Technical Standards for National Education Assessment 2020

Bhutan Council for School Examinations and Assessment
Royal Government of Bhutan

Technical Standards for National Education Assessment 2020

Assessment and Monitoring Division
Bhutan Council for School Examinations and Assessment
Thimphu: Bhutan
2021

This Technical standards is developed by Assessment and Monitoring Division (AMD), Bhutan Council for School Examinations and Assessment (BCSEA) in collaboration with the Australian Council for Educational Research (ACER), India under NEA project.

The project is funded by Global Partnership for Education (GPE) under Education Sector Programme Implementation Grant (ESPIG) with support from United Nations International Children's Fund (UNICEF) and Save the Children International-Bhutan (SCI).

Acknowledgement

Bhutan Council for School Examinations and Assessment acknowledges the invaluable contributions made by the following officials in developing this guideline.

Authors:

1. Arjun Kumar Gurung, Project Manager NEA project, BCSEA
2. Karma Jigme Lepcha, Examination and Assessment Officer, SED, BCSEA
3. Kinley Dema, Principal Education Monitoring Officer, AMD, BCSEA
4. Lhamo Sonam, Education Monitoring Officer, AMD, BCSEA

ACER India

1. Dr. Mee Young Han, Senior Research Fellow
2. Mr. Amit Pathak, Research Fellow

Logistic and Administrative

1. Sangay. Tenzin, CPO, BCSEA
2. Bhim Kumar Pradhan, Project finance, BCSEA
3. Kezang. Dema, Adim Assistant, BCSEA

Contents

1.1 Introduction

Technical Standards is a key document that clearly describes standards of technical quality for all aspects of the learning assessment that indicates how standards are used as a part of quality monitoring and reporting. It lays down the standards of technical quality that serves as guidelines to establish the qualities or requirements of a given process or output. In the context of National Education Assessment (NEA), the technical standards explicitly indicate the expectations on the quality of the assessment as approved by the Core Team of Bhutan Council for School Examinations and Assessment (BCSEA). This document is designed to guide and provide foundation for the proper implementation of the NEA at various grade levels of learning – III, VI and IX. It ensures that the processes followed during the implementation of NEA are of high quality and the results are qualified as valid and reliable.

1.2 Purpose of Technical Standards

The objective of developing technical standards for the NEA is to establish a set of procedures for the collection of data in a systematic manner. The Technical Standards assist BCSEA and related agencies by providing guidelines on expectations of data quality and NEA programme implementation.

The technical standards complement the NEAF. It is essential to critically review the NEA processes based on the technical standards post assessment, to evaluate their suitability, to assess the extent to which the standards are adhered and to document any variations in the implementation of the NEA.

Documenting the adherence to, and deviation from the technical standards will provide transparency on the assessment quality, and thus improve the confidence of stakeholders in the assessment results and support the use of the assessment data in education policy making and planning.

1.3 Fundamental Principles

The technical standards described in this document are based on the international best practices and they are simple, measurable and enforceable. The standards are tailored to give all participating agencies and stakeholders a clear understanding of the assessment processes and expected outcomes. This document is built on the following four major and interrelated principles.

Consistency

Implementation of test instruments and data collection should happen in a consistent manner to enable fair and reliable comparisons of student performance across the country. For example, test conditions, data collection and data processing should all remain consistent throughout the implementation to avoid inadvertent errors or bias. This ensures differences in student performance are due to differences in their ability rather than factors unrelated to individual abilities.

Precision

Assessment instrument, data collection and processing should be precise to avoid unintentional errors. Systematic and random errors are the two main types of errors. Systematic errors are caused by variation in data collection and submission, for example using different conditions (different time duration, poor light or ventilation in some classrooms, etc.) during testing. Random errors are caused by unknown and unpredictable factors, they mostly arise during data collection and preparation process. For example, coding procedures may leave room for interpretation. It is important that strict data collection and preparation standards are developed to ensure better estimation of student achievement.

Generalizability

The NEA is a sample-based, however data collected will be used to make inferences about the general achievement of students in the country. Therefore, it is essential that individuals to be tested, test materials

developed, sampling strategy and other components should be selected in a way that valid inferences can be made about the target population of interest at large.

Timeliness

A series of tasks and activities are interrelated while conducting the NEA. It is important that all tasks are conducted as per the pre-defined timeline and survey results are disseminated in time, in order to guide important policy decisions. A failure to implement one task could have a cascade effect, thereby delaying the whole program.

1.4 Key Technical Standards

These standards are developed to address data and management related aspects of the NEA for both the Field Trial and Main Survey. The key technical standards of each aspect of the project that relate to major processes and outcomes are described as given below.

1.4.1 NEA Management

Outcomes of the NEA depend on the elements of the assessment and on how the NEA is managed and executed. Therefore, clear guidelines must be established at the national level. A team of experienced managers and coordinators must also be set up to execute the NEA smoothly and timely in different Dzongkhags/Thromdes.

Standard 1

A core team must be established at the national level, with sufficient expertise in large-scale assessment to oversee NEA implementation.

Standard 2

An advisory board comprising of representatives of key stakeholders must be established to advise on matters of policies, politics, culture, technical or any other issues that might arise in the process of implementing the NEA. Clear protocols for test implementation and data collection must be outlined at the national level.

Standard 3

Each Dzongkhag/Thromde must have coordinators appointed who are familiar with the assessment processes – test administration, coordination, and data collection. The coordinators must work according to the NEA guidelines and any deviation from the described protocol must be immediately communicated to the NEA implementing agency.

1.4.2 Communication

Standard 1

Concerned officials must be available to respond in a timely manner to queries related to issues that might arise during the implementation of the NEA. Adequate and proper communication channels must be set-up for sending out common instructions or messages and responding to specific queries from different administrative levels.

1.4.3 Timelines

BCSEA with the relevant stakeholders and ACER, India, must jointly agree and adhere to timelines for the preparation and implementation of NEA.

Standard 1

Each of the following elements of assessment must be prepared and finalised within the agreed time-frame:

- technical standards
- assessment framework
- cognitive instruments
- contextual questionnaires
- field administrative manual
- field trial
- main survey
- data collection
- data coding and management
- data analysis
- reporting and dissemination

1.4.4 Security of Data and Materials

Standard 1

All materials related to NEA must be maintained securely and confidentially. No personnel involved in the assessment are allowed to copy or circulate NEA documents or data without prior permission from BCSEA, unless the materials are made publicly available.

Standard 2

Assessment materials must be made available to respective officials only through secure central NEA management system. Necessary measures must be adopted to control access to materials by allotting unique IDs and password. In the absence of an online system to access materials, distribution of materials must be supervised and monitored at the nodal centres.

Standard 3

Student data must be securely maintained and de-identified before sharing with anyone.

Standard 4

Access to primary or raw data must be provided for analysis only after prior permission from BCSEA.

Standard 5

Officials involved in the implementation of NEA at the national level or the Dzongkhag/Thromde level must sign a confidentiality agreement. Any breach of security must be immediately informed to the NEA implementing agency.

1.4.5 Population and Sampling

The following standards aim to achieve level of precision at par with the large-scale international studies such as PISA and TIMSS. Meeting the standards given in this section will ensure that assessment data is collected from appropriate and representative sample, so that inferences drawn from the study are valid for the entire target population. These standards do not apply to the Field Trial but to the Main Survey, because the main purpose of field trialling is limited to check test item functionality.

Standard 1

A detailed NEA Sampling Plan must be developed to specify level of precision, exclusions, and stratification variables.

Standard 2

The target population for grades III, VI, IX studying in government and private schools across Bhutan must be identified and documented. While it is not possible to include every child from the target pop-

ulation in the assessment due to various forces, a representative sample must be taken from the target population as per the appropriate sampling technique. The sampling exercise must begin only after making important decisions regarding school-level and within-school exclusions, sampling strata, sample size, replacements and weighting.

Standard 3

Exclusions within each reporting unit must not be greater than 5% of the target population to ensure coverage of at least 95% of the target population. The exclusion criteria must be pre-defined in the NEA Sampling Plan, such as geographical inaccessibility, logistical issues, CWDs within-school exclusion due to disability that severely impedes their ability to perform in the assessment.

Standard 4

As per the international best practice, the sample size must be large enough to yield a 95% confidence interval for study estimates. To achieve the necessary level of precision and accuracy in prediction, systematic sampling must be used for identifying a representative sample and for reporting on sub-groups of interest. To maintain consistency, a single data source must be used across all sampling units when building the sample frame.

Standard 5

The NEA must follow a systematic multi-level sampling as used in the large scale international studies like PISA and TIMSS. The sampling must have two levels - school level, and student level. Schools must be sampled at the first level by the method called Probability Proportional to Size (PPS) while students must be sampled at the second level by Simple Random Sampling (SRS).

Standard 6

School response rate[1] must be at least 85% to ensure representativeness of the sample by minimising non-participation at the school level. Representativeness of the sample is most critical in achieving accurate predictions about the target population. If the said response rate is not achieved in a reporting unit, then this data may not be included in the national data. An adjudication must be launched to determine the circumstances and reasons behind the low response rate. Data from such a reporting unit can be included in the national data only if it does not undermine the national data after such scrutiny.

An acceptable solution can be used to achieve the recommended school response rate through a systematic use of replacement schools. Replacement process, its pre-conditions, and approval process for including data from low response rate reporting units must be defined in the detailed NEA Sampling Plan.

Standard 7

Student response rate[2] must be above 80% of all the sampled students across all of the responding schools. This is to ensure representativeness of the sample by minimising non-participation at the student level.

Standard 8

Student response rate for a school must be more than 50% of all the sampled students within the school, failing which it is regarded as non-participating.

Standard 9

Absent sampled students must not be replaced with non-sampled students.

Standard 10

Field trials must have at least 200 students per test item to meet the minimum number of students for calculating reliable item parameters.

---

1 The percentage of sampled schools that actually participate in the assessment is called as the school response rate.

2 The percentage of sampled students who actually participate in the assessment is called as the student response rate.

Standard 11

Within-school exclusions can be allowed due to the reason that some students have such extreme disabilities that it severely impedes their ability to perform at the NEA. It should not be more than 2.5% of the NEA desired population.

1.4.6 Instrument and Material Preparation

The outcomes of the NEA are not only dependent on the extent to which the procedures and protocols are adhered to, but are also dependent on the quality of the assessment instrument and the contextual instrument used. Therefore, it is essential that the instrument meets the requirements and specifications as discussed and agreed on, at the onset of the assessment cycle.

Standard 1

The assessment framework must provide explicit details about the information of the construct, target population to be assessed, test design and test blue-print along with the cognitive processes and contextual factors that need to be investigated.

Standard 2

A rigorous process for developing and designing the test instrument must be adopted. All test items must relate to the specifications laid out in the assessment framework and meet all quality assurance processes such as panelling and piloting.

Standard 3

After all items have been piloted or trialled, psychometric analysis of each item must be performed to analyse its appropriateness for inclusion in the Main Survey. Therefore, a sufficient number of extra items must be made available to allow for the inevitable failure of a certain proportion of items at the piloting or trialling phase.

Standard 4

The contextual questionnaires must be relevant to the context and appropriately targeted to each cohort of informants. It must meet the quality assurance processes such as panelling and piloting.

Standard 5

Assessment materials must be printed using high quality paper. The print quality, format and layout for all instrument materials must be of the same standard.

Standard 6

Any deviation in the standards related to design or quality of assessment materials must be recorded as reference for the current as well as for future cycles of the NEA.

1.4.7 Test Administration

In order to have a fair comparison of data across the country, it is important that the test conditions for all students are as similar as possible. To reinforce this, a test administration manual must be developed. The manual specifies all elements of test administration such as test timing, test conditions, student tracking, assigning test booklets etc.

Standard 1

Respective officials from different stakeholders must be oriented and trained to monitor and administer NEA. All test administrators must follow the instructions as stated in the test administrator's manual both during the Field Trial and the Main Survey.

Standard 2

Test administrators must not have any personal or professional relationships with the schools or the students they are administering. This is to ensure that the relationships between test administrators and students do not affect the credibility of the test session.

Standard 3

Some test administration sessions (10-15 %) must be randomly monitored by the trained monitors to ensure quality assurance.

1.4.8 Data

Data standards refer to manual and machine-based data collection, processing, coding and scoring procedures. Meeting these standards minimizes errors in the database and subsequent delays in processing due to error checking and correction mechanism.

Standard 1

The coding scheme given in the scoring guide must be followed as per the direction from the item developers. If OMR is used in the study, the OMR processing must undergo field trials along with the testing instruments to ensure adherence to the standards given in this section.

Standard 2

Coders must be hired and trained as per the agreed coding guidelines and procedures. If OMR is used in the study, the test administrators must be trained to ensure that OMR sheets are filled by the students as per the required standards.

Standard 3

Data must be entered into the system using NEA data management system. The data are then stored in '.csv' or '.dat' or '.xlxs' format in the required folder structure and file naming convention as per the file and folder structure given in the NEA data entry manual.

If OMR is used, OMR scanning process must ensure the output in '.csv' or '.dat' format in the required folder structure and file naming convention as per the file and folder structure given in the NEA data entry manual.

Standard 4

Data cleaning and preparation phase address anomalies in data through a series of iterative and interactive processes meant for detecting and correcting errors in the raw data. The process must include the following:

- Missing value analysis;
- Data validation of mandatory and optional data fields;
- Omission or duplication of data;
- Consistency and credibility of data based upon pre-specified acceptable responses for all variables; and
- Cross-verification of invalid/missing/duplicate/omitted responses.

Standard 5

In order to assess correlation of various contextual factors with student achievement scores, data must be collected through contextual questionnaires. Sample students, teachers, school principals, and DEOs/TEOs must fill the respective contextual questionnaires /OMR sheets as per required standards.

Standard 6

Only the teachers and schools associated with sample students must fill the respective contextual questionnaires.

1.4.9 Psychometric Standards

Psychometric standards quantitatively define the technical quality of the test data and the interpretation that can be derived from the test result. It ensures that the statistical analyses, comparisons and inferences are based upon appropriate statistical procedures that are internationally acceptable.

Standard 1

The NEA data must be analysed using Item Response Theory (IRT), a robust technique used in credible large scale assessments worldwide to compute item parameters and student ability/scores.

Standard 1.1

After the Pilot NEA, cognitive data must be analysed using IRT run1 and run 2. The selection of items for the Main NEA must be adhered to the following fit indexes:

- Item-Rest Correlation/item discrimination should be in between 0.2 to 0.6.
- Weighted MNSQ should be in between 0.8 to 1.5.
- Item Delta(s)/ item difficulty should be in the threshold of -3 to +3 on the logit scale.
- PV Avg/ ability difference between the key and other option should be minimum of 0.15 and the PV Avg of the key should be higher than the other options.
- Pt Bis for key should be positive and remaining options should be negative.

Standard 2

A performance scale must be developed on the basis of IRT analysis followed by creating level descriptors for various achievement levels of students so that a meaningful snapshot of student performance can be drawn.

Standard 3

Each domain must have its own scale.

Standard 4

Appropriate scaling method must be employed to enable equating across time and context of assessments. In 2021, the Main NEA was carried out at grade III only. Therefore, the initial performance scale was developed for grade III. As other grades will be added in the future NEA cycles, the subsequent analyses will be extended in the performance scale to measure learning progression of students from grade III to IX.

Standard 5

Each performance scale must be adjusted to have scores ranging between 0 to 600, with a mean score of 300 and standard deviation of 50.

Standard 6

In addition to IRT analysis, further appropriate analysis must be carried out using univariate, bivariate and multivariate techniques as may be required. For data interpretation, measurement and sampling errors must be reported along with the point estimate values. Standard errors and sampling weights must be computed using appropriate statistical methods.

Standard 7

Observed differences are indicated as statistically significant if the applicable hypothesis testing rejects the null hypothesis at $\alpha = 0.05$ or less. The conclusions of interest are supported by two tailed tests of significance, followed by post-hoc tests as may be required.

Note: The key aspects of psychometric standards related to reliability, validity, important indicators of item characteristics under IRT and Differential Item Functioning (DIF) should be incorporated after Field Trial.

1.4.10 Reporting

The technical standards for reporting must ensure that the results are reported as per international and institutional standards containing information that are relevant for various stakeholders.

Standard 1

Results and recommendations must be presented in reports which are consistent with institutional reports. The level of details included in each report may vary as per the intended audience and type of report.

Standard 2

All reports must allow readers to understand the nature and limitations of the outcomes presented. Therefore, appropriate uncertainty estimates, such as confidence level and standard errors, must be reported at appropriate places alongside graphs and tables inside the reports.

Standard 3

Methodologies for each assessment component, data, interpretations, recommendations and policy insights must be documented in the technical report.

Standard 4

Results of psychometric analysis of student performance and their correlation with the contextual questionnaires must be reported as per the instrument design.

Standard 5

After the completion of the Main NEA, a full set of data must be generated along with data dictionaries. Based on the defined requirements and prior permission, this database must be made available to the interested researchers for secondary analysis.

"An educated and enlightened society guided by GNH principles"

Vision - Ministry of Education, Bhutan

"An internationally recognized centre of excellence in educational assessment that provides quality services to build the integrity and profile of the education system leading to an improvement in the quality of learning in Bhutan"

Vision - Bhutan Council for School Examinations and Assessment, Bhutan