

Comparison of ASER and NCERT's National Achievement Survey (NAS)–Class V

Currently two large-scale learning assessments are conducted in India. Pratham/ASER Centre's Annual Status of Education Report (ASER) has been brought out annually since 2005. NCERT's National Achievement Survey (NAS) is conducted every three years, beginning in 2001-2002 for different grade levels.

ASER and NAS are designed for different purposes and employ different methodologies. This note describes and compares these methodologies so that informed conclusions can be reached. The note is based on ASER 2005-2012¹ and the NAS report for Class V, Cycle 3, 2010-2011 released in 2012.²

Objectives, sampling and coverage

ASER is designed to generate district, state, and national level estimates of children's schooling status for all children age 3-16, and estimates of basic ability in reading and arithmetic for all children age 5-16. It is designed as a household-based survey so as to include all children: those enrolled in government schools, private schools, other types of schools, and those not enrolled in school.

ASER aims to cover all rural districts each year. It employs a two-stage sample design. At the first stage, 30 villages are selected in each rural district from the Census 2001 directory using Probability Proportional to Size (PPS). In the second stage, 20 households in each village are randomly selected. All children age 3-16 in sampled households are surveyed. All children age 5-16 are assessed.

ASER 2012 reached 331,490 households in 568 districts. 595,139 children in the age group 3-16 were surveyed and 448,467 children age 5-16 were assessed.

NAS-Class V aims to "provide reliable information on the achievement of students in the elementary sector of education in government and government-aided schools" (p.3). It is a school-based survey intended to assess grade level competencies of children enrolled in Std. V in government and government-aided schools.

NAS aims to cover all 35 states and Union Territories. It employs a three-stage cluster design (p.11). In the first stage, districts are selected using PPS. In the second stage, schools within sampled districts are selected, again using PPS. In the third stage, students are randomly selected within sampled schools.

DISE 2007-08 was used as sample frame for NAS-Class V. The report notes significant discrepancies between DISE data and actual school enrolments (p.22). NAS-Class V, Cycle 3 was implemented in 31

¹ See www.asercentre.org for ASER reports from 2005 to 2012 and additional details on methodology.

² See www.ssatcfund.org/Home/Publications.aspx for the NAS-Class V report for cycle 3, which used a different methodology from earlier cycles. All page numbers referred to in this note refer to this report.

states and Union Territories. It covered 122,543 children from 6,602 urban and rural schools across 27 states and 4 Union Territories. (p.1)

Tools and testing

ASER assesses early reading and basic arithmetic ability, which are foundational skills fundamental to literacy and numeracy acquisition. Early reading ability implies the acquisition of letter knowledge, ability to decode Std. 1 and 2 level words and fluently read Std. 1 and 2 level passages. ASER tools are designed to assess mastery of these foundational skills and are not intended to differentiate within each mastery level. For instance, amongst the group of children identified as fluent readers of Std. 2 level text, the ASER assessments are not designed to differentiate between their ability to read and to comprehend.

The highest level tested in reading is a Std. II level text. The highest level tested in arithmetic is a 3-digit by 1-digit division problem, usually taught in Std. III or IV. Tools and testing procedures are available in the public domain.

NAS-Class V assessed grade level competencies of Std. V students in language (including reading comprehension), mathematics and environmental science (p.3). NAS-Class V, cycle 3 test forms are based on common core content and competencies identified from an analysis of state textbooks (p.4). 40 multiple choice test items were constructed for each subject. The language test additionally included a writing task (p.10). Tools, testing procedures, and grading rubrics for the writing task are not in the public domain.

Test administration

ASER is a household survey. ASER reading and arithmetic assessments are administered one on one in an oral format. Children are tested at home. All children are given the same test, regardless of age or grade.

NAS-Class V is a pen and paper test administered to a group of students in school.³ In most schools, children were tested in two out of the three subjects (p.177). The cover of the test booklet has instructions for students indicating how to record or modify their responses (p.11).

Process implementation and monitoring

ASER is conducted each year by surveyors from partner organizations in each district. These include DIETs, teacher training colleges, universities, NGOs and others. Surveyors receive intensive three-

³ The NAS-Class V report states that within each school, children were selected from class registers using simple random sampling (implemented via a lottery). This seems to imply that only children present in school on the day of the test were included (p.177).

day training in preparation for the survey, including a full day of practice in the field. ASER devotes considerable time and effort to ensuring data quality through carefully designed and implemented training, monitoring, and recheck procedures, details of which are provided in each year's ASER report and on the ASER Centre website.

In addition to an assessment of surveyors' understanding of the process before the survey rollout, quality checks comprise two main processes: monitoring of survey teams during the actual field survey, and recheck of data after the survey has been conducted. These processes are implemented by the central and state ASER teams and Master Trainers in each state. In ASER 2012, more than half of all villages were monitored and/or rechecked. External process audits of the ASER field work and data collection process are also conducted periodically.

NAS-Class V was coordinated by state agencies like SCERTs and SIEs (p.16). Data collection was done by DIET students. The report notes the possibility of insufficient training and practice given to field investigators (p.22). The training manual is not in the public domain. No information on recheck procedures is available in the report.

Accuracy of estimates

ASER estimates are self-weighting at the district level. At the state and national levels, estimates are weighted by the appropriate population weights.

ASER does not report standard errors and margins of error for its state and national estimates. However, a study done on the precision of ASER learning and enrollment estimates shows that margins of error are well within 5% at the state level.⁴ In addition, a detailed check of sample sizes is done for smaller states where sample sizes can be small for some sub-populations. Where the number of observations in the sample is found to be insufficient, estimates are not presented in the report.

Since 2011 ASER reports also present estimates at divisional levels. These estimates are presented with the associated standard errors and margin of error.

NAS-Class V estimates are not weighted. The report notes: "Unfortunately, due to discrepancies in the DISE data, limitations in the sampling method and loss of information at the sampling and administration stages of the survey, it was impossible to estimate sample weights for the survey. Therefore, student responses of class V (NAS) data were equally weighted within their state/UT data and each state/UT carried equal weight as a reporting unit" (p.178). The report notes that this posed problems for aggregation of data and generation of estimates (p.178-179). In particular, it states, "It

⁴ Ramaswami and Wadhwa 2010, "Survey Design and Precision of ASER Estimates". Available at http://img.asercentre.org/docs/Aser%20survey/Technical%20Papers/precisionofaserestimates_ramaswami_wadhwa.pdf

is important to note that such results are not the average for the pupils nationally since states with larger populations are not weighted more highly, as they would be, for a national or group pupil average" (p.179).

The NAS-Class V report presents the standard errors associated with the estimates that are reported.

Availability of results

ASER findings are made available in the same school year that the fieldwork was conducted. The survey is conducted between September and November of each year and the report is published the following January. District, divisional, state, and national level estimates are in the public domain.

NAS-Class V report was released in July 2012. Fieldwork was conducted between November 2010 and March 2011. (p.xxi)

Test reliability and validity

ASER tests assess achievement of mastery rather than the relative standing of children in relation to their peers. Reliability in this case refers to the consistency of the decision-making process in assigning children to a mastery level across repeated administrations of the test. In addition, since examiners assign each child to a mastery level, it is important to also estimate the consistency of the decision-making process across examiners, which in technical terms is referred to as inter-rater reliability. A series of studies⁵ indicate substantial reliability of decisions across repeated measurements (test-retest) and satisfactory inter-rater reliability.⁶

The validity of the ASER reading test (that is, whether the test actually measures the constructs it is intended to measure) was examined using the Fluency Battery as a criterion measure for estimating the validity of the ASER Hindi language tool. The Fluency Battery is a test of early reading ability adapted from the Early Grade Reading Assessment (USAID, 2009) and the Dynamic Indicators of Basic Early Literacy Skills (University of Oregon Center on Teaching and Learning, 2002). The Fluency Battery is a test of early reading ability similar to the ASER language tool, but it is a longer and more

⁵<http://img.asercentre.org/docs/Aser%20survey/Tools%20validating%20the%20aser%20testing%20tools%20oct%202012%202.pdf>

⁶The test-retest correlation coefficients for the ASER-reading test for all children from Grades 1-5 is .95 and for the ASER math test is .90. More importantly the average Cohen's kappa estimate for decision consistency across repeated test administrations for the ASER-reading test is .76 and for the ASER-math test is .71.

The inter-rater reliability estimated using Cohen's Kappa for a group of 590 children is .64 for the ASER reading test and .65 for the ASER-math test on average, also indicating 'substantial' agreement. The average and median weighted Kappa across all pairs of examiners is .82 and .81 respectively for the ASER-reading test and is .79 and .80 for the ASER-math test indicating 'almost perfect' agreement for the ASER-reading test and 'substantial' agreement for the ASER-math test.

detailed assessment comprising 6 subtests. Children's reading is timed using a stopwatch and scores represent number of units (akshars/words/nonwords) read accurately in one minute. The ASER language assessment is strongly associated with the Fluency Battery. The magnitude of the correlation coefficients range from .90 to .94 (a correlation coefficient of 1 indexes a perfect and positive association between two measures).⁷

NAS-Class V report does not discuss validity or reliability of the tests utilized.

Comparisons over time

ASER has used the same sampling and assessment procedures since 2007. The reading assessment tool has not changed since 2005. All estimates generated since 2007 are comparable.

NAS-Class V, cycle 3, used Item Response Theory (IRT) to analyse the data, unlike earlier two cycles of the survey which used Classical Test Theory (CTT) (p.17). The report points out that the results of the most recent cycle are therefore not comparable with earlier years. (p.23)

CONCLUSIONS

ASER and NAS surveys are very different in test content, methodology, sampling, purpose, and years for which the results are reported. More importantly, the results are also computed very differently. Since estimates generated by each of these assessments neither cover the same populations nor assess the same content, their results are not comparable.

⁷ http://img.asercentre.org/docs/Aser%20survey/Tools%20validating_the_aser_testing_tools__oct_2012__2.pdf