schedule reliably and may be less personal. Alternative methods and their relative cost-effectiveness are an empirical question for future work.

We hope this piece motivates further creative low-tech and cost-effective approaches to assessment. 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.

9. Account for sample bias. A challenge to conducting phone-based assessments is that there may be systematic biases in sample selection. While access to phones is nearly universal in Botswana, it is likely that households who do not respond to phone surveys differ from those who do. For example, non-responders may lack access to a phone or may live in a crowded household where it is difficult to speak quietly on the phone. The problem of sample bias applies both for validating phone-based assessments (via face-to-face assessments) as well as for data collection. Concurrent validity assessments may be flawed if a significant proportion of the face-to-face sample do not respond to phone-based assessments. The first approach to this problem is to document the bias. If socioeconomic indices are available for participating households, these data can be used to understand how responders and non-responders differ. If bias is a concern, a sample of non-responders can be selected for follow-up using different assessment methods (e.g., asking a neighbor to lend them their phone) and the data from this sub-group could be weighted accordingly in final analyses.[38]

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. That said, researchers should always remember that just as face-to-face assessments have limitations (high cost), so do phone-based assessments (more difficult to assess children with certain disabilities, like hearing loss). We have proposed an initial set of

practical lessons for phone-based assessments based on literature and the experience of piloting a phone-based assessment in Botswana. We are continuing to learn in our own practice: for example, we are experimenting with randomizing assessors to avoid any systematic enumerator fixed effects and with randomizing the set of questions posed to each child to measure both the reliability of constructs and to back out a measure of sampling error. Our hope is that future research will employ, critique and validate these lessons and contribute to a communal effort to develop best practices in this area. 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 Studentsduring 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. Romero C, Ventura S, Bra P de. Using mobile and web-based computerized tests to evaluate university students. Computer Applications in Engineering Education. 2009;17(4):435–47.

16. Sahin F. Using Mobile Phones for Educational Assessment. In: Encyclopedia of Mobile Phone Behavior [Internet]. 2015 [cited 2020 Jul 6]. Available from: www.igi-global.com/chapter/using-mobile-phones-for-educational-assessment/130132

17. Teachers College, Columbia University Institutional Review Board. Young 1ove's activities of learning interventions and assessment (20-299 protocol).

18. Angrist N, Bergman P, Brewster C, Matsheng M. Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana. 2020.

19. 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

20. Uwezo Kenya. Are Our Children Learning? Annual Learning Assessment Report [Internet]. Twaweza; 2012. Available from: https://ww.w.twaweza.org/uploads/files/UwezoKE-ALAReport2012.pdf

21. 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.

22. Reubens A. Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children [Internet]. USAID; 2009 [cited 2020 Jul 6]. Available from: https://pdf.usaid.gov/pdf_docs/Pnads439.pdf

23. 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-surveys

24. 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

25. 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

26. Brennan RL. Educational Measurement, Fourth Edition. Rowman & Littlefield Publishers; 2006.

27. Coaley K. An Introduction to Psychological Assessment and Psychometrics. Second edition. SAGE Publishing; 2014.

28. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. BMJ. 1997 Feb 22;314(7080):572.

29. Sackett PR, Lievens F, Berry CM, Landers RN. A cautionary note on the effects of range restriction on predictor intercorrelations. J Appl Psychol. 2007 Mar;92(2):538–44.

30. USAID. Early Grade Reading Assessment [Internet]. 2011. Available from: https://www.globalreadingnetwork.net/sites/default/files/eddata/EGRA_FAQs_25Oct11.pdf

31. Blimpo MP, Evans D, Lahire N. Parental human capital and effective school management : evidence from The Gambia [Internet]. The World Bank; 2015 Apr [cited 2020 May 13] p. 1–56. Report No.: WPS7238. Available from: http://documents.worldbank.org/curated/en/923441468191341801/Parental-human-capital-and-effective-school-management-evidence-from-The-Gambia

32. Sabarwal S, Evans D, Marshak A. The permanent input hypothesis : the case of textbooks and (no) student learning in Sierra Leone [Internet]. The World Bank; 2014 Sep [cited 2020 May 13] p. 1–37. Report No.: WPS7021. Available from: http://documents.worldbank.org/curated/en/806291468299200683/The-permanent-input-hypothesis-the-case-of-textbooks-and-no-student-learning-in-Sierra-Leone

33. 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.

34. 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

35. 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

36. Romero M, Sandefur J, Sandholtz WA. Outsourcing Education: Experimental Evidence from Liberia. American Economic Review. 2020 Feb;110(2):364–400.

37. Romero M, Sandefur J. Beyond Short-term Learning Gains: The Impact of Outsourcing Schools in Liberia after Three Years [Internet]. Center For Global Development. 2019 [cited 2020 Jun 11]. Available from: https://www.cgdev.org/publication/beyond-short-term-learning-gains-impact-outsourcing-schools-liberia-after-three-years

38. Baird S, Hamory J, Miguel E. Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1). 2008 Aug 1 [cited 2020 Jul 6]; Available from: https://escholarship.org/uc/item/3cw7p1hx