

Analysis of National Learning Assessment Systems [ANLAS]

ANLAS Ethiopia:
Country Report

June 2019

ANLAS Ethiopia National Team

Mesaye Demessie Zeleke (focal point)

Yilikal Wondimeneh Demissie

Arega Mamaru Yewore

Nega Gichile Bongasse

Abraham Mengistu Sertse

Talefe Eshete Awoke

Aregawi Gidey Mesele

Bekele Geleta Ayana

Effa Gurmu Bati

Getachew Abebe Tsigie

Belay Endeshaw Gizaw

Robel Getachew Worku

Table of Contents

ANLAS Ethiopia National Team	1
Acknowledgments.....	3
Abbreviations.....	4
1 Introduction.....	5
1.1 ANLAS purpose	6
1.2 The ANLAS model	6
2 ANLAS in Ethiopia	8
2.1 Implementation process.....	8
2.2 National team training and planning.....	9
2.3 Stakeholder briefings.....	9
2.4 Completing the analytical tables	9
2.5 Preparing and disseminating a country report to relevant stakeholders to inform education sector planning	12
3 Findings of ANLAS.....	13
3.1 Evaluation of key areas.....	13
3.2 Aspects of improvement and recommendations.....	15
4 Using the ANLAS findings	38
Appendix A: The national team	39
Appendix B: Stakeholder and document mapping tables	40
References	43

Acknowledgments

The Global Partnership for Education (GPE), under the leadership of Ramya Vivekanandan and the support of Medjy Pierre-Louis, has initiated the Analysis of National Learning Assessment Systems (ANLAS). It has assigned the development of ANLAS to the Australian Council for Educational Research (ACER).

The ANLAS development process has three phases: (1) development of the ANLAS toolkit (July 2018 to January 2019); (2) pilot implementation of ANLAS in three GPE partner countries (January to June 2019); and (3) revision and refinement of ANLAS based on observations and experiences from the pilot implementations (June to August 2019).

Ethiopia is one of three countries participating in the piloting of ANLAS. Each piloting country contributed significantly to the consultative development process of ANLAS, allowing the testing of both the content of the ANLAS toolkit and the process of using these resources to analyze the national learning assessment system and to make recommendations for improvement to inform strategies within education sector plans.

The Ethiopia Ministry of Education endorsed the ANLAS study and nominated knowledgeable and responsible national team members. In this regard, special acknowledgment should go to H.E. Dr. Tilaye Getie, Minister of Education, and H.E. Ato Mohamed Ahmedin, State Minister of Education, for their concern and support of the team in all respects. The team also extends its thanks to Ato Araya Gebre-Egziabher, Director General of the National Educational Assessment and Examinations Agency (NEAEA), for his facilitation and for creating a conducive environment for the team and the study as well.

Sally Robertson (ACER) supported the national team throughout the piloting process and Ursula Schwantner (ACER) provided support during the in-country visit.

Last but not least, the national team gratefully acknowledges all the ANLAS participants drawn from teachers, senior experts in education, school leaders, training providers, development partners, and parents for providing the necessary information, active participation and their concern.

Abbreviations

ACER	Australian Council for Educational Research
A4L	Assessment for Learning
CTE	College of Teacher Education
DCP	developing country partner
DFID	U.K. Department for International Development
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EMIS	education management information system
ESA	education sector analysis
ESDP	Education Sector Development Program
ESP	education sector plan
ETP	Education and Training Policy
FGD	focus group discussion
GAML	Global Alliance to Monitor Learning
GEQIP	General Education Quality Improvement Program
GEQIP-E	General Education Quality Improvement Program for Equity
GER	gross enrollment rate
GPE	Global Partnership for Education
GP-LA	Good Practice in Learning Assessment
MoE	Ministry of Education
n.d.	no data
NEAEA	National Educational Assessment and Examinations Agency
NLA	National Learning Assessment
NLSA	National Large-Scale Assessment
PISA	Program for International Student Assessment
REB	Regional Education Bureau
READ	Russia Education Aid for Development
SABER	Systems Approach for Better Education Results
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SNNP	Southern Nations, Nationalities, and Peoples' Region
TIMSS	Trends in International Mathematics and Science Study
UIS	UNESCO Institute for Statistics
UN	United Nations
UNICEF	United Nations Children's Fund
USID	Universal, Sustainable, Inclusive Design for social change

1 Introduction

Ethiopia is located in the area of northeast Africa known as the Horn of Africa. It covers 1.1 million square kilometers and has an estimated population of more than 90 million with highly diversified ethnic groups, languages, culture and topography. The country has adopted federal governance with nine regions and two city administrations. A prime minister heads the government, and executive power is exercised by the government. The federal legislative power is vested in both the government and the two chambers of parliament.¹



The current Education and Training Policy (ETP) was enacted in 1994 to alleviate the problems of access, curriculum relevance, equity, quality and related challenges of the education sector.² As a result of the policy priority and financial input for the sector, Ethiopia made significant progress toward achieving the target set for primary education to be universal by 2015. For instance, the gross enrollment rate (GER) of primary education reached 109.3% in 2017/18. During this time, the number of students and teachers totaled 26,905,580 and 583,461, respectively. The number of public and private schools had also reached 41,441 in 2017/18.³

The use of assessment in Ethiopia started in 1938 with National Examinations; however, National Learning Assessment (NLA) was implemented much later, in 2000. Although classroom assessment was introduced along with modern education in 1908, it was given more attention in the 1994 ETP. To monitor and evaluate the quality of education on a regular basis, the National Educational Assessment and Examinations Agency (NEAEA) was re-established by the Council of Ministers Regulation No. 260/2012 by integrating the two assessment systems, National Examinations and NLA.⁴ National Examinations is a census-based assessment that takes place in grades 10 and 12; NLA is a sample-based assessment undertaken in the exit cycle of primary and secondary education (grades 4, 8, 10 and 12). Early grade literacy and numeracy assessment, on the other hand, is conducted at grades 2 and 3. Both the National Examinations and the NLA have been administered on the basis of the national education and training policy and curricula of the country.

Though tremendous achievement has been observed on access of education for all citizens, the quality of education as indicated in successive NLAs and National Examinations is still a critical challenge for the nation. To this end, the assessment system should be strengthened to support evidence-based intervention and decision-making.

¹ National Planning Commission Central Statistical Agency, *National Strategy for the Development of Statistics (2015/16–2019/20)* (Addis Ababa: Central Statistical Agency, 2016).

² Transition Government of Ethiopia, *Education and Training Policy* (Addis Ababa: Ministry of Education, 1994).

³ Ministry of Education, *Education Statistics Annual Abstract* (Addis Ababa: Ministry of Education, 2018).

⁴ Council of Ministers. "National Educational Assessment and Examinations Agency Establishment Council of Ministers Regulation No. 260/2012," *Federal Negarit Gazeta* 15 (January 26, 2012).

1.1 ANLAS purpose

Learning assessments are increasingly used by education systems around the world to inform evidence-based policymaking and reform processes, and to improve teaching and learning practice.⁵ To ensure the learning assessment systems' sustainability and effectiveness, it is important to regularly review and evaluate their actual state and progress.

For the Global Partnership for Education (GPE), an important priority is the quality of learning assessment systems as a key component of strong education systems and as a vital means to monitor and improve learning outcomes. GPE is a multi-stakeholder partnership and funding platform that focuses on supporting the efforts of more than 65 developing country partners (DCPs) in building effective education systems to improve equity and learning. In addition to its efforts to reinforce learning assessment systems through its grants to DCPs, GPE works to strengthen learning assessment systems and to promote a holistic measurement of learning through its Assessment for Learning (A4L) initiative, launched in 2017.

The Analysis of National Learning Assessment Systems (ANLAS) is a key component of GPE's A4L initiative, aiming to support partner countries to build effective monitoring systems for evidence-based education sector planning.⁶

ANLAS is designed to enable country partners to undertake a comprehensive analysis of national learning assessment systems, covering learning assessments in school education from primary or basic education to secondary education, in all schools within the system (public, private and community), and at central as well as decentralized levels. The aim of the qualitative analysis is to identify areas and recommendations for improvement of the national assessment system, to inform the development and implementation of improvement strategies as part of the wider education sector planning process.

As part of the intervention, to guide and assist developing country partners in undertaking ANLAS, a framework and toolkit have been developed. The toolkit consists of process tools to support the implementation of ANLAS, analytical tools to document and analyze the learning assessment system, and reporting and dissemination tools.

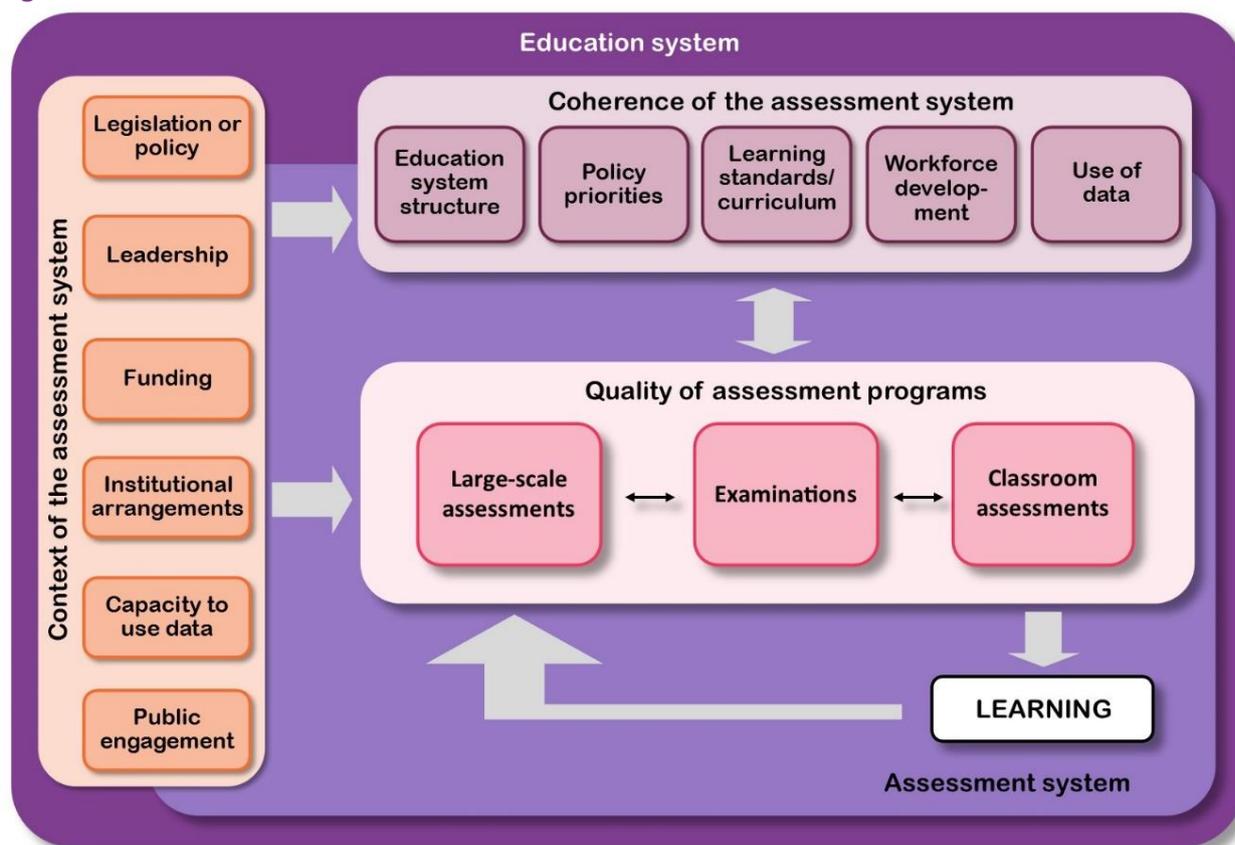
1.2 The ANLAS model

The ANLAS model illustrated in Figure 1 provides the analytical framework for the qualitative analysis of national learning assessment systems.

⁵ John C. Cresswell, "System-Level Assessment and Educational Policy" (Melbourne: Australian Council for Educational Research, 2017).

⁶ Global Partnership for Education, "Concept Note: Analysis of National Learning Assessment Systems (ANLAS)" (Washington, DC: Global Partnership for Education, 2018).

Figure 1: The ANLAS model



The qualitative analysis includes the following *three focus areas*:

1. **Context** refers to the broader context of the assessment system and the extent to which it is supportive of the assessment system.⁷ Context covers six key areas as shown in the ANLAS model.
1. **Coherence** refers to the extent to which the assessment system is aligned with, or coherent with important aspects of the broader education system.⁸ Coherence covers five key areas as shown in the ANLAS model.
2. **Quality** refers to the overarching key quality concepts for learning assessment programs: technical rigor, fitness for purpose, clarity and consistency of purpose objectivity and independence, transparency and accountability, and ethicality and fairness. The assessment programs covered in ANLAS include large-scale assessments (national, international/regional), examinations and classroom assessments. Eight key areas are

⁷ Marguerite Clarke, "What Matters Most for Student Assessment Systems: A Framework Paper," Systems Approach for Better Education Results (SABER) Student Assessment Working Paper No. 1 (Washington, DC: World Bank, 2012); Network for African Learning Assessments (NALA), "Learning Assessments Systems Evaluation Framework (Draft)," accessed January 4, 2019, http://www.adeanet.org/adeapmp/sites/default/files/activities/learning_assessments_systems_framework_v18.pdf.

⁸ Clarke, "What Matters Most."

operationalized for quality of large-scale assessments and examinations, and seven key areas for quality of classroom assessments.

An important **cross-sectional element** of ANLAS is the domain of **21st-century skills**. These skills—or “general capabilities” and “transversal competencies,” as they are often referred to—are an increasing policy priority for education systems around the world.⁹ The diagnostic nature, the cross-sectional consideration of 21st-century skills and the integration of the analysis into the education sector planning process are essential, distinct features of the initiative.

2 ANLAS in Ethiopia

Education is used as a means for making the nation competent in the current knowledge-based economy. As such, Ethiopia has been making every effort for Quality Education for All and ensuring that learning is taking place. Ethiopia has been conducting national large-scale sample-based assessments at particular grades to check students’ achievement levels in a timely and technically defensible manner to produce relevant data that help monitor and improve the quality of education. National Examinations, which are census based, have also been practiced for many decades to select or certify students as they move from one level of the education system to the next. Classroom assessment is also used to provide real-time information that supports ongoing teaching and learning.

To compare the country’s education system performance with other countries, participating in an international and/or regional assessment like the Program for International Student Assessment (PISA)/Trends in International Mathematics and Science Study (TIMSS) or the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) is highly valuable. However, for various reasons, Ethiopia has not yet participated in any international or regional assessments.

ANLAS provided a useful opportunity to undertake a comprehensive review of these different programs that make up Ethiopia’s national learning assessment system.

2.1 Implementation process

The implementation of ANLAS in Ethiopia was undertaken in six steps, commencing in July 2018 and ending in June 2019. Table 1 1 outlines the dates and duration of the six steps.

Table 1: ANLAS implementation process in Ethiopia

ANLAS implementation step	Date and duration
1. Nominating a focal point and establishing a national team	August 16–23, 2018 One week (the number of national team members was increased on January 23, 2019)
2. Familiarization with the ANLAS toolkit	January 14–23, 2019 9 days

⁹ Global Partnership for Education, “Concept Note.”

3. Team training and planning	January 23–25, 2019 3 days
4. Stakeholder briefings	January 29, 2019 1 day
5. Completing the analytical tables: description, evaluation, recommendations for improvement, and synthesis of findings	February 11–March 8, 2019 4 weeks
6. Preparing and disseminating a country report to relevant stakeholders to inform education sector planning	March 13–June 27, 2019 15 weeks

2.2 National team training and planning

The national team consisted of 12 members, including the focal point, representing the Ministry of Education (MoE) and NEAEA. The members, along with their roles and organizations, are listed in Appendix A.

The national team training was undertaken to develop a shared understanding of the ANLAS framework, tools and processes. The national team planning was aimed to develop a detailed ANLAS implementation plan and to identify the key stakeholder groups to involve.

2.3 Stakeholder briefings

The stakeholder briefings aimed to create a common understanding of the objectives and implementation process of ANLAS among the national team and key stakeholders. The stakeholder groups involved were federal and regional government education senior official representatives, NEAEA directors and senior experts, development partners and public relations officers. Thirty-two stakeholders participated in the briefing session held in the NEAEA hall in Addis Ababa, Ethiopia.

2.4 Completing the analytical tables

Analytical tables and instructions for completing them were provided to document and guide the analysis of the three ANLAS focus areas. Within each focus area, several key areas were analyzed by using sets of guiding questions. For each key area, a quality objective was defined, against which the key area was described and evaluated. These quality objectives are indicated as part of the ANLAS findings. The analysis of the national learning assessment system consisted of three major steps:

1. Describing the key areas for each ANLAS focus area
2. Evaluating the key areas for each ANLAS focus area, identifying aspects for improvement and making recommendations
3. Synthesis of the findings

This section describes these three steps. The findings from the analysis are presented in section

2.4.1 Describing the key areas of the national learning assessment system

The following assessment programs were included in the analysis:

1. National Examinations (for grade 10, certification of secondary school and grade 12 admission for higher educational institutes administered in each year)
2. Subnational Examinations, responsible for each region (grade 8, certification of primary school administered in each year)
3. National Learning Assessment (given for lower grades 2 and 3 like EGRA and EGMA and exit levels of primary and secondary at grades 4, 8, 10 and 12 for every four-year intervals)
4. Classroom assessments

No regional or international assessments are currently implemented in Ethiopia. However, the analysis provided some useful insights into future options for these assessments.

The description of the national learning assessment system took place alongside the evaluation of the key areas. The processes and sources of data used to undertake this analysis are described below.

2.4.2 Evaluating the key areas against quality objectives and making recommendations for improvement

The sources of data used in ANLAS were public documents and stakeholder consultations through interviews and focus group discussions (FGDs). Semi-structured interviews and FGDs were held with a total of 76 stakeholders representing education leaders, experts, teachers of primary and secondary schools, cluster supervisors, parents and development partners like USAID, UNICEF, DFID and the World Bank, as shown in Table 2. Primary schoolteachers and leaders were consulted from selected cluster schools in the three regions and one city administration: Oromia regional state (Qobo Luto Primary School), Amhara regional state (Chacha Primary School), SNNP regional state (Tiya Primary School) and Addis Ababa City Administration (Minelik II Primary School). Training providers also came from different regions' Colleges of Teacher Education (CTEs) (Asela CTE, Jigjiga CTE, Arbaminch CTE, Harari CTE, Hosaena CTE, Dilla CTE, Bonga CTE, Adwa CTE and Abiy Adi CTE). Seven senior experts from Regional Education Bureaus (REBs)—SNNP 3, Tigray 1, Oromia 2 and Somali 1—participated in the ANLAS FGDs.

The national team briefly explained the ANLAS purpose and its toolkits to the stakeholders and then the participants responded with their ideas. The national team facilitated and recorded their responses based on the ANLAS stakeholder mapping tables. Efforts were also made to triangulate the information provided by the interviewees using subsequent interviews or public documents. Furthermore, different documents like the current Education and Training Policy, policy frameworks for assessment and examinations, manuals, guidelines and assessment reports were used (see Appendix B).

Table 2: ANLAS consulted stakeholders

Number	Stakeholder group	Number of participants
1	National or subnational level officials	
	Federal Ministry of Education senior experts and directors	7
	NEAEA senior experts and directors	13
	Regional/Subnational Education Bureau senior experts and directors (SNNP 3, Tigray 1, Oromia 2 and Somali 1)	7
2	Training providers	
	Colleges of Teacher Education (Asela CTE, Jigjiga CTE, Arbaminch CTE, Harari CTE, Hosaena CTE, Dilla CTE, Bonga CTE, Adwa CTE and Abiy Adi CTE)	9
3	School leaders and teachers	
	Primary school leaders and supervisors	5
	Primary schoolteachers	19
	Secondary schoolteachers	9
4	Development partners	
	USAID	2
	UNICEF	2
	DFID	1
	World Bank	2
	Total	76

The stakeholders were contacted through letters and phone calls in order to gauge their interest in participating in the consultations. Following this, a blend of face-to-face contact/briefings and email/written communication were used in order to summarize the analytical tables.

The participants' responses were cross-checked and triangulated with pertinent official documents. Moreover, the data collectors (the national team) read the points or summary notes of the FGD to the participants for confirmation of consensus reached.

For each ANLAS focus area, its key areas were evaluated against quality objectives (see section 3). Three evaluation categories are differentiated as shown in Table 3.

Table 3: Evaluation categories

Category	The quality objective is ...
1	Achieved (currently no improvements needed)
2	Partly achieved (improvements can be made)
3	Not achieved (improvements are required)

For key areas in evaluation category 2 or 3, recommendations for improvement were identified. These recommendations aim to inform the development and implementation of improvement strategies as part of the wider education sector planning process.

2.4.3 Limitations

Though the national team composition was from relevant departments of the Ministry of Education and as a result there was no significant challenge in accessing pertinent documents for the assessment and coverage of the focus and key areas, the availability of the key stakeholders for consultations at various levels was a serious challenge. It was very demanding for the national team to get key informants for face-to-face discussion for more than half a day. To tackle this problem, various data collection strategies/modalities, such as identifying the relevant guiding questions from the analytical tables for the specific stakeholder group, were used. In addition, some key stakeholders couldn't attend the stakeholders' briefing and consultation session; therefore, the national team was engaged for additional time for briefings about the intention and implementation of ANLAS.

2.4.4 Synthesis

The information from the analysis of the three ANLAS focus areas was consolidated. This involved developing an overview table that describes the evaluation categories, the aspects and the recommendations for improvement for all three focus areas. This information is presented in section 3.2.

2.5 Preparing and disseminating a country report to relevant stakeholders to inform education sector planning

After completing the data collection, evaluation and synthesis, the national team organized the report into three focus areas. The national team also organized the necessary evidence from different official documents that support issues related to this diagnostic study.

This report was developed to present and disseminate the findings from the analysis to all stakeholders—government officials, development partners, senior education experts and teachers—and the public at large through panel discussion, mass media, and eventually the relevant websites (e.g. NEAEA and GPE websites).

The study findings have also been used to develop Ethiopia's sixth education sector development plan (ESDP VI) to fill the gaps of assessment system, using documents and discussion with policymakers and MoE officials, since September 2019. Further information

about how the ANLAS findings will be used is provided in section 4.

3 Findings of ANLAS

This section provides a summary of the outcomes of the qualitative analysis of the national learning assessment system in Ethiopia.

3.1 Evaluation of key areas

The three focus areas are (1) context of the assessment system, (2) coherence of the assessment system and (3) quality of large-scale assessments, examinations and classroom assessments. The evaluation of each key area in the three focus areas is summarized as follows.

The first focus area, context of the assessment system, has six key areas. One key area, institutional arrangements, was rated as achieved. The other five key areas—legislation/policy, leadership, funding, capacity to use assessment data and public engagement—were rated as partly achieved, indicating improvements may be needed in these areas.

The second focus area, coherence of the assessment system, has five key areas. The first key area, structure of the education system, was rated as achieved. The four other key areas (education policy priorities, learning standards and curriculum, school education workforce development and use of data) were rated as partly achieved, which may need some improvements.

The third focus area is about the quality of large-scale assessments, examinations and classroom assessments. The focus area has two parts (A and B). Part A refers to large-scale assessments and examinations. It consists of three subcomponents: national large-scale assessment (NLSA), national examinations and subnational examinations. The quality of large-scale assessment has eight key areas: (1) organization and resources, (2) assessment framework, (3) assessment instruments, (4) sampling, (5) field operations, (6) data management, (7) data analysis, and (8) reporting and dissemination.

In NLSA, five of the eight key areas (assessment instruments, sampling, field operations, data management and data analysis) were achieved and three (organization and resources, assessment framework, reporting and dissemination) were partly achieved. For the national examination, five key areas (assessment instruments, field operations, data management, data analysis, reporting and dissemination) were achieved and two key areas (organization and resources, assessment framework) were partly achieved. In the subnational examination, two key areas (field operations and reporting and dissemination) were achieved and the other five key areas (organization and resources, assessment framework, assessment instruments, data management and data analysis) were partly achieved.¹⁰

Part B refers to classroom assessment and has seven key areas: (1) guidelines, (2) training, (3) resources and tools, (4) assessment methods, (5) assessment content, (6) quality assurance and (7) use of assessment data. All the key areas were partly achieved. Table 4 shows the results of the evaluation categories for each ANLAS focus area and key area.

¹⁰ Since the national and subnational examinations are census-based, sampling is not an applicable key area.

Table 4: Evaluation results

1. Achieved (currently no improvements needed)	2. Partly achieved (improvements can be made)	3. Not achieved (improvements are required)
Focus area 1: Context of the assessment system		
4. Institutional arrangements	1. Legislation or policy	
	2. Leadership	
	3. Funding	
	5. Capacity to use assessment data	
	6. Public engagement	
Focus area 2: Coherence of the assessment system		
1. Structure of the education system	2. Education policy priorities	
	3. Learning standards and curriculum	
	4. School education workforce development	
	5. Use of data	
Focus area 3.A: Quality of large-scale assessments and examinations		
<i>Program 1 – National Large-Scale Assessment (NLSA)</i>		
3. Assessment instruments	1. Organization and resources	
4. Sampling	2. Assessment framework	
5. Field operations	8. Reporting and dissemination	
6. Data management		
7. Data analysis		
<i>Program 2 – National Examination</i>		
3. Assessment instruments	1. Organization and resources	
5. Field operations	2. Assessment framework	
6. Data management		
7. Data analysis		
8. Reporting and dissemination		

Program 3 – Subnational Examination		
5. Field operations	1. Organization and resources	
8. Reporting and dissemination	2. Assessment framework	
	3. Assessment instruments	
	6. Data management	
	7. Data analysis	
Focus area 3.B: Quality of classroom assessments		
	1. Guidelines	
	2. Training	
	3. Resources and tools	
	4. Assessment methods	
	5. Assessment content	
	6. Quality assurance	
	7. Use of assessment data	

3.2 Aspects of improvement and recommendations

Based on the analysis of the data obtained from different stakeholder groups and documents, some major gaps that need improvements were identified and possible solutions were recommended by the stakeholders in group discussions to enhance the effective implementation of the assessment systems. The findings, improvements and recommendations in each focus area are briefly discussed below.

Focus area 1: Context of the assessment system

The context of the assessment system comprises six key areas. Five of them were evaluated as partly achieved. One key area (institutional arrangement) was evaluated as achieved. The details for each key area are described below and summarized in Table 5.

With regard to legislation or policy, the data obtained from the analysis of stakeholders and document consultations revealed that the country has an Education and Training Policy. The policy guides the overall education system and policy frameworks, manuals and guidelines; internal directives are also developed for NLA and public examinations under the education policy. However, the analysis identified that there is a gap between policy provision and implementation of policy. In particular, there is inconsistency of the evaluation criteria in the policy document and the grading system in national examinations. For instance, while there are fixed criteria for evaluation in the education policy, norm-referenced grading system is practiced in grade 10.

According to the information obtained from the FGDs, there is encouraging political commitment by leadership on promoting and supporting the assessment system by developing

policy documents, allocating budget, establishing institutional arrangements that can help to carry out the assessment, communicating to the public about the purpose of the assessment and promoting different development partners to engage in the assessment system. In spite of this, the leadership has a limited capacity to use the assessment results for making evidence-based decision and intervention.

In regard to the key area of funding, it was found that the government covered only 24% of the NLSA budget; donor sources covered the rest (76%). However, the government allocated a greater budget amount (75%) for National Examinations, with the remaining 25% collected from private students. Moreover, the response of key informants showed that there is no clearly defined budget for classroom assessment. In conclusion, there is budget inadequacy for the large-scale assessment, national examinations and classroom assessment practices.

Regarding the institutional arrangements, the study found that the government has a separate institution (NEAEA) responsible for designing, analyzing, implementing and using data for evidence-based decision-making. In addition, various departments in the ministry are expected to document and utilize the assessment results for better intervention.

The purpose of the learning assessment is to determine the level of students' academic achievement, to identify factors influencing achievement and finally to generate recommendations for policymaking to improve education quality. The government has the capacity to use data from learning assessments for evidence-based decision-making in education policy and practice. However, there is a limitation in dissemination and utilization of the assessment results for intervention during formative evaluation.

Concerning the public engagement, there is a good experience in community participation to support the assessment system. However, there is a gap in having common understanding about the purpose of the assessment. Yet, there is a limitation in addressing updated guidelines and manuals for stakeholders like teachers, principals and supervisors.

Focus area 2: Coherence of the assessment system

Focus area two—coherence of the assessment system—comprises five key areas. Four of them were evaluated as partly achieved. One key area (structure of the education system) was evaluated as achieved. Aspects and recommendations for improvement were identified for each area that was partly achieved. These are described below and summarized in Table 5.

Concerning the second key area (education policy priorities), although 21st-century skills are well addressed in the Ethiopian curriculum framework, the assessment system as well as the implemented curriculum, like students textbooks and other instructional materials, are not articulated in line with the curriculum framework such that 21st-century skills are not well addressed. More specifically, these 21st-century skills are not adequately assessed within the large-scale assessment system since the instruments are only multiple-choice test items, which are not suitable to assess most performance skills. Therefore, it is recommended that there is a greater focus to incorporate the constructed response test items that are helpful to assess the 21st-century skills in the future. There is also a lack of fair alignment among the assessment programs to provide relevant information for policy priorities. The education management information system (EMIS) includes only the performance data from the national and regional examinations in its annual report. Therefore, it is recommended that the EMIS should organize and provide the aligned data from major NLA findings and classroom assessment results.

With regard to learning standards and curriculum, aligning classroom assessment with the curriculum standards (competencies) and making continuous classroom assessment part of the instruction are key areas that require improvement. Attempts should be made to assess language practical skills and activities like listening, writing and speaking in language as well as laboratory and other hands-on performances. The data from classroom assessment should be properly documented and utilized for further investigations of policies, standards, curriculum and instructional methods that the teacher used. In this case, provision of capacity building and development of appropriate instructional materials and improvement of the assessment system are suggested.

Concerning school education workforce development, alignment of the assessment programs with national official documents and integrating use of data with capacity-building programs are issues that need improvement, which is why the students' results in classroom assessment and national assessments were significantly different according to the respondents. This was supported by the discrepancies noted between scores in each cycle of the NLA at each grade level and performance in classroom assessments. As recommendations, aligning training materials with the official documents and incorporating the findings of the assessment data in capacity-building programs were forwarded.

Assessment data are used by various key stakeholders for evidence-based decision-making in education policy and practice. However, producing quality classroom data and relevant data to particular regions according to their contexts, disseminating the findings to all stakeholders on time (particularly to the lower levels of education systems such as zones, districts and schools), limitation in using the data (recommendations) (for example, curriculum revisions and instructional improvements), and establishing accountability to use the data from the assessment systems were areas identified as drawbacks.

Hence, producing quality and relevant data, improving its accessibility, establishing policies and regulations regarding accountability, and providing appropriate trainings on using data are recommended.

Focus area 3.A: Quality of large-scale assessments and examinations

Large-scale assessment in Ethiopia's context indicates National Large-Scale Assessment (NLSA) programs, including EGRA and EGMA, supported by government and donors. It is carried out to provide information on education system accomplishment and the factors contributing to educational outcomes, to inform education policy and practice. Large-scale assessments in Ethiopia use a sample of the target population.

Similarly, examination in Ethiopia's perspective describes both National and Subnational Examinations under the auspices of the federal government. National Examinations are conducted at the national level to make decisions for certification at the end of grade 10 and for university entrance selection purposes at the end of grade 12. Subnational Examinations are conducted at subnational level to make decisions for certification at the end of grade 8. The National and Subnational Examinations use a census approach to test all eligible students.

In this part of analysis, within "quality of large-scale assessments and examinations," evaluations against quality objectives for eight key areas have been made. Accordingly, in the NLSA program, five key areas—namely, assessment instruments, sampling, field operations, data management and data analysis—were informed by different stakeholders to be achieved;

the three key areas of organization and resources, assessment framework, and reporting and dissemination were described as being partly achieved.

In the National Examination program, five key areas—assessment instruments, field operations, data management, data analysis, and reporting and dissemination—were identified as being achieved. Organization and resources and assessment frameworks are the two key areas where improvements could be made in the National Examination.

In the Subnational Examination program, the two key areas of field operations and reporting and dissemination were determined by the stakeholders to be achieved. On the other hand, organization and resources, assessment framework, assessment instruments, data management and data analysis were deemed to be partly achieved.

Based on the data-gathering operation in the large-scale assessment and examination programs, no key area was categorized as not achieved.

In addition to the aforementioned evaluation part, the following recommendations for improvement in the eight key areas have been forwarded. Organization and resources is a common key area that requires improvement in the three large-scale assessment programs (that is, the NLSA, National Examination, and Subnational Examination programs). It was discovered that there have been limited capacities in the assessment management and staff and inadequate physical facility provisions to complete the diverse task in the assessment programs. Thus, it was recommended that assessment management and staff should be strengthened and should upgrade their skills in the assessment and related field with short- or long-term training in collaboration with government and donors. In addition, the government should provide adequate office rooms, essential software and scanners to the assessment staff in the three large-scale assessment programs.

Assessment framework is also a common key area that was partly achieved in the three large-scale programs. There is no assessment framework prepared for the NLSA, National Examination and Subnational Examination programs that provides a rationale for contextual data collected and specifies the design of the assessment program. Hence, it was recommended that an assessment framework should be prepared for the programs.

Assessment instruments is a key area common to the NLSA and National Examination programs that currently requires no improvement. However, this key area was recognized as partly achieved in the Subnational Examination program: It was discovered that there was no piloting of test instruments to ensure the reliability, validity and fairness of the tests. Therefore, it was recommended that reviewing and pre-testing of test instruments ought to be in place in the subnational program.

The key area of sampling presently needs no improvement in the NLSA program. This key area was not an issue in the National and Subnational Examination programs because the data-gathering approach is census based.

Field operations is the one common key area that was achieved in all three large-scale assessment programs.

The key area of data management needs no improvement in the NLSA and National Examination programs; however, this key area was only partly achieved in the Subnational Examination program. It was identified that there is no data management manual for Subnational

Examination that ensures appropriate quality data structure and documentation. Thus, it was suggested that a data management manual should be prepared for the Subnational Examination program.

Data analysis is a key area that needs no improvement in the NLSA and National Examination programs; however, it was only partly achieved in the Subnational Examination program. There is no sound data analysis technique in the Subnational Examination to provide valid and useful analytical inferences about the population of interest; therefore, a recommendation was made that the subnational education bureaus should provide scoring and data processing machines to the Subnational Examination program.

Reporting and dissemination is the other common key area that needs no improvement in the National and Subnational Examination programs. However, improvement could be made in this area in the NLSA program for the effective use of assessment data and results to different stakeholder groups. Hence, it was recommended that the NLSA findings should be disseminated in a timely manner to different stakeholders according to their needs.

International assessment

As indicated in the ESDP V plan, Ethiopia planned to join both regional and international assessments, but that has not happened yet. So, the analysis of international assessment against the quality indicators of ANLAS was not able to be conducted.

However, during stakeholder consultations, the participants emphasized the importance of regional and international assessments for monitoring the quality of the Ethiopian education system. Hence, it is recommended that Ethiopia should join the SACMEQ regional assessment and either the TIMSS or PISA international assessment to receive additional information about the education system from an international, comparative perspective and to learn from other systems of similar as well as different contexts.

Focus area 3.B: Quality of classroom assessments

Concerning the overall quality of classroom assessment practices, consultations with key stakeholders and document analysis revealed that all seven key areas were partly achieved and that there are gaps that require immediate action. Some of the identified gaps include the following:

- Problem of accessibility of the guidelines
- Shortage of resources and tools such as curriculum materials, especially syllabi and teachers' guides
- Limited or no use of some assessment methods (example: interviews, questionnaires, peer assessment, self-assessment)
- Lack of 21st-century skills as a package
- Weak mechanism of system-level monitoring of the quality of classroom assessment
- Limited use of classroom assessment data for school interventions and improvement of students' learning

Regarding guidelines, there is a problem of accessibility for most secondary and primary schoolteachers; however, the Ministry of Education and Regional Education Bureaus should be able to address the demand of schools by printing and distributing the guidelines. There is also a

need to develop new guidelines that incorporate practical examples in each specific domain/subject for both primary and secondary schools. The guidelines should also incorporate 21st-century skills.

The capacity building for teachers and school leaders (denoted as the 'training' key area) is regularly conducted through pre-service and in-service programs. However, a significant proportion of teachers and school leaders do not benefit from these programs. There is a lack of refresher and timely trainings. Thus, it is recommended that all teachers should get the opportunity of trainings through in-service programs. In addition, refresher and timely trainings (hands-on trainings) on classroom assessment practices should be organized and conducted at the national and subnational levels. Similar trainings are required to be in place at the cluster resource centers for both secondary and primary schools. Awareness creation programs for teachers about the importance of classroom assessment practices should be made.

Concerning the provision of resources and tools to schools, key informants confirmed shortages of curriculum materials, in particular, teachers' guides and syllabi. Thus, it is recommended that education offices at each level need gathering/predicting data of all learners, proper planning, and timely printing and distribution of curriculum materials to all schools.

Assessment methods and assessment content are areas that require improvement. As it was confirmed by key informants, only assessment methods such as quizzes and teacher-made tests are commonly used in most schools. Assessment methods such as interviews, questionnaires, peer assessment, use of standardized tests and self-assessment are not commonly practiced. Regarding assessment content, there is a lack of inclusion of 21st-century skills in the curriculum and/or lack of assessment of 21st-century skills. To fill this gap, it is advisable to enrich the curriculum with 21st-century skills methods such as critical thinking/reasoning, creativity/creative thinking, problem solving, metacognition, collaboration, communication and global citizenship, information literacy, ICT digital literacy and communication and incorporate these skills in the assessment system.

Another key area that require improvement is monitoring mechanisms (denoted as the 'quality assurance' key area). The study indicated that there is a weak system-level mechanism of monitoring the quality of classroom assessment practices. Hence, it is recommended that strong system-level mechanisms of monitoring the quality of classroom assessment should be established.

In regard to the last key area with respect to the use of assessment data, its use by teachers and education officials at each level to modify their teaching and make appropriate interventions, respectively, is not satisfactory. The capacity of teachers should be enhanced on how to use assessment data to modify their teaching and improve students' learning. Similarly, the concerned bodies at national and subnational levels should use classroom assessment data for school interventions. Hence, classroom assessment practices should be used for teacher supervision, teacher performance evaluations, external moderation systems and regular reviews of the quality of education and awarding government funding for research and development. In addition, the MoE, REBs and education officials at each level should enhance the capacity of teachers on how to use assessment data to modify their teaching and improve students' learning and establish systems to enable this, as well as to allow school leaders to develop goals and programs for school improvement and to report to the district officials that manage and support schools.

Table 5 lists the evaluation category, aspects of improvement, and recommendations identified for each of the three ANLAS focus areas.

Table 5: Summary of recommendations

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 1: Context of the assessment system			
<p>1. Legislation or policy The assessment system is guided by legislation or policy.</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> • The policy frameworks, manuals and guidelines that guide the assessment system are not enacted by education law. • There is inconsistency between the policy provision and practices; particularly, there is inconsistency of the criteria set in policy documents that students are required to score a minimum of 50% achievement level with that of the norm-referenced approach used by the grading system. • Lack of subject-specific assessment guidelines at school level. 	<ul style="list-style-type: none"> • Federal education law should be developed to strengthen and protect the assessment systems and resolve disputes that may arise within. • The education policy objectives should be revised and reformulated in a way to react the identified gaps. • Consistent evaluation and grading system with the criteria set in the education policy must be applied in public examinations. • Up-to-date subject-specific assessment guidelines and standards must be addressed to all schools.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 1: Context of the assessment system			
<p>2. Leadership</p> <p>The government demonstrates senior leadership and political will in support of the assessment system. The importance of the assessment system is recognized by all key stakeholder groups.</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> • Lack of appropriate management system that can respond to the increasing number of subjects in national examination, particularly in mother tongue languages. • Some officials are assigned to areas of technical expertise by political decision and not based on competence, which induces unenthusiastic impact to deliver the intended mission. • Ministry of Education (MoE) has given little attention to the assessment system, especially in funding, disseminating, communicating and utilization of assessment results. 	<ul style="list-style-type: none"> • Number of subjects for national examination must be fixed nationally in order to maintain fairness and quality assessment. • Government should establish a merit-based system of hiring education officials so that they can give technical and professional support in the area. • MoE should establish strong monitoring and support system for assessment programs on funding, disseminating, communicating and utilization of assessment results.
<p>3. Funding</p> <p>The government provides sufficient and stable funding for the assessment system.</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> • Inadequacy of budget for the design, implementation, analysis, printing and dissemination of practices. • Lack of budget for assessment capacity building and infrastructure. 	<ul style="list-style-type: none"> • The government should allocate appropriate budget for effective assessment practices, capacity building and technology infrastructure for the assessment system.
<p>4. Institutional arrangements</p> <p>The government has institutional arrangements in place for designing, implementing, analyzing and using data from various learning</p>	<p>1. Achieved</p>		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 1: Context of the assessment system			
assessments.			
<p>5. Capacity to use assessment data</p> <p>The government has the capacity to use data from learning assessments for evidence-based decision-making in education policy and practice.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> There is skill gap in utilization of different software's for analyzing, interpreting assessment data. Limited capacity of the curriculum development and implementation department in utilization of assessment findings for formative evaluation. Structural gap of EMIS in the MoE regarding documentation and dissemination of data results. 	<ul style="list-style-type: none"> Continuous training on utilization of software applications must be given to experts in NEAEA. Intensive awareness and capacity-building programs must be given to education officials and/or experts of curriculum department on the importance and utilization of the assessment results for formative evaluation. MoE needs to develop a system that enhances the capacity of EMIS in a way that information is efficient, effective and accessible to all key stakeholders
<p>6. Public engagement</p> <p>The public understands, is engaged in and supports the assessment system.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Not all key stakeholders have a common understanding on the purpose and importance of National Learning Assessment. Communication gap in disseminating assessment results. 	<ul style="list-style-type: none"> MoE should give an intensive orientation and awareness for key stakeholders on the purpose of national assessments. MoE/NLA department should sufficiently communicate the assessment practices and results via print and nonprint medias.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 2: Coherence of the assessment system			
<p>1. Structure of the education system</p>	1. Achieved		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 2: Coherence of the assessment system			
The assessment system provides performance data in key learning domains and relevant contextual data, at key stages of primary and secondary school education, and for relevant levels of the education system.			
<p>2. Education policy priorities</p> <p>The assessment system provides relevant data to inform education policy priorities. The assessment programs that form part of the assessment system are aligned to jointly provide the relevant data to inform education policy priorities.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> 21st-century skills of learning domains are indicated in the curriculum framework. However, they are not properly addressed in the assessment systems since both NLA and national/regional examinations are using only multiple-choice test items, which do not give the room for assessing these skills. Data from NLSAs and classroom assessments are not well aligned to inform education policy priorities. 	<ul style="list-style-type: none"> 21st-century skills such as problem-solving skills, critical thinking, communication skills and teamwork in the learning domains should be more incorporated in the assessment system, using constructed response test items in the future. EMIS should design a mechanism to collect and organize classroom assessment data to align with NLSAs to inform the system for evidence-based decision-making.
<p>3. Learning standards and curriculum</p> <p>The learning domains/ subjects in the assessment system are aligned with official learning standards and/or curriculum.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Gap in teachers' skills to align classroom assessment with the curriculum standards (competencies). Continuous classroom assessment is not currently implemented as part of teachers' instruction. Some competencies are not assessed in large-scale assessments, such as listening, speaking, practical activities. 	<ul style="list-style-type: none"> Intensive capacity building should be provided to teachers and school leaders on classroom assessment techniques and their alignment with curriculum standards (competencies). Instructional materials should be developed in such way that they incorporate more guidance for continuous classroom assessment.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 2: Coherence of the assessment system			
			<ul style="list-style-type: none"> NEAEA and the Curriculum Development and Implementation Department should devise a way to assess 21st-century skills.
<p>4. School education workforce development</p> <p>Training and capacity-building programs for the school education workforce are aligned with the legislation or policy framework for the assessment system, and official learning standards and/or curriculum that guide the assessment system.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> There is a gap in alignment between training and capacity-building materials with national official documents like curriculum documents. Lack of using assessment data for capacity-building programs. School principals and teachers do not consider continuous classroom assessment as part of instructional process; rather, it is assumed as a summative and additional task. This is due to lack of capacity and attitude to assessment. Lack of monitoring and evaluation to check the impact of trainings provided on assessment programs. 	<ul style="list-style-type: none"> Teacher training institutions and the MoE should prepare classroom assessment training materials that align with official documents. Capacity-building materials should incorporate the findings of the assessment data. Teachers and educational leaders' development directorate should work on enhancing capacity building and raise awareness of teachers and school leaders on assessment programs. MoE should devise a mechanism for monitoring and evaluating the impact of training on assessment programs.
<p>5. Use of data</p> <p>Assessment data is used by various key stakeholders for evidence-based decision-making in education policy and practice.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Lack of timely dissemination of assessment results to various key stakeholders. Lack of accountability for the assessment findings. Limited use of assessment data for intervention. 	<ul style="list-style-type: none"> NEAEA should improve accessibility of data that address the needs of various stakeholders in a timely manner. MoE should establish policies and accompanying rules and regulation for accountability. MoE should provide training/awareness creation on using data for various stakeholders.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
<i>Program 1 – National Large-Scale Assessment (NLSA)</i>			
<p>1. Organization and resources</p> <p>There is an assessment team with dedicated staff who are appropriately skilled and adequately resourced to complete the diverse tasks associated with key assessment program areas.</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> Currently, there is limited assessment capacity of management, experts and staff (in the area of assessment-related fields like psychometrics and special need) responsible for the NLSA. At present, there is inadequate physical facilities provision in the NLSA program. 	<ul style="list-style-type: none"> Assessment management body, experts and staff in the NLSA program should strengthen their skills by training in the field of assessment and related fields, be it short or long term, with the help of government and donors. Office rooms, different essential software, scanners should be provided for assessment staff in the NLSA program by the government.
<p>2. Assessment framework</p> <p>There is documentation that clearly describes the knowledge and skills to be assessed in the key learning domains/subjects, provides a rationale for the contextual data collected and specifies the design of the assessment program. The document is made available to individuals/groups who are involved in assessment instrument development, key stakeholders and the</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> At present, there is no assessment framework for the NLSA program. But there is a policy framework for the NLSA program. 	<ul style="list-style-type: none"> Assessment framework should be prepared for NLSA program by NLSA for future use.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
public.			
3. Assessment instruments Quality assurance mechanisms are in place to ensure the test and any contextual data collection instruments are reliable, valid and fair.	1. Achieved		
4. Sampling The sample, through the use of scientific sampling methods, helps to guarantee appropriate and estimable levels of statistical precision and validity in the interpretation of assessment results.	1. Achieved		
5. Field operations Quality assurance mechanisms are in place to ensure field operations are standardized, monitored and documented, to ensure that the data are collected under the same conditions, independent of	1. Achieved		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
the administration context, in an efficient and secure manner.			
6. Data management Quality assurance mechanisms are in place to ensure the final database is free from discrepancies and errors, appropriately structured and documented.	1. Achieved		
7. Data analysis Technically sound and appropriate data analysis techniques are used to provide analytical results that permit valid and useful inferences about the population(s) of interest. Analytical results are fully documented and reproducible.	1. Achieved		
8. Reporting and dissemination Appropriate products and approaches to reporting and dissemination are	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> The dissemination strategy is not effective to address the use of NLSA data and results. The NLSA findings could not be disseminated to different stakeholders on time. 	<ul style="list-style-type: none"> Dissemination strategy should address the effective use of assessment data and results to different stakeholder groups. The NLSA findings should be disseminated in a timely manner to different stakeholders according to their

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
tailored to the different stakeholder groups and promote appropriate and effective use of the assessment data and results by those groups.			needs.
<i>Program 2 – National Examination</i>			
<p>1. Organization and resources</p> <p>There is an assessment team with dedicated staff who are appropriately skilled and adequately resourced to complete the diverse tasks associated with key assessment program areas.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> At present, there is limited assessment capacity of management, experts and staff (in the area of assessment-related fields like psychometrics and special need) responsible for the National Examination. At present, there is inadequate physical facilities provision in the National Examination program. 	<ul style="list-style-type: none"> Assessment management body, experts and staff in the National Examination program should strengthen their skills by training in the field of assessment and related fields, be it short or long term, with the help of government and donors. Office rooms, different essential software, scanners should be provided by the government to assessment staff in the National Examination program.
<p>2. Assessment framework</p> <p>There is documentation that clearly describes the knowledge and skills to be assessed in the key learning domains/subjects, provides a rationale for the contextual data collected and specifies the design of the assessment program.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Currently, there is no assessment framework for the National Examination program. However, there is a policy framework for the National Examination program. 	<ul style="list-style-type: none"> Assessment framework should be prepared for the National Examination program by the National Examination program for future use.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
The document is made available to individuals/ groups who are involved in assessment instrument development, key stakeholder groups and the public.			
3. Assessment instruments Quality assurance mechanisms are in place to ensure the test and any contextual data collection instruments are reliable, valid and fair.	1. Achieved		
4. Sampling The sample, through the use of scientific sampling methods, helps to guarantee appropriate and estimable levels of statistical precision and validity in the interpretation of assessment results.	As the National Examination followed a census approach to collect data, this key area was not applicable.		
5. Field operations Quality assurance mechanisms are in place to	1. Achieved		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
<p>ensure field operations are standardized, monitored and documented, to ensure that the data are collected under the same conditions, independent of the administration context, in an efficient and secure manner.</p>			
<p>6. Data management Quality assurance mechanisms are in place to ensure the final database is free from discrepancies and errors, appropriately structured and documented.</p>	<p>1. Achieved</p>		
<p>7. Data analysis Technically sound and appropriate data analysis techniques are used to provide analytical results that permit valid and useful inferences about the population(s) of interest. Analytical results are fully documented and reproducible.</p>	<p>1. Achieved</p>		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
8. Reporting and dissemination Appropriate products and approaches to reporting and dissemination are tailored to the different stakeholder groups and promote appropriate and effective use of the assessment data and results by those groups.	1. Achieved		
<i>Program 3 – Subnational Examination</i>			
1. Organization and resources There is an assessment team with dedicated staff who are appropriately skilled and adequately resourced to complete the diverse tasks associated with key assessment program areas.	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> There is currently limited assessment capacity of management, experts and staff (in the area of assessment-related fields like psychometrics and special need) responsible for the Subnational Examination. At present, there is inadequate physical facilities provision in the Subnational Examination program. 	<ul style="list-style-type: none"> Assessment management body, experts and staff in the Subnational Examination program should strengthen their skills in the field of assessment and related fields by training, be it short or long term, with the help of national states, federal government and donors. Office rooms, different essential software, scanners should be provided by the national states to assessment staff in the Subnational Examination program.
2. Assessment framework There is documentation that clearly describes the knowledge and skills to be	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Currently, there is no assessment framework for the Subnational Examination program. 	<ul style="list-style-type: none"> Assessment framework should be prepared for the Subnational Examination program by the Subnational Examination program for future use.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
<p>assessed in the key learning domains/subjects, provides a rationale for the contextual data collected and specifies the design of the assessment program. The document is made available to individuals/groups who are involved in assessment instrument development, key stakeholder groups and the public.</p>			
<p>3. Assessment instruments Quality assurance mechanisms are in place to ensure the test and any contextual data collection instruments are reliable, valid and fair.</p>	<p>2. Partly achieved (Improvements are required)</p>	<ul style="list-style-type: none"> At present, there is no piloting of test instruments in the Subnational Examination program. 	<ul style="list-style-type: none"> Reviewing and pre-testing of test instruments ought to be in place in the Subnational Examination program to ensure the tests are reliable, valid and fair.
<p>4. Sampling The sample, through the use of scientific sampling methods, helps to guarantee appropriate and estimable levels of statistical precision and validity in the</p>	<p>As the Subnational Examination program followed a census approach to collect data, this key area was</p>		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
interpretation of assessment results.	not applicable.		
5. Field operations Quality assurance mechanisms are in place to ensure field operations are standardized, monitored and documented, to ensure that the data are collected under the same conditions, independent of the administration context, in an efficient and secure manner.	1. Achieved		
6. Data management Quality assurance mechanisms are in place to ensure the final database is free from discrepancies and errors, appropriately structured and documented.	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Currently, there is no data management manual in the Subnational Examination program. 	<ul style="list-style-type: none"> Data management manual should be developed to ensure the final database is free from error and appropriately documented in the Subnational Examination program.
7. Data analysis Technically sound and appropriate data analysis techniques are used to provide analytical results	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> At present, there is a lack of marking, scoring and data-processing machines in the Subnational Examination program. 	<ul style="list-style-type: none"> The Subnational Education Bureau should provide marking, scoring and data-processor machines that help provide valid and useful inferences of the target population in the Subnational Examination program.

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.A: Quality of large-scale assessments and examinations			
that permit valid and useful inferences about the population(s) of interest. Analytical results are fully documented and reproducible.			
<p>8. Reporting and dissemination</p> <p>Appropriate products and approaches to reporting and dissemination are tailored to the different stakeholder groups and promote appropriate and effective use of the assessment data and results by those groups.</p>	1. Achieved		

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.B: Quality of classroom assessments			
<p>1. Guidelines</p> <p>There are national- or subnational-level documents that provide guidelines or recommendations for classroom assessment.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> • Problem of accessibility of the centrally developed classroom assessment guidelines or manuals for schools. • Lack of classroom assessment guidelines with practical examples specific to each subject at primary and secondary schools. 	<ul style="list-style-type: none"> • The guidelines are required by all secondary and primary schools. Thus, the concerned bodies at the national and subnational levels should try to address the demand of schools by printing and distributing the guidelines. • There is also a need for developing guidelines that incorporate practical examples in each specific domain/subject at both primary and secondary schools. The guidelines should also incorporate 21st-century skills as a package.
<p>2. Training</p> <p>Training programs are provided to (future and current) teachers and school leaders to build their capacity to assess students' learning.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> • Lack of refresher and timely training focusing on classroom assessment practices. • Lack of training in the importance of classroom assessment. 	<ul style="list-style-type: none"> • Refresher and timely trainings (hands-on trainings) on classroom assessment practices should be organized and conducted at national and subnational levels. Similar trainings are required to be in place at the cluster resource centers for both secondary and primary schools. • Awareness creation for teachers about the importance of classroom assessment practices <i>should be made</i>.
<p>3. Resources and tools</p> <p>Teachers use a variety of resources and tools for classroom assessment.</p>	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> • Shortage of curriculum materials (syllabi, teachers guide, textbooks) 	<ul style="list-style-type: none"> • Adequate funding is required for printing and distribution of curriculum materials. Education offices at each level need gathering/predicting data of all learners, proper planning, and timely printing and distribution of curriculum materials to all schools. Thus, both federal and regional governments should allocate sufficient budget to alleviate this problem.
<p>4. Assessment methods</p>	2. Partly	<ul style="list-style-type: none"> • There is limited or no use of multiple methods, 	<ul style="list-style-type: none"> • Teachers should be trained on how to use different

Key areas	Evaluation category	Aspects for improvement	Recommendations for improvement
Focus area 3.B: Quality of classroom assessments			
Teachers use multiple assessment methods to support valid and reliable assessment of students' learning.	achieved (Improvements are required)	especially interviews, standardized tests, student self-assessment and questionnaires, by teachers in both primary and secondary schools.	assessment methods. To do this, both teacher training institutions and Regional Education Bureaus should work together to enhance the capacity of teachers on classroom assessment practices.
5. Assessment content The knowledge and skills assessed are clearly defined.	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> 21st-century skills were not sufficiently incorporated in the curriculum materials. The knowledge and skills to be assessed in the art are not clearly defined in primary schools. ICT is not part of the school curriculum in primary schools. 	<ul style="list-style-type: none"> 21st-century skills should be incorporated as a package in curriculum. The knowledge and skills to be assessed in the art should be clearly defined. ICT should be part of the school curriculum in primary schools.
6. Quality assurance There are formal mechanisms at the school and national/subnational levels to monitor the quality of classroom assessment practices.	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Monitoring mechanism is not intended to promote the quality of classroom assessment practices. Rather, it is for the sake of evaluating a teacher's performance. Weak system of monitoring the quality of classroom assessment practices 	<ul style="list-style-type: none"> Strong system level of monitoring the quality of classroom assessment practices should be established by MoE. The monitoring mechanism should aim to promote the quality of classroom assessment practices in schools. It should not be for the sake of evaluating a teacher's performance.
7. Use of assessment data Data from classroom assessment are used to improve teaching and learning.	2. Partly achieved (Improvements are required)	<ul style="list-style-type: none"> Lack of awareness of teachers in identifying learning gaps and modifying their teaching practices. Lack of appropriate interventions by using classroom assessment data. 	<ul style="list-style-type: none"> Awareness creation for teachers on how to use the assessment data to identify learning gaps and modify their teaching practices. School leaders and local educational authorities should plan and implement appropriate interventions based on the data from classroom assessments.

4 Using the ANLAS findings

ANLAS is designed to inform the development and implementation of improvement strategies as part of the wider education sector planning process. This section first explains the national education sector planning process in Ethiopia, then it describes how the ANLAS findings are planned/envisaged to be used to inform education sector planning.

As required by the current Education and Training Policy (ETP) of 1994, education sector plans (ESPs) known as medium-term Education Sector Development Programs (ESDPs) have been developed and implemented in Ethiopia, each lasting a set period of time. The current ESDP V is Ethiopia's fifth medium-term plan, serving as the central strategy document for educational development in the country from 2015/16 to 2019/20. To monitor and improve the quality education, an attempt has been made to measure education outcomes through a robust assessment system since ESDP IV.

The ANLAS model could help as a theoretical foundation to adapt and adopt for the Ethiopian education sector in general and the assessment system in particular. It can be used to inform the entire policy cycle, including education sector analysis, planning and endorsement, implementation, monitoring and sector reviews, and evaluation.

Therefore, this ANLAS diagnostic study will be used to develop the next ESDP, ESDP VI, which will be implemented after 2020. The ANLAS report can also be used as a source document for curriculum revision, teacher training institutions, and teacher continuous professional development (CPD). Table 6 provides details of the time frames for the current and next education sector plans.

Table 6: Education sector planning process

Current education sector plan (ESDP V)	2015/16–2019/20
Next education sector plan (EDSP VI)	2020/21–2024/25
Relevant education sector planning stages to consider ANLAS recommendations	2020/21

Appendix A: The national team

Table A1: The national team

Name	ANLAS Role	Organization
Mesaye Demessie Zeleke	Focal Point	Deputy Director General, NEAEA
Yilikal Wondimeneh Demissie	National Team Member	National Learning Assessment Directorate, NEAEA
Arega Mamaru Yewore	National Team Member	National Examination Development and Administration Directorate, NEAEA
Nega Gichile Bongasse	National Team Member	National Curriculum Development and Implementation Directorate, MoE
Abraham Mengistu Sertse	National Team Member	Planning and Resource Mobilization Directorate, MoE
Talefe Eshete Awoke	National Team Member	Teachers and Educational Institutions Leaders Licensing Directorate, MoE
Aregawi Gidey Mesele	National Team Member	National Learning Assessment Directorate, NEAEA
Bekele Geleta Ayana	National Team Member	National Learning Assessment Directorate, NEAEA
Effa Gurmu Bati	National Team Member	National Learning Assessment Directorate, NEAEA
Getachew Abebe Tsigie	National Team Member	National Examination Development and Administration Directorate, NEAEA
Belay Endeshaw Gizaw	National Team Member	National Examination Development and Administration Directorate, NEAEA
Robel Getachew Worku	National Team Member	National Examination Development and Administration Directorate, NEAEA

Appendix B: Stakeholder and document mapping tables

Table B1: Key stakeholder groups

Key stakeholder group	Focus area 1 Context	Focus area 2 Coherence	Focus area 3.A Quality of large-scale assessments and examinations	Focus area 3.B Quality of classroom assessments
National or subnational level officials				
• Senior curriculum development and implementation experts	✓	✓	✓	✓
• Senior educational assessment experts	✓	✓	✓	✓
• Senior examination development experts	✓	✓	✓	✓
• Senior examination data processor experts	✓	✓	✓	✓
• Regional and zonal educational experts	✓	✓	✓	✓
• Senior planning and resource mobilization experts	✓	✓	✓	✓
Training program providers				
• Teachers from Colleges of Teacher Education	✓	✓	✓	✓
School leaders and teachers				
• Senior teachers from primary schools	✓	✓	✓	✓
• Senior teachers from secondary schools	✓	✓	✓	✓
• Primary and secondary school principals	✓	✓	✓	✓
Development partners				
• World Bank senior education expert	✓	✓	✓	✓
• DIFD senior education expert	✓	✓	✓	✓
• UNICEF senior education expert	✓	✓	✓	✓

Table B2: Key documents

Document	Focus area 1 Context	Focus area 2 Coherence	Focus area 3.A Quality of large-scale assessments and examinations	Focus area 3.B Quality of classroom assessments
Education and Training Policy, MoE, 1994	✓	✓	✓	✓
Curriculum Framework for Ethiopian Education (KG-Grade 12), MoE, 2010	✓	✓	✓	✓
Policy Framework for National Large-Scale Assessment, NEAEA, 2012	✓	✓	✓	✓
Policy Framework for National Examinations Assessment, NEAEA, 2012	✓	✓	✓	✓
Ethiopia Early Grade Reading Assessment/ English, USAID/RTI, 2010	✓	✓	✓	✓
Ethiopia Early Grade Reading Assessment/ Mother Tongue, USAID/AIR, 2010, 2014, 2018	✓	✓	✓	✓
Early Grade Mathematics Assessment: Baseline Study Report, NEAEA, 2014	✓	✓	✓	✓
National Learning Assessment of Grade 4 & 8 Reports, NEAEA, 2004, 2008, 2012, 2016	✓	✓	✓	✓
National Learning Assessment of Grade 10 & 12 Reports, NEAEA, 2010, 2014, 2018	✓	✓	✓	✓
Education Statistics Annual Abstract, MoE, 2016, 2017	✓	✓	✓	✓
Contextual Data Questionnaires (EGRA, EGMA, NLA), NEAEA, 2014, 2017	✓	✓	✓	✓
National Examinations Students Application Form, 1979–2018	✓	✓	✓	✓
Education Sector Development Program V (ESDP V), MoE, 2015	✓	✓	✓	✓
General Education Quality Improvement Program (GEQIP), MoE, 2008	✓	✓	✓	✓
General Education Quality Improvement Program for Equity (GEQIP-E), MoE, 2017	✓	✓	✓	✓
NEAEA Annual Reports, 2018	✓	✓	✓	✓
Syllabus for each subject by grade level, MoE, 2008	✓	✓	✓	✓
Table of Specifications, NEAEA (n.d.)	✓	✓	✓	✓
Item Development Manual, NEAEA, 2011	✓	✓	✓	✓

Document	Focus area 1 Context	Focus area 2 Coherence	Focus area 3.A Quality of large-scale assessments and examinations	Focus area 3.B Quality of classroom assessments
Item Development Evaluation Checklist, NEAEA, 2016,2018	✓	✓	✓	✓
National School Classification Framework, MoE, 2013	✓	✓	✓	✓
General Education Inspection Framework, MoE, 2014	✓	✓	✓	✓
Continuous Assessment Manual for First Cycle Primary (1–4), MoE and UNICEF (n.d.)	✓	✓	✓	✓
Competency-Based Assessment for Learning, UNICEF, 2017	✓	✓	✓	✓
Continuous Assessment and How to Use It, USAID, 2006	✓	✓	✓	✓
Assessment, Measurement and Evaluation in Primary Schools (Course Module), MoE, 2014	✓	✓	✓	✓
Continuous Professional Development for Primary and Secondary Teachers, Leaders and Supervisors, MoE, 2010, 2011, 2012	✓	✓	✓	✓
School Annual Reports, 2018	✓	✓	✓	✓
Education Sector Development Program IV (ESDP IV), MoE, 2010	✓	✓	✓	✓
Education Road Map (draft), 2018	✓	✓	✓	✓
Final READ Trust Fund Report 2008–2015, World Bank, 2015	✓	✓	✓	✓

References

- Australian Council for Educational Research Centre for Global Education Monitoring (ACER-GEM), and UNESCO Institute for Statistics (UIS). "Principles of Good Practice in Learning Assessment." ACER-GEM and UIS, 2017.
- Clarke, Marguerite. "What Matters Most for Student Assessment Systems: A Framework Paper." Systems Approach for Better Education Results (SABER) Student Assessment Working Paper No. 1, World Bank, Washington, DC, 2012.
- Council of Ministers Regulation No. 260/2012. "National Educational Assessment and Examinations Agency Establishment Council of Ministers Regulation No. 260/212." *Federal Negarit Gazeta* - No. 15. Addis Ababa: Berhanina Selam Printing Enterprise, January 26.
- Cresswell, John C. "System-Level Assessment and Educational Policy." Melbourne: ACER, 2017.
- Education Strategic Center. "Ethiopian Education Development Roadmap (2018–30): An Integrated Executive Summary." Addis Ababa: Ministry of Education, 2018.
- Global Partnership for Education (GPE). "Concept Note: Analysis of National Learning Assessment Systems (ANLAS)." Washington, DC: GPE, 2018.
- Ministry of Education (MoE). *Education Statistics Annual Abstract*. Addis Ababa: MoE, 2018.
- Ministry of Education (MoE). *Education Sector Development Program V (EDSP V), 2016-2020*. Addis Ababa: MoE, 2015.
- National Planning Commission Central Statistical Agency. *National Strategy for the Development of Statistics (2015/16–2019/20)*. Addis Ababa: CSA, 2016.
- Network for African Learning Assessments (NALA). "Learning Assessments Systems Evaluation Framework (draft)." n.d. http://www.adeanet.org/adeapmp/sites/default/files/activities/learning_assessments_systems_framework_v18.pdf.
- Transition Government of Ethiopia. *Education and Training Policy*. Addis Ababa: MoE, 1994.