Access to digital education and learning in the times of COVID-19: a latent class analysis

Santanu Pramanik, Reem Ashraf, Bijay Chouhan, Sonalde Desai

Introduction

The COVID-19 pandemic has not only affected physical and mental wellbeing of people, it has impacted people's livelihood, led to stagnation of economic growth and posed an unprecedented challenge to the education system. According to the UN Educational, Scientific and Cultural Organization (UNESCO) estimates, about 190 countries have closed schools nationwide during the early phase of the pandemic [1]. Although the evidence on the effectiveness of school closures to control transmission is weak in the context of SARS and an early evaluation in the context of COVID-19 also supports that [2]. Many countries gradually opened up schools at least partially, but some were forced to close schools again because of outbreaks in local communities [3].

India is one of those countries whose educational institutions, including schools, have been shut down in order to contain the transmission of COVID-19 since March 25 when the first nationwide lockdown was announced. All schools in India have remained closed for about six months and with no set date for reopening yet, it is unclear how long this closure will continue. In the case of Ebola outbreak, schools were closed between five and eight months in three West African countries. The length and intensity of the Ebola pandemic is perhaps the only health crises in the past to come close to the school closures being experienced in 2020 due to COVID-19 [3].

As a response to this crisis and with the objective of minimizing learning disruption, some schools in India have shifted their base to virtual platforms to conduct classes online or share learning materials through other web-based applications. However, not all schools and their teachers have the required infrastructure to follow this path, neither does all the children have the resources to access such remote learning facilities. Lately, the Government also took some initiative for remote learning in order to broaden the access base to educate students through digital platforms. The scheme *PM eVidya* is meant to be a one-stop solution for all learning needs of students during the pandemic. It includes studying through e-content, special radio podcast for visually and hearing-impaired students, and a dedicated educational TV channel for students who do not have access to internet.

Research question

Months of school closure can have serious negative impact on children's education, learning and labour market prospects [4]. Prior study suggests that school closures affect children from low socioeconomic status (SES) households more severely relative to children from higher SES [5]. The shift of responsibility of children's learning from schools to households often disproportionately affect those on the margins. Learning gaps widen as temporary alternative methods of remote learning remain inaccessible to the most marginalized [3]. In order to understand the impact of school closure on children's learning, as an intermediate step it is crucial to estimate the level of access to digital learning for school-age children from various socioeconomic background. As per our knowledge, surveys collecting data on actual use of digital learning during periods of school closure are sparse in India. We predicted level of access to digital mode of learning for children in the age group of 6 to 18 years old in the Delhi National Capital Region during the time of school closures using latent class analysis approach. We used data on explanatory variables determining the level of access to digital learning from existing surveys for this prediction problem.

Data and Research methods

Data

The Delhi Metropolitan Area Study (DMAS) baseline survey was conducted by researchers at NCAER National Data Innovation Centre during Feb-May 2019. The target geographical area for DMAS was the Delhi National Capital Region (NCR) which comprises 31 districts spread over four states, viz., Haryana (13 districts), Delhi (9 districts), Rajasthan (2 districts), and Uttar Pradesh (7 districts) [6]. Although it may not be apparent from the name, Delhi NCR is a highly diverse region including the metropolitan areas of Delhi as well as rural areas of districts in Haryana, Rajasthan, and Uttar Pradesh. Within a state, we considered a multi-stage stratified cluster sampling design. Districts, clusters, and households were selected at the first, second, and third stages of sampling. Clusters or the secondary sampling units (SSUs) were defined as census villages in rural areas and NSS Urban Frame Survey (UFS) blocks in urban areas. The goal of the sampling design was to select representative random sample at each stages of selection.

A sample of 5,253 households from the NCR were interviewed face-to-face for the DMAS baseline survey. Total number of individuals belonging to 5,253 households were 27,447. For our analysis, we prepare the analytic sample which includes households having at least one child in the age group of 6 to 18 and at least one of the parents lived with the child in the household. This results in 3,133 households and 6,510 children in our analytic sample.

Latent class analysis (LCA)

We used latent class analysis to classify school-age children to one of K mutually exclusive classes representing different levels of digital access. LCA uses the observed explanatory variables (manifest) in the data to classify units to groups (latent). Conditional upon values of this latent variable (group membership), responses to all of the manifest variables are assumed to be statistically independent. The model, in effect, probabilistically groups each observation into a latent class, which in turn produces expectations about how that observation will respond on each manifest variable.

Manifest variables

To classify the children into (latent) levels of digital access, we used the following manifest variables:

Table 1. List of manifest variables used to classify school-age children to one of K mutually exclusive classes representing different levels of digital access

Manifest variable	Type of variable
Household size	Count
Number of children in the age group of 6-18 in the household	Count
Number of mobiles (any type) used in the household	Count
Anyone in the household has a smartphone	Binary
Ownership of cable TV	Binary
Ownership of laptop/ computer	Binary
Hours of electricity	Count
Number of rooms in the house	Count
Number of household members who know how to use computers	Count
Number of household members who use internet	Count
Household’s economic status as determined by the wealth quintile	Ordinal categorical
Both parents have secondary or above education	Binary
None of the parents have above primary education	Binary

Covariates used for predicting class membership

The latent class regression model allows inclusion of covariates to predict individuals’ latent class membership. This is a one-step technique for estimating the effects of covariates, because the coefficients on the covariates are estimated simultaneously as part of the latent class model.

Table 2. List of covariates used to predict individuals’ latent class membership

Covariates	**Type of variable**
Child’s age	Continuous
Child’s gender	Nominal categorical
Child’s grade	Ordinal categorical
State of location	Nominal categorical
Area of residence	Nominal categorical (2 categories: Rural/ Urban)
Household caste group	Nominal categorical (3 categories)
Household religion	Nominal categorical (3 categories)

Analysis: To fit the latent class model, we used the poLCA package of R. poLCA is a software package for the estimation of latent class and latent class regression models for polytomous manifest variables, implemented in the R statistical computing environment [7].

Preliminary findings
Preliminary fitting of models suggest 4-class latent class model based on the minimum BIC, maximum likelihood and parsimonious model fitting criteria. Estimated class population shares are 26%, 20%, 33% and 22%. However, the latent classes and their ordering need to interpret in the context of the manifest variables which requires some more thinking.

Next steps
This is a work in progress. Next steps involve the following work to complete the paper. First, we plan to interpret the latent classes predicted by the model in the context of the manifest variables used for the analysis and understand the ordering of level of digital access. Then only it would be possible to provide a meaningful interpretation of the coefficients corresponding to the covariates used to predict individuals' latent class membership. The poLCA package, used for the analysis, allows only polytomous variables as manifest variables in the LCA model. Hence, the continuous variables were transformed into categorical variables having certain number of categories. This led to some loss of information. We plan to compare findings from poLCA package with other packages such as MCLUST package of R. A review of three latent class cluster analysis packages: Latent GOLD, poLCA, and MCLUST suggests that MCLUST outperforms other packages based on the criteria of within cluster homogeneity and between cluster heterogeneity [8].

References

1. UNESCO. *Global monitoring of school closures caused by covid-19. Available at https://en.unesco.org/covid19/educationresponse*. 2020 23 Sept, 2020].
2. Viner, R.M., et al., *School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review.* The Lancet Child & Adolescent Health, 2020.
3. Smith, W.C., *Potential long-term consequences of school closures: Lessons from the 2013-2016 Ebola pandemic.* 2020.
4. Azevedo, J.P., et al., *Simulating the potential impacts of covid-19 school closures on schooling and learning outcomes: A set of global estimates*. 2020, The World Bank.
5. Van Lancker, W. and Z. Parolin, *COVID-19, school closures, and child poverty: a social crisis in the making.* The Lancet Public Health, 2020. **5**(5): p. e243-e244.
6. National Capital Region Planning Board, *Annual Report, 2016-17*. 2017, Ministry of Housing and Urban Affairs, Government of India: New Delhi.
7. Linzer, D.A. and J.B. Lewis, *poLCA: An R package for polytomous variable latent class analysis.* Journal of statistical software, 2011. **42**(10): p. 1-29.
8. Haughton, D., P. Legrand, and S. Woolford, *Review of three latent class cluster analysis packages: Latent Gold, poLCA, and MCLUST.* The American Statistician, 2009. **63**(1): p. 81-91.