EOL Policy Brief Series 1: The Implications of Remote Assessment

Introduction

The Covid-19 pandemic related travel and movement restrictions limited access to communities and from the perspective of social accountability in education, created an imperative for remote assessment. Street Child Nepal had carried out remote assessments as a part of its evaluation of educational interventions during the pandemic. Since assessment is key to Education Out Loud's (EOL) objective of advocacy and community mobilisation centred around educational outcomes, a pilot of a citizen led assessment has been carried out remotely using mobile phones by Street Child Nepal to assess the effectiveness of this method. This brief considers the implications of moving assessment online, given the possibility of further restrictions on movement and accessing communities.

Pedagogical Implications

The 3 sections below consider psychological-pedagogical aspects of assessment which could be impacted as the mode of assessment changes from in-person to remote (phone-based).

Increasing Cognitive Load:

As per the theory of cognitive load, in a remote assessment carried out on the phone, it can be hypothesised that cognitive load may be higher. This is a hypothesis conditional on the assumption that during in person testing, best possible testing conditions would be better than best possible testing conditions on the phone. If so, this can lead to higher cognitive load and lower working memory(Sweller et al., 1998). As a result, the working memory that is available to solve the mathematical problem or decoding the letters / abugidas for reading is lower(Sweller et al., 1998). This may lead to a higher degree of incorrect responses from children.

Examples of features, present only in online assessment, that can contribute to higher cognitive load, are paragraphs and stories for reading which are sent across multiple text messages. This may lead to 'noise' where the cognitive load of the child will increase by having to watch a parent navigate the interface of the phone, as well as other notifications that emerge in the background in the case of a smartphone. A mitigation strategy could involve ensuring the text to be read can fit into one text message. However, this character limitation would have implications on the paragraph / story component of the assessment.

Effect on Interest

We must also consider the implications of assessing remotely on the interest of the child. Interest is defined as '*the psychological state of engaging or the predisposition to reengage with particular classes of objects, events or ideas over time.'* (Hidi & Ann Renninger, 2006, p. 112). Interest has an affective component which interacts with a cognitive component, and it is the outcome of an interaction between a person and particular content. The potential for interest is in the person but the content and the environment define the direction of interest and contribute to its development(Hidi & Ann Renninger, 2006). It is hypothesised that the arrival of an individual to a house and the sense of occasion of the test may affect situational interest positively. It is not to say that a phone assessment would not trigger situational interest – the novelty of a phone-based assessment ought to hypothetically develop situational interest.

Situational interest is motivating and can play a part in a child's willingness to carry out the CLA. Research has also shown situational interest has a positive impact on reading comprehension (Hidi(1990) and Hidi & Baird (1998) as cited in (Hidi & Ann Renninger, 2006)), and therefore is a crucial factor to consider when assessing online. Further, psychological research has illustrated the impact of situational interest to help with inference, linking prior schema with new knowledge and focusing attention (McDaniel et al. (2000) , Hidi (1995), Kintsch (1980) as cited in (Hidi & Ann Renninger, 2006). Any phone-based assessment should ensure the situational interest among children is high.

Academic Self-Efficacy

Finally, academic self-efficacy refers to the belief a child may have of themselves regarding their competency to perform a specific academic task(Bong & Skaalvik, 2003). Research upon this concept has shown that it has malleable properties and can be easily influenced and is to do with perception in a specific future task. And like situational interest, research in self-efficacy has shown that it is predictive of performance and motivation, as per Bandura (1986) and Pajares and Miller (1994), as cited in Bong & Skaalvik (2003).

It has also been shown that self-efficacy can be influenced by verbal persuasion, especially if the persuasion is given by an individual who is an expert in that field(Bong & Skaalvik, 2003). Therefore, persuasion by the assessor at the outset of the task may help influence the child's self-efficacy regarding the test. It may also have an impact on positive affect. Alternatively, if the assessor cannot establish her/his credentials, their persuasion may have limited effect. Like situational interest, a strategy that allows the assessor the build a rapport, establish their credentials, and persuade a child to take the test would have to be designed to increase self-efficacy.

Overall, if online assessment affects these cognitive aspects of a child being assessed, we may find the items in the assessment illustrating Differential Item Functioning (DIF). DIF would occur when children with the same latent ability do not have the same probabilities of responding to a question correctly due to a feature of assessment. Therefore, to avoid DIF, assessment and engagement strategies with the child would have to lower cognitive load, ensure high self-efficacy and pique situational interest.

Psychometric Properties

Validity of data in remotely conducted CLAs

The validity of data is crucial as it is a cornerstone from which policymakers, educators and academics will choose the extent to which they engage with the data and infer implications for policy making.

In an in-person mode of assessment, validity is easily ensured by statistically testing validity and after effective training is given to assessors (as argued by Vagh in 'Is simple, quick and cost effective, also valid? Evaluating the ASER Hindi Reading Assessment in India' (UNESCO Institute of Statistics, 2016)). The assessor can ensure that no other person or resource assists the child in answering the question. The same cannot be ensured on the phone because it would be impossible to say for sure that the child's responses were completely based on independent cognitive skills, and there was no collaboration or assistance from a peer, parent, or sibling.

In the context of implementing a remote phone-based learning intervention in Botswana, protocols of assessment require that a child ought to explain the answer in order to be marked correct, as well as a time cap of 2 minutes when responding. This ensures the probability that someone assists the child is minimized (Angrist, Bergman, Brewster, et al., 2020). The study in Botswana had undertaken 2 parallel testing strategies, both drawn from the same population, and similar results make it confident of the validity of phone-based assessments. However, the assessments were not concurrent, understandably, given the context, and not formal validity assessments (Angrist, Bergman, Evans, et al., 2020). Finally, the method for verification, which asks children to articulate their mathematics responses is testing more than knowledge of arithmetic skills – must acknowledge that it may be cognitively more demanding than the test itself. This does not preclude the value of these methods but is critically appraising methods to verify and validate test results. There are statistical methods to identifying systematic errors and correcting for them. These methods can work for data on aggregated levels for policy purposes and not on household / hamlet levels.

Inter-rater and parallel reliability of remote CLAs

Studies have been carried out on ASER (Bhattacharjea et al., 2021) and Uwezo to check for their inter-rater reliability and both studies illustrated a high degree of inter-rater reliability (ACER Centre for Global Education Monitoring, 2015). It would be crucial to see whether this inter-rater reliability is maintained, when judging reading and numeracy skills remotely, particularly when the assessor also must judge whether the child is performing the task independently or not. Additionally, in-person assessments and in-phone assessments need to have high parallel reliability which ensure that in person assessments and phone assessments are equivalent as a study carried out in Sierra Leone did (Crawfurd et al., 2021), .

Protection and Safeguarding considerations

We must also consider the protection and safeguarding aspects when an assessment is carried out remotely. During online classes conducted during the Marginalised No More Project run by Street Child of Nepal from 2020-21, several parents expressed some concern when their daughters were taught online by a male teacher. This was even though parents had been made aware of classes being conducted online and had access to community focal points who could address their queries and concerns. Further discussion and reassurances were required to ensure sustained high participation of these participants. It is quite likely that parents may have some reservations about a female student being assessed by a male assessor. It might be a basis for not wanting to participate. Mitigation

steps, which would vary as per the demands of local context, must be taken. Oversight and accountability such as a supervisor monitoring a sample of calls, or recordings of several calls would need to be carried out. Research practices that apply in person for children would have to be adapted to the remote(phone-based) context.

Implications for Inclusion

Phone based assessments can lead to a sampling bias as it would lead towards only selection of those with access to a phone. By having a phone, and a working phone connection, the sampling bias is towards the privileged side of the technological / digital divide as was the case in the tutoring intervention in Kenya ((Angrist, Bergman, Evans, et al., 2020)). Since several studies have documented the correlation between socio-economic status and attainment, a sample that includes only those with phones is likely to be non-representative of the population. (Schueler & Rodriguez-Segura, 2021)

On the other hand, the use of technology can also be an enabler to access remote communities particularly from the *Himal* region of Nepal*,* as well as some parts of the *Pahad* region. In the caste segregated spaces and villages found in parts of the *Terai and Pahad,* phone assessment can overcome caste barriers. It also overcomes reservations communities or families may have of increased risk of infectious diseases by letting an assessor into their living spaces. Therefore, while accessing those without a phone in a remote assessment is hard to circumvent, the ability to access and assess communities which normally would have been excluded is valuable.

Critical Perspectives:

Providing feedback

Several large CLAs do not give immediate feedback in the interests of saving time in training and when conducting assessment. Beyond data collection, evidence of community mobilisation for stakeholder engagement and advocacy is very limited (Results for Development Institute, 2015). In that context, the parent and community's '*gain*' for participating in the survey is indirect and subject to policy makers taking onboard the data and implementing interventions that would particularly benefit their children.

A phone or in-person assessment ought to give feedback and provide time for parents to ask questions to increase parental and community engagement with the results. The two-way flow of information, responsive and adapted to the needs of the parents ameliorates the somewhat extractive nature of assessment. Perhaps for an assessment exercise tangibly owned and operated by the community, implementation can include extensive training on understanding the assessment process and results, within a broader framework of citizenship, accountability and rights. Not only would this leave the community with a pedagogical product, but it would also increase their engagement with the CLA exercise too. The magnitude of such a task, if carried out, would be extensive as it would have to overcome many social and education barriers. While it would be resource intensive and outside the scope of the EOL, it is nevertheless posited as the normative value for a community is immense.

Further, phone-based assessments can be more efficient from the perspective of time and less extractive as unlike in person assessments conducted at home, there are no social expectations of hospitality that several communities in South Asia feel obliged to adhere to when having visitors at home. Notwithstanding that, we must account for research fatigue that emerges from extensive research that is carried out by a growing social science discipline among vulnerable communities. We must question whether expecting parents to co-operate and repeatedly create an environment and observe protocols for effective, efficient, and economical assessment is reasonable.

Increased Cultural Distance

The notion of assessment as defined by contemporary standards, itself is alien and possibly illogical to some cultures (Sternberg, 2007). What is being assessed is non-neutral, as has been discussed adequately in the literature as well as the EOL paper series. The act of assessment itself is non-neutral. Sternberg (2007) describes how certain indigenous communities where people solve problems as a team would find it unnatural and strange that the assessment must be carried out alone. It would be important to explore whether the presence of an assessor who is a member of their community reduces this disconnect. Overall, it is important to consider whether remote assessment exacerbates the communities disconnect with assessment?

Conclusion

In conclusion, strategies for remote assessment must account for pedagogical implications of remote assessment and ensure interest, self-efficacy remains high and cognitive load remains low among students. Safeguarding training and operating procedures should adapt to an online mode of testing. High psychometric properties for the assessment framework are essential if these results are to be used for evidence-based policy making. While remote assessment can exclude those without phones, it should strive to include those communities who are normally too far to travel to. Taking cognisance of gaps in the past, remote interaction ought to be harnessed as a force multiplier to increase in engagement and mobilisation of communities with the data at pre-assessment and post-assessment stages. Remote assessments should be part of a wider remote engagement and agenda setting process, which harnesses technology to enable community participation in social accountability in education.

Bibliography

ACER Centre for Global Education Monitoring. (2015). *REPORT ON THE CONCURRENT VALIDITY AND INTER-RATER RELIABILITY STUDIES OF UWEZO* (Issue June).

Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020). Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana. *SSRN Electronic Journal*, *44*(August). https://doi.org/10.2139/ssrn.3663098

Angrist, N., Bergman, P., Evans, D. K., Hares, S., Jukes, M. C. H., & Letsomo, T. (2020). Practical lessons for phone-based assessments of learning. *BMJ Global Health*, *5*(7). https://doi.org/10.1136/bmjgh-2020-003030

Bhattacharjea, S., Saeed, S., Timalsina, R., & Ahamed, S. (2021). *Citizen-led Assessments : A Model for Evidence-based Advocacy and Action to Improve Learning* (Using Assessment Data in Education Policy and Practice: Examples from the Asia-Pacific, Issue 4).

Bong, M., & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different are they really? *Educational Psychology Review*, *15*(1), 1–40.

Crawfurd, L., Evans, D. K., & Hares, S. (2021). Teaching and Testing by Phone in a Pandemic. In *Pandemic Working paper* (Issue 591).

Hidi, S., & Ann Renninger, K. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Results for Development Institute. (2015). *Bringing Learning to Light: The Role of Citizen-led Assessments in Shifting the Education Agenda* (Issue June). https://hewlett.org/wp-content/uploads/2016/08/R4D - Bringing Learning to Light - June 2015.pdf

Schueler, B. E., & Rodriguez-Segura, D. (2021). A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya. In *EdWorkingPaper No. 21-432* (Issue 21).

Sternberg, R. J. (2007). Culture, instruction, and assessment. *Comparative Education*, *43*(1), 5–22. https://doi.org/10.1080/03050060601162370

Sweller, J., Merrienboer, J. J. G. Van, & Paas, F. (1998). Cogntive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3). https://doi.org/10.1023/A

UNESCO Institute of Statistics. (2016). *Understanding What Works in Oral Reading Assessments*. https://doi.org/10.15220/978-92-9189-196-2-en