

SUMMATIVE EVALUATION OF GPE'S ASSESSMENT FOR LEARNING (A4L) INITIATIVE

January 2021

by Lindsay Read &
Kate Anderson

UNBOUNDED
ASSOCIATES

Table of Contents

EXECUTIVE SUMMARY	4
INTRODUCTION	12
Background and purpose of evaluation	12
Methodology	12
Limits of the evaluation	15
CONTEXT AND PROGRAM DESIGN.....	16
GPE context	16
A4L context.....	16
PART 1 - CONTRIBUTION ANALYSIS	20
A4L's Theory of Change	20
A4L'S Activities and Outputs.....	22
A4L'S Contributions to Short-Term Outcomes	25
Effectiveness of ANLAS.....	25
Effectiveness of regional workshops.....	26
Effectiveness of efforts to integrate 21 st century skills.....	27
Contribution rating.....	27
A4L'S Contributions to Medium-Term Outcomes.....	29
Effectiveness of ANLAS.....	29
Relevance of ANLAS.....	30
Effectiveness of support to regional networks	31
Relevance of support to regional networks.....	32
Effectiveness and relevance of efforts to integrate 21 st century skills.....	33
Contribution rating.....	33
A4L'S Contributions to Long-Term Goals	35
Overall relevance of A4L	35
Relevance of A4L to GPE country-level support	36
Relevance of A4L to global efforts	36
Sustainability of A4L	37
Contribution rating.....	39
PART 2 - EFFICIENCY ANALYSIS	41
Management	41
Communications	42
Collaboration with foundations	44
LESSONS LEARNED AND RECOMMENDATIONS	46
What worked well	46
What hindered impact	46

Recommendations	47
APPENDICES	49
Appendix 1. Additional details about contribution analysis and matrix key	49
Appendix 2. List of stakeholders interviewed	51
Appendix 3. Generic interview protocol for GPE Secretariat.....	53
Appendix 4. Generic interview protocol for implementing partners	55
Appendix 5. Generic interview protocol for country partners	57
Appendix 6. Generic interview protocol for funders	59
Appendix 7. Generic interview protocol for external stakeholders.....	61
Appendix 8. List of reviewed project documents	62
Appendix 9. Web scan protocol	1
Appendix 10. Evaluation matrix	1
Appendix 11. Subsample of recommendations to strengthen the context and coherence of the assessment systems in ANLAS pilot countries	1
Appendix 12. Web scan results	1

Acknowledgements

The authors would like to thank Ramya Vivekanandan and Raphaëlle Martínez of the GPE Secretariat for their excellent leadership and guidance on this evaluation. We are also grateful to Daniela Ukuni of Unbounded Associates for her coordination and research assistance. This report benefitted from interviews with a wide range of stakeholders noted in Appendix 2. The final report represents the views of the authors and not necessarily those of the GPE Secretariat. Any errors or omissions are solely the responsibility of the authors.

Suggested citation: Read, L. & Anderson, K. (2021). *Summative evaluation of GPE’s Assessment for Learning (A4L) initiative*. Washington, DC: Unbounded Associates.

About Us

Unbounded Associates is a woman-owned small business that applies evidence in context-specific, culturally appropriate ways to improve outcomes for children and the families and systems that support them. We support donors, multilateral agencies, non-governmental organizations, governments, networks and researchers to find strategic ways to increase impact, communicate successes, and chart new paths in global education and human services. www.unboundedassociates.org

EXECUTIVE SUMMARY

Background

The Global Partnership for Education (GPE) is a multilateral partnership and funding platform that brings together developing country governments, donor nations, multilateral organizations, civil society, teachers' organizations, and the private sector to strengthen education systems worldwide.

In 2017, GPE launched the Assessment for Learning – Strengthening Learning Assessment Systems (A4L) initiative, which aims to strengthen learning assessment systems and promote the holistic measurement of learning at the national and global levels. The A4L initiative responded to demand from GPE partner countries for support was seen as an essential means to achieve the core goal of the GPE 2020 Strategic Plan to achieve improved and more equitable learning outcomes.

The initiative had a budget of US\$2.74 million, encompassing contributions from two foundations – \$2.24 million from Porticus and \$500 thousand from Dubai Cares – to support three major components of work:

1. **Piloting of a diagnostic tool for national learning assessment systems.** The Secretariat contracted the Australian Council for Educational Research (ACER) to design and launch a diagnostic toolkit – the Analysis of National Learning Assessment Systems (ANLAS) – to systematically gather and assess information about a country's learning assessment system to inform the education sector planning process. The tool was piloted in three GPE partner countries – Ethiopia, Vietnam, and Mauritania.
2. **Support to regional assessment networks.** To build a foundation for continued country-level work, GPE provided direct support to two regional assessment networks – the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) and the Teaching and Learning Educators' Network for Transformation (TALENT). The networks have hosted multiple workshops to build capacity for learning assessment, produced research and learning products related to the measurement of learning and facilitated the sharing of knowledge and expertise.
3. **Support for the measurement of 21st century skills.** Recognizing an increasing global and national focus on measuring skills beyond core academic subjects of literacy and numeracy, GPE undertook a landscape analysis of 21st century skills to elucidate the role GPE could take in supporting partner countries to embed 21st century skills into their education systems.

The GPE Secretariat commissioned this summative evaluation to review the results achieved under A4L, understand whether and how A4L support is contributing to the goal of strengthening learning assessment systems, and review contextual challenges to implementation. The findings of the evaluation will help shape future GPE programming.

This evaluation is guided by the OECD DAC evaluation principles of effectiveness, relevance, sustainability, and efficiency. The report will explore the extent to which A4L achieved its intended outputs, outcomes and goals, how relevant the project is at country and global levels, whether the

activities were implemented as planned, to what extent the initiative has catalyzed support in the area of learning assessment systems on the part of GPE, and whether targeted financing is an valuable modality to support work in specific thematic areas.

Methodology

The evaluation adopts a theory-based approach using Contribution Analysis, which examines a theory of change against results observed and builds a credible contribution story. The overall aim of the report is to understand why results did or did not occur and the role A4L played in the larger landscape of influencing factors. Contribution claims are derived from the A4L logic model.

A4L Logic Model

Activities	Outputs	Outcomes	Contribution to Systemic Change (Goal)
		Short term -> Medium term -> Longer term	
<ol style="list-style-type: none"> 1. Pilot diagnostic of national learning assessment systems in 3 countries. 2. Reflect on lessons learned/scale up of diagnostic. 3. Conduct a global landscape review of WCD measures in order to identify current stakeholder efforts, opportunities for collaboration and what role GPE could play and support the development of new tools to assess WCD. 4. Support 2 regional assessment networks (NEQMAP and TALENT) to build capacity at country level and share knowledge. 	<ol style="list-style-type: none"> 1. 3 technical diagnostic reports on learning assessment systems in 3 countries finalized. 2. Short report on lessons learned/scaling of diagnostic finalized. 3. Landscape report on WCD metrics, including 21st century and socio-emotional skills, completed. 4. Assessment tool developed and implemented in at least 2 countries. 5. 8 regional capacity building workshops conducted, covering topics such as systems strengthening and holistic measurement of learning. 6. At least 3 knowledge products (newsletters, webinars, online information etc.) produced / disseminated to promote knowledge sharing at regional level 	<p>Short-term:</p> <ol style="list-style-type: none"> A. Clear recommendations for how to strengthen learning assessment systems in 3 countries. B. Increased understanding of efforts to measure WCD, such as 21st century and socio-emotional skills. C. Increased measurement of social-emotional skills. <p>Medium-term:</p> <ol style="list-style-type: none"> D. Good practices on learning assessments systems embedded in countries, leading to stronger systems. E. Tool development for the measurement of WCD metrics, such as 21st century and social-emotional skills is advanced through collaborative efforts. F. Countries are exposed to models for integrating holistic measures of learning. 	<p>Long-term:</p> <ul style="list-style-type: none"> ● Strengthened learning assessment systems, where learning data is used by teachers, schools, and government officials to inform instruction, teacher training, and system wide policies to improve learning for all children. ● Socio-emotional skills and other 21st century skills are measured, monitored, and promoted at national and global level as part of a holistic measurement of learning.

Source: Porticus 2017 Annual Grant Report

The evaluation is primarily informed by a series of semi-structured interviews with 46 project stakeholders, including current and former members of the GPE Secretariat, foundation representatives, implementing partner staff, national contacts from ANLAS pilot countries, national focal points from within regional networks, and key stakeholders from relevant international agencies. Key informant

interviews were complemented by a review of program documents provided by the Secretariat, the analysis of quantitative data from the GPE Results Framework, and a web scan of articles and web pages mentioning A4L. The evaluation team also administered a short questionnaire to assess the operational efficiency of the program.

A full evaluation matrix that illustrates how the various lines of inquiry with their respective data collection methods and tools were used and triangulated to address the evaluation questions can be found in Appendix 10.

Key Findings and Conclusions

Activities and Outputs

Despite slight delays in A4L's activities, implementation remained relatively on track and all outputs from the first two components of the initiative – piloting of the ANLAS diagnostic tool and support to regional assessment networks – were achieved as expected. A report summarizing lessons from three ANLAS pilot countries was released in October 2019 and the two regional networks each held four capacity-building workshops, produced a number of knowledge products related to learning assessments and facilitated the sharing of knowledge and expertise.

In contrast, activities conducted under the third component of the A4L project – aimed at supporting broader measurements of learning (i.e., 21st century skills and whole child development) – deviated substantially from initial plans. The A4L project was initially meant to focus on the uptake and systematization of classroom assessment tools focused on 21st century skills, building on work undertaken by the Brookings Institution's OAA initiative. However, this uptake and systematization did not occur, mainly because the program design depended on a handoff of assets between GPE and Brookings rather than joint deliverables and expectations were not fully aligned in terms of what the work would look like in practice.

Another activity within the third component of work – a landscape review on 21st century skills – was also adjusted slightly from its original conception. First conceived as a review on the measurement of 21st century skills, it was reconceptualized as an *overall* global landscape review on 21st century skills with an updated goal to consider the role GPE could take in supporting its partner countries in embedding 21st century skills within their systems. This adjustment was understood as being better aligned with the value-add of the GPE partnership model and more responsive to countries' interest in the issue.

Outcomes

Component 1 - ANLAS

ANLAS supported the three pilot countries to undertake a comprehensive analysis of their national learning assessment systems and to develop recommendations that can be used to inform strategies to improve their assessment systems. The three pilot countries have used ANLAS as an input to the following activities as part of the sector planning process:

- Ethiopia: To develop the country’s next Education Sector Development Program (ESDP) VI plan (planned for 2020/21-2024/25); and to be used as a source document for curriculum revision, teacher training institutions, and teacher continuous professional development
- Vietnam: To develop the new Education Development Strategy for 2021-2030; and support the development of assessment regulations based on a new general education curriculum, to be piloted during the 2020-21 school year
- Mauritania: To inform the design of the Plan d’actions triennal (Triennial Action Plan) (PAT) 2019-2021, the Programme National de Développement du Secteur Éducatif (PDNES) III (2021-2030) and the country status report expected in 2020

This success is partly owed to the thoughtful selection of pilot countries in alignment with the timing of sector planning processes. In all cases, the timing of the ANLAS pilot meant discussions and findings from the diagnostic exercise were able to feed directly into the upcoming sector plans.

Yet, while the comprehensive nature of ANLAS is useful to countries, some stakeholders think that demand for the tool may be limited. First, some respondents felt that the ANLAS tool was “supply-driven,” rather than responsive to country interests and demands. Global, regional, and country-level stakeholders suggested that many governments are already aware of major shortcomings in their learning assessment systems and most, particularly countries with more nascent systems, may not require the level of detail that ANLAS offers and limited funds may be better spent on more pressing issues. Second, some stakeholders felt that the ANLAS toolkit duplicates, rather than complements, other efforts in the sector. For instance, the ANLAS toolkit adds to a list of existing frameworks, tools and rubrics available to support the analysis of learning assessment systems. Third, respondents suggested that the diagnostic tool is burdensome to national country-teams in that it is both time- and resource-intensive.

Component 2 - Support to Regional Workshops

National, regional, and global stakeholders all agreed that GPE support to the regional networks filled a significant gap, particularly as this is an area of work where not many other donors show interest. Regional workshops are resource intensive, and financial support from GPE allowed networks to greatly expand the scope and scale of work.

Many country-level stakeholders noted the value of regional workshops and knowledge exchange activities in facilitating peer learning as well as an awareness of best practices and intervention strategies from countries facing similar hurdles in strengthening learning assessment systems. Through the platform of the NEQMAP and TALENT regional networks, countries have had multiple opportunities to gain technical knowledge, share examples of best practices, and learn from the experiences of peers. Support to these regional workshops have created a structure that allows for systematic knowledge exchange between countries rather than ad hoc engagements, ensuring that participants gain a full picture of country issues and strategies. National government representatives spoke particularly highly of activities focused on strengthening the alignment and coordination across agencies working on curricula, teacher training, and assessment.

There is also some evidence that countries have taken this knowledge and embedded it at the country-level. For instance, Bhutan is now developing its first national education assessment framework using knowledge gained from the technical workshops about item development, sampling, test implementation, and dissemination of assessment results. Similarly, Cambodia has also made efforts to

strengthen its assessment framework and improve communications across departments within the Ministry of Education.

However, some felt that funding provided to the TALENT network was not commensurate with need or the ambition of the project, and support could have been made more relevant to the African context. Both NEQMAP and TALENT received the same amount of financial support despite higher operational costs in Africa for coordinating meetings and activities. **In addition, the fast pace of A4L deliverables made embedding work within countries more difficult.** This was particularly the case for the TALENT network, due to the added difficulty of coordinating a larger number of countries across a wider area compared to the Asia-Pacific region.

Component 3 - Support for the measurement of 21st century skills

As the third component of the A4L initiative was scaled back and no longer includes activities related to the uptake of classroom assessment tools focused on 21st century skills, **GPE's level of contribution to an increased measurement of socio-emotional skills is likely small, though not insignificant.** There is evidence that A4L slightly magnified the efforts of the OAA project through the convening of side meetings in the margins of regional network workshops as well as contributing knowledge and time to discussions on country selection and project design. But stakeholders noted that the OAA would have progressed in a similar vein without GPE's direct support.

However, **the landscape review of 21st century skills has been effective in its goal as a signal to country partners and the global community about GPE's interest in this space and as an input into GPE's strategic thinking.** Though discussions are ongoing, stakeholders suggest that 21st century skills may be featured more prominently in GPE's upcoming strategic plan, due in part to the landscape review exercise, which raised awareness and stimulated internal discussions. The new 2025 GPE goal will be "To accelerate access, learning outcomes and gender equality through equitable, inclusive and resilient education systems fit for the 21st century."

Impact

There is some indication that A4L has catalyzed support for learning assessment systems within the GPE Secretariat. A4L activities, particularly the landscape review of 21st century skills, have stimulated discussions about the strategic direction of GPE's priorities and have been included in negotiations regarding GPE's next strategic plan. There is also now a permanent position for a thematic lead to sustain momentum and broaden the portfolio of work. Several members of the Secretariat said that the support by foundations and the program specialist position afforded by A4L helped create the space for learning assessment systems to be featured as a thematic area in the Knowledge and Innovation Exchange (KIX).

Beyond GPE, stakeholders all agreed that A4L laid a good foundation for the strengthening of learning assessment systems worldwide, but further capacity building is needed to truly make a long-term impact at the country level. For instance, the long-term success of support to regional workshops could be limited without additional efforts from GPE to strengthen national capacity. Regional workshops (by design) tend to be one-off, short events with limited participation, and as such this limits their effectiveness as a vehicle to support long-term impact at the country-level.

For some stakeholders, the pathway for integrating A4L at the country-level has not been specified and it is unclear whether A4L complements GPE’s support at the country level. Representatives from the GPE Secretariat’s country teams shared confusion for what their role should be in relation to A4L’s products – whether as a neutral arbiter of discussions at the local level to assist countries in making their own decisions on priorities and sequencing, or to promote the use of a specific tool or methodology. **For some country partners, it is also unclear what mechanisms exist to request additional support for A4L work at the national level.** This could have been due to lack of internal communication and coordination within the Secretariat, or misalignment between the grant cycles and the period in which the ANLAS tool was developed.

Stakeholders also noted that there is a need for better synergies and coordination between the different initiatives and actors. Respondents felt that communication could be further formalized and established at all stages of project development. They noted that the area of 21st century skills is a space that could particularly benefit from expanded and strengthened partnerships, as it is a comparatively nascent and contested space. **The need for better coordination is particularly true at the country level where national stakeholders are sometimes uncertain how different efforts by multiple country partners (ex: GPE, World Bank, regional and international assessment networks) interact and complement each other.**

Efficiency

Overall, A4L stakeholders felt the day-to-day management of the initiative was well-organized. The A4L team at the Secretariat, especially the senior program specialist, was well-regarded overall as efficient and having good leadership. Yet, respondents were split in their perceptions of whether the time allocated for planning and design of A4L was adequate. While the GPE was flexible in terms of accommodating delays from partners and granting no-cost extensions to the regional networks, **implementing partners said they would have liked to see even more flexibility in their agreements with GPE,** for example to adjust the timing of certain activities or hire additional staff so they could better focus on implementation.

Collaborating with foundations was considered to be an important enabler of additional resources to the learning assessment thematic area. Several interviewees mentioned that the foundation funding on the theme of learning assessment was instrumental in allowing GPE to hire a program specialist and build a thematic area on learning assessment. It signaled the importance of a theme that is increasingly gaining traction both globally and at the country-level, and is fundamental to the current GPE funding model. **But notwithstanding these benefits, many stakeholders felt that the transaction costs were simply too high, and the GPE model does not allow for in-depth engagement with donors on a project.**

Lessons Learned and Recommendations

What worked well

Key drivers of A4L’s success included its leadership and management, particularly the project lead. Despite significant delays at the beginning of the project, implementation remained relatively on track and all outputs from components 1 and 2 were achieved as expected. Providing funding for a permanent senior education specialist and thematic lead for learning assessment has also allowed GPE to meaningfully participate in global and regional discussions related to learning assessment.

Another driver of success was the thoughtful selection of ANLAS pilot countries based on needs and alignment to timing with sector planning processes. All of the government stakeholders who participated in A4L activities reported significant needs related to strengthening learning assessment, and these needs aligned with the features of an assessment system put forth in ANLAS. Country stakeholders also welcomed the timely and relevant support to regional networks as a structured mechanism for peer learning and knowledge exchange related to joint challenges faced in reforming learning assessment systems.

The technical quality of partners such as ACER and Brookings, as well as the convening power of UNESCO Bangkok and UNESCO Dakar also strengthened the impact of A4L's activities. However, there was a tension between the technical strength of activities such as ANLAS and OAA and the accessibility and simplicity of tools requested by DCPs.

What hindered impact

Participants in the evaluation noted that while there were extensive negotiations with the funders prior to the start of the initiative, there was the limited time offered during the grant period for project development with implementing partners and funders, including necessary discussions to align expectations. Similarly, the equal allocation of funding across regional implementing agencies was not commensurate with the different needs of the organizations or the ambition of the project.

There was also a misalignment between the perceived promotion of one thematic area (i.e., strengthening learning assessment systems) over others, and how that related to the neutral role GPE is expected to play in facilitating the development of the education sector planning process.

While there was overall demand for A4L activities as noted by the GPE board members approving the initiative, the degree to which there was demand for the activities in the specific countries where they were implemented was questioned.

For one of the A4L funders, there were high expectations on the level of the funder's engagement in the initiative and the amount of communications and other requests that were expected of the A4L team. This was in contrast to another funder, whose limited requests to the GPE Secretariat helped ensure that staff time was spent on project implementation and not donor relations and reporting.

Recommendations

Building from these lessons, the evaluation team makes the following recommendations to the GPE Secretariat:

- Leverage the A4L activities and tools to provide direct support to the education sector planning process.
- Use KIX to fill a global gap for knowledge and innovation around 21st century skills.
- Consider adapting a "lite" or "screening" version of ANLAS as an entry point for countries with nascent learning assessment systems.

- Support deepened regional capacity building efforts on learning assessment systems with ongoing, just-in-time technical support.
- Amplify research and knowledge products generated by regional networks on GPE's website and other external and internal communications mediums.
- Set guidelines for how GPE engages with private foundations, including a minimum contribution level, minimal reporting requirements, expectations around the degree of engagement.

INTRODUCTION

Background and purpose of evaluation

The purpose of this summative evaluation is to review the results achieved under the Global Partnership for Education's (GPE) Assessment for Learning – Strengthening Learning Assessment Systems (A4L) initiative, due to conclude in July 2020.¹ This external evaluation was commissioned by the GPE Secretariat (Secretariat) and will focus on understanding whether and how A4L support is contributing to the goal of strengthening learning assessment systems and review contextual challenges to implementation.

The primary audience of this evaluation is the GPE Secretariat and developing country partners (DCPs). A secondary audience includes implementing partners (the Australian Council for Educational Research [ACER], the Brookings Institution, NEQMAP/UNESCO Bangkok, TALENT/UNESCO Dakar); Porticus and Dubai Cares as funders of the initiative; and the broader education community. The information will help shape future GPE programming related to learning assessment systems.

The core review period for this evaluation runs from the project's formal launch in 2017 to March 2020, when the evaluation process began. As some of the components of the initiative are ongoing, the evaluation team will only evaluate those activities that have been completed by March 2020 but will allow for the mention of ongoing and recently completed activities where relevant.

Methodology

This summative evaluation is guided by the OECD DAC evaluation principles of effectiveness, relevance, sustainability, and efficiency.² The report will explore the extent to which A4L achieved its intended outputs, outcomes and goals, how relevant the project is at country and global levels, whether the activities were implemented as planned, to what extent the initiative has catalyzed support in the area of learning assessment systems on the part of GPE, and whether targeted financing is an valuable modality to support work in specific thematic areas. The evaluation is guided by the following questions:

Table 1. List of evaluation questions

Domains	Evaluation Questions
Effectiveness	1. To what extent has A4L achieved its intended outputs, including completing the ANLAS toolkit, producing a landscape report on 21 st century skills, supporting regional networks, convening capacity building workshops, and producing knowledge products?

¹ No-cost extensions have been granted to TALENT and NEQMAP until the end of 2020 due to delays caused by the COVID-19 crisis.

² According to the OECD DAC evaluation criteria, effectiveness refers to the extent to which the intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance; relevance refers to the extent to which the objectives of an intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies; efficiency is the measure of how economically resources (funds, expertise, time, etc) are converted to results; and sustainability refers to the continuation of benefits from an intervention after assistance has been completed.

Source: https://www.oecd.org/dac/evaluation/seco_guidelines.pdf

<p>Effectiveness (con't)</p>	<ol style="list-style-type: none"> 2. To what extent has A4L achieved its intended short-term and medium-term outcomes, including embedding good practices at the country level and advancing the measurement of whole child development (WCD)³ metrics? 3. To what extent has A4L contributed toward its long-term goals of strengthening learning assessment systems and promoting the measurement of socio-emotional and other 21st century skills at the national and global levels? 4. What were the most critical variances between planned and actual project activities? What contributed to these variances? 5. What aspects of project design and organization most contributed to its success and/or shortcomings?
<p>Relevance & Sustainability</p>	<ol style="list-style-type: none"> 6. To what extent has A4L activities addressed the needs and priorities of country, regional and global actors in terms of strengthening learning assessment systems? 7. How do A4L activities complement and/or add value to GPE's support to learning assessment systems at the country level? 8. How do A4L activities complement and/or add value to ongoing global efforts to strengthen learning assessment systems? 9. To what extent has A4L catalyzed support and investment in the areas of learning assessment systems (at both global and country level)? 10. To what extent has A4L allowed for new or strengthened collaboration among actors at global, regional, and country levels in regard to strengthening learning assessment systems?
<p>Efficiency</p>	<ol style="list-style-type: none"> 11. To what extent has the management and coordination of A4L among the GPE Secretariat, implementing agencies, and funders been well-organized, consistent and efficient in terms of staff-time and resources? 12. To what extent has the Secretariat been efficient in terms of communication and dissemination around A4L activities and achievements? 13. To what extent has the collaboration with foundations added value for GPE in supporting work in specific thematic areas?

To judge the initiative's effectiveness, relevance, and sustainability, the evaluation adopts a theory-based approach using Contribution Analysis (CA). This approach, which is a blend of various theory-based methods such as process tracing and outcome harvesting, examines a theory of change against results observed and builds a credible contribution story by demonstrating plausible associations between inputs and subsequent changes.⁴

The A4L theory of change is shown in the initiative's logic model in table 2 below.

³ Whole Child Development is the preferred terminology of the primary A4L funder, Porticus, in reference to 21st century skills. Various stakeholders also refer to 'transversal skills' and 'socio-emotional skills.' For the purpose of this evaluation, 21st century skills, transversal skills, socio-emotional skills and WCD are all used interchangeably.

⁴ Befani, B. and O'Donnell, M. (2016): Choosing Appropriate Evaluation Methods: A Tool for Assessment and Selection. Bond. Available at: https://www.bond.org.uk/sites/default/files/caem_narrative_final_14oct16.pdf

Table 2. A4L Logic Model

Activities	Outputs	Outcomes	Contribution to Systemic Change (Goal)
		Short term -> Medium term -> Longer term	
<p>1. Pilot diagnostic of national learning assessment systems in 3 countries.</p> <p>2. Reflect on lessons learned/scale up of diagnostic.</p> <p>3. Conduct a global landscape review of Whole Child Development (WCD)⁵ measures in order to identify current stakeholder efforts, opportunities for collaboration and what role GPE could play and support the development of new tools to assess WCD.</p> <p>4. Support 2 regional assessment networks (NEQMAP and TALENT) to build capacity at country level and share knowledge.</p>	<p>1. 3 technical diagnostic reports on learning assessment systems in 3 countries finalized.</p> <p>2. Short report on lessons learned/scaling of diagnostic finalized.</p> <p>3. Landscape report on WCD metrics, including 21st century and socio-emotional skills, completed.</p> <p>4. Assessment tool developed and implemented in at least 2 countries.</p> <p>5. 8 regional capacity building workshops conducted, covering topics such as systems strengthening and holistic measurement of learning.</p> <p>6. At least 3 knowledge products (newsletters, webinars, online information etc.) produced / disseminated to promote knowledge sharing at regional level</p>	<p>Short-term:</p> <p>A. Clear recommendations for how to strengthen learning assessment systems in 3 countries.</p> <p>B. Increased understanding of efforts to measure WCD, such as 21st century and socio-emotional skills.</p> <p>C. Increased measurement of social-emotional skills.</p> <p>Medium-term:</p> <p>D. Good practices on learning assessments systems embedded in countries, leading to stronger systems.</p> <p>E. Tool development for the measurement of WCD metrics, such as 21st century and social-emotional skills is advanced through collaborative efforts.</p> <p>F. Countries are exposed to models for integrating holistic measures of learning.</p>	<p>Long-term:</p> <ul style="list-style-type: none"> Strengthened learning assessment systems, where learning data is used by teachers, schools, and government officials to inform instruction, teacher training, and system wide policies to improve learning for all children. Socio-emotional skills and other 21st century skills are measured, monitored, and promoted at national and global level as part of a holistic measurement of learning.

Source: Porticus 2017 Annual Grant Report

The overall aim is to understand why results did or did not occur and the role the project played in the larger landscape of influencing factors. The evaluation team has used a color-coded rubric to categorize the results of the contribution analysis along three lines of inquiry: 1) Level of contribution; 2) Significance of the outcome; and 3) Strength of the evidence underlying the claims. The level of contribution refers to the extent the outcome would or would not have happened in the absence of the intervention; significance is understood as the importance of the effect (e.g., how many people benefited from the change or the size or scale of the benefit); and the strength of the evidence refers to how decisive and descriptive the available evidence is in relation to the outcome of interest. Each component has been rated on a scale from low (red), medium (yellow), to high (green). The color coding is intended as a qualitative orientation tool to readers, rather than as a quantifiable measure. Additional details about CA can be found in Appendix 1, including a rubric key.

⁵ Whole Child Development is the terminology used by one of the A4L funders, Porticus, to reference 21st century skills.

The first stage of this evaluation involved drafting an inception report and confirming the research questions and methodology with the GPE Secretariat. The evaluation is primarily informed by a series of semi-structured interviews with current and former members of the GPE Secretariat, foundation representatives, implementing partner staff, national contacts from ANLAS pilot countries, national focal points from within regional networks, and key stakeholders from relevant international agencies. An initial list of contacts was provided by the Secretariat and additional names were gathered using a snowball method. A total of 46 respondents were interviewed by Zoom or telephone between April 27, 2020 and May 22, 2020. Interviews were conducted in English and French with simultaneous translation. Appendix 2 provides a full list of contacts and their associated organization. Appendices 3 to 7 provide the generic interview protocols that were used to guide the interviews for each stakeholder group.

Key informant interviews were complemented by a review of program documents provided by the Secretariat, including strategy documents, grant proposals, annual reports, previous evaluation results, and published knowledge products. Appendix 8 provides a full list of reviewed documents. The evaluation team also gathered and analyzed quantitative data from the GPE Results Framework and conducted a web scan of articles and web pages mentioning A4L. Details of the web scan protocol can be found in Appendix 9.

In addition to an assessment of the effectiveness, relevance and sustainability of A4L, a key line of inquiry for the evaluation is to assess the operational efficiency of the program. To gain insight into this strand of questioning, the evaluation team administered a short questionnaire to a subsample of 21 stakeholders, from key informant interviews involved in implementation, administration, and communication-related tasks to discern the amount of time spent on operational activities and their view on the initiative's value. We received 17 responses to the questionnaire, including 3 representatives from the 2 foundations that supported A4L, 5 members of the GPE Secretariat, and 9 staff members across the 3 implementing agencies. A copy of the questionnaire can be found in Appendix 10.

Qualitative and quantitative data were organized in relation to the evaluation questions outlined above, then categorized and analysed based on whether the information confirmed or contradicted the program's contribution claims (see figure 2). A full evaluation matrix that illustrates how the various lines of inquiry with their respective data collection methods and tools were used and triangulated to address the evaluation questions can be found in Appendix 10.

Limits of the evaluation

This evaluation has several limitations. First, it collected data on informants' viewpoints, opinions, and perspectives, and it is possible that some informants offered views which were biased in one way or another. This challenge, however, is not unique to this evaluation. We mitigated these challenges by interviewing multiple individuals within organizations and with similar roles, to ensure as much as possible we were gathering all of the perspectives.

Second, COVID-19 made scheduling interviews difficult, particularly for country partners, so no interviews could be conducted in person, and for some interviewees, they were also unable to participate via Zoom calls. For these participants, we mitigated the effects by requesting answers to the evaluation questions via email, but in those cases the evaluation team could not guide questioning or follow-up on questions to dig deeper.

Third, we conducted two interviews in French with simultaneous translation, and a third interview scheduled in French could not be conducted because the interviewees were having connection issues. For those interviewees, we mitigated the situation by collecting responses via email. There were also two country partners who did not respond to requests for interviews or email questionnaires.

Fourth, some A4L activities are ongoing and are not captured by this evaluation. For instance, a webinar on the 21st century landscape review report took place in May 2020, while the ANLAS toolkit has been translated into additional languages and the regional networks are completing additional strategic activities.

CONTEXT AND PROGRAM DESIGN

GPE context

The Global Partnership for Education is a multilateral partnership and funding platform established in 2002 that brings together developing country governments, donor nations, multilateral organizations, civil society, teachers' organizations, and the private sector to strengthen education systems worldwide. As a partnership and a fund, GPE mobilizes investments, both external and domestic, in nearly 70 partner countries to develop and implement education sector plans to support a quality basic education for every child, with a specific focus on the poorest and most vulnerable children. The GPE Secretariat is guided by a constituency-based Board of Directors, which sets policies and strategies and approves performance-based funding grants, including education sector plan development grants (ESPDGs), program development grants (PDGs), and education sector program implementation grants (ESPIGs).

GPE's present operations are guided by its current strategic plan (2016-2020) known as GPE 2020, which focuses heavily on education quality and equity in response to mounting evidence of a learning crisis facing developing countries. In this context, the GPE Board approved the Knowledge and Good Practice Exchange (KGPE) in June 2016, which aims to support national capacities to plan and design improvements in national education systems. KGPE activities were designed to "leverage improvements in sector planning, mutual accountability through inclusive and evidence based policy dialogue, and by enhancing the capacity of developing countries to more effectively utilize GPE grants and finance."⁶ The KGPE strategy employed a unique financing modality that piloted the use of targeted financing from foundations.

A4L context

Assessment for Learning (A4L) was one of two thematic initiatives launched under the KGPE strategy.⁷ Learning assessment is seen as an essential means to achieve the goals of the GPE 2020 Strategic Plan, which has a core goal of 'improved and more equitable learning outcomes.' Learning assessments provide the means for countries to monitor learning, can inform evidence-based policymaking and reform processes, and improve teaching practices.

⁶ GPE (2016). Meeting of the Board of Directors June 14-15, 2016. Oslo, Norway (BOD/2016/06 DOC16). Available: <https://www.globalpartnership.org/sites/default/files/2016-06-bod-doc-16-knowledge-and-good-practice-exchange.pdf>

⁷ The second initiative was the Better Early Learning and Development at Scale (BELDS) initiative, which aims to improve global knowledge and resources available to support countries with effective planning and implementation of Early Childhood Care and Education programs.

A4L aims to strengthen learning assessment systems to be capable of yielding policy-relevant evidence for sector planning and administration and to promote the holistic measurement of learning at the national and global levels. In 2015, when the Sustainable Development Goals were being finalized, the GPE Secretariat and Board members had participated in several multi-stakeholder dialogues on learning measurement, including the Learning Metrics Task Force. It became evident that countries needed targeted support to generate the types of learning data needed to report on the SDG indicators, as well as to inform their own national educational goals. The A4L initiative responded to demand from GPE partner countries for support to embed learning assessment systems with the capacity to inform educational policies and increase the availability of learning data at country, regional and global levels.

The initiative is intended to complement the country-level work of GPE. The GPE funding model requires countries to have a system or mechanism in place to monitor student learning outcomes, or a time-bound plan to develop such systems or mechanisms. In addition, the funding model requires countries that do not meet this requirement to use GPE Education Sector Plan Implementation Grant (ESPIG) funding for this purpose. Of the 34 implementation grants active in 2018, 88% invested in activities related to learning assessments, representing some \$39 million of grant allocations (out of a total of \$582 million allocated to learning activities).⁸ The quality of countries' learning assessment systems is also monitored through an indicator in GPE's Results Framework, on which data is collected every two years. Significantly, 48% of GPE partner countries were classified as having an established learning assessment system that met quality standards in 2018, up from 40% at baseline in 2015.⁹

A4L also dovetails with GPE's Knowledge and Innovation Exchange (KIX), which seeks to strengthen national education systems by facilitating knowledge sharing, innovation, and capacity strengthening among country partners. Strengthening Learning Assessment Systems is one of the six thematic areas for which organizations can apply for KIX funding, and the concept papers outlining potential activities within this thematic area builds on the efforts of A4L. It is expected that A4L activities could be amplified or strengthened in some manner under KIX. Additional financing may be provided by foundations, the private sector, and traditional donors to further drive the work of KIX, as well as technical collaborations with new partners.

A4L Design

The A4L initiative was launched by the Secretariat in July 2017 and is due to conclude in July 2020. The initiative had a budget of US\$2.74 million, encompassing contributions from two foundations – \$2.24 million from Porticus and \$500 thousand from Dubai Cares – to support three major components of work:

1. **Piloting of a diagnostic tool for national learning assessment systems.** The Secretariat contracted the Australian Council for Educational Research (ACER) to design and launch a diagnostic toolkit – the Analysis of National Learning Assessment Systems (ANLAS) – to systematically gather and assess information about a country's learning assessment system to inform the education sector planning process. The goal of the toolkit is to foster the use of evidence-based strategies for building robust learning assessment systems, and incorporate these strategies in the design of ESPs. The tool was piloted in three GPE partner countries – Ethiopia, Vietnam, and Mauritania – which were selected from a shortlist of countries that have

⁸ Global Partnership for Education (2019). Results Report 2019. GPE: Washington, DC.

⁹ Ibid.

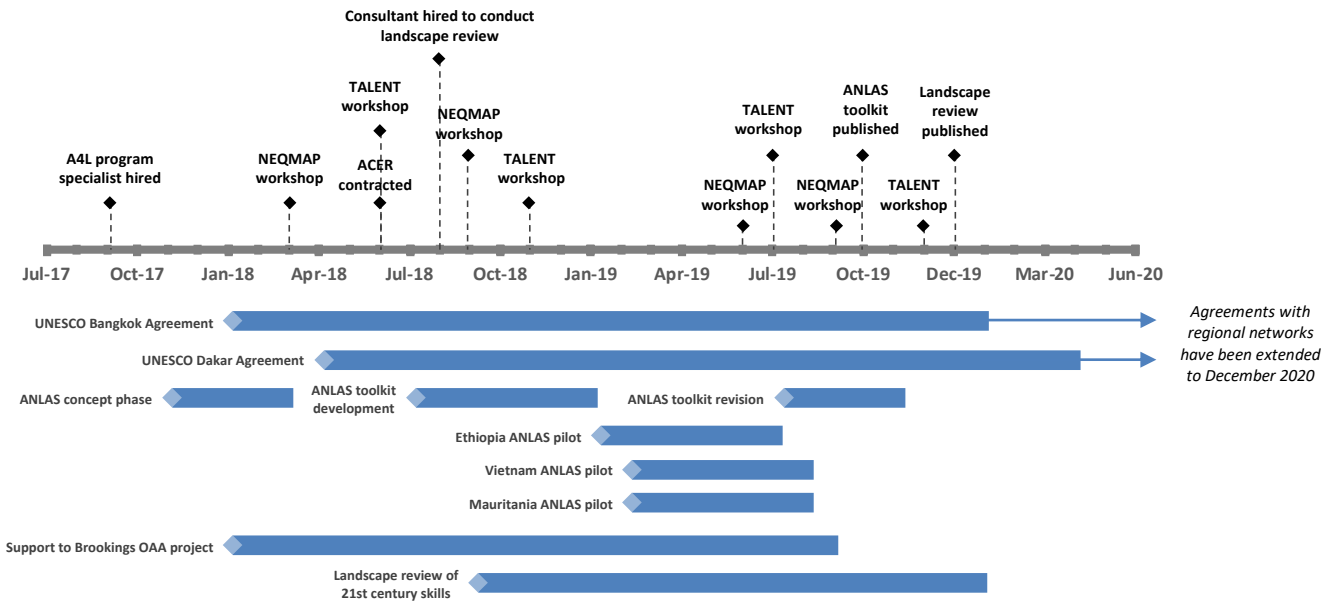
demonstrated interest and demand for strengthening learning assessment systems, particularly in relation to the assessment of 21st century skills, and that have a new Education Sector Plan starting in the 2020-2022 period. The countries are working to incorporate findings of their ANLAS exercise into their next Education Sector Plans. A final version of the toolkit was released on the GPE website in October 2019.

2. **Support to regional assessment networks.** To build a foundation for continued country-level work, GPE provided direct support to two regional assessment networks – the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP) and the Teaching and Learning Educators’ Network for Transformation (TALENT).¹⁰ These networks, coordinated by the UNESCO field offices in Bangkok and Dakar respectively, bring together learning assessment stakeholders of the countries of each region to share knowledge around strengthening learning assessment systems and promote holistic measurements of learning. The networks have hosted multiple workshops to build capacity for learning assessment, produced research and learning products related to the measurement of learning and facilitated the sharing of knowledge and expertise.
3. **Support for the measurement of 21st century skills.** Recognizing an increasing global and national focus on measuring skills beyond core academic subjects of literacy and numeracy, GPE undertook a landscape analysis of 21st century skills to elucidate the role GPE could take in supporting partner countries to embed 21st century skills into their education systems. The landscape review provided a deep dive on education sector plans and assessment systems in 15 GPE partner countries, as well as a snapshot of related initiatives by select global and regional actors. The report was published on the GPE website in January 2020. Another activity in this same vein, the implementation of new assessment tools that promote socio-emotional skills building from the Brookings Institution’s Optimizing Assessment for All (OAA) initiative, was planned but subsequently dropped.

The A4L team was comprised of 100% of a Programme Specialist (later recruited at WB/GPE’s ‘G’ level or Senior Education Specialist), 15% of the time of a ‘Learning Specialist’ and 10% of the time of a ‘Senior Advisor’. The team also included a short-term consultant during the majority of FY19. The senior program specialist charged with coordinating the initiative was recruited and commenced work at the Secretariat on September 1, 2017. A timeline of activities under review for this evaluation is illustrated in Figure 1.

¹⁰ Originally, it was planned that NALA, which is a task team of the Association for the Development of Education in Africa (ADEA), would be the recipient of GPE support to regional networks in Africa. But in the months prior to the formal commencement of the A4L activities, it was determined that TALENT, which is coordinated by UNESCO Dakar, would be the recipient. This decision was made because TALENT focuses on regional networking for capacity development, research and knowledge sharing and thus is more aligned with the goals of A4L support for regional assessment networks. It was determined that working with TALENT also allows for channelling the A4L support through an existing grant agent (UNESCO), while ADEA is not an existing grant agent. Source: GPE (2018). Meeting of the Finance and Risk Committee October 10-11, 2018. Washington, DC.

Figure 1. Timeline of A4L activities



PART 1 - CONTRIBUTION ANALYSIS

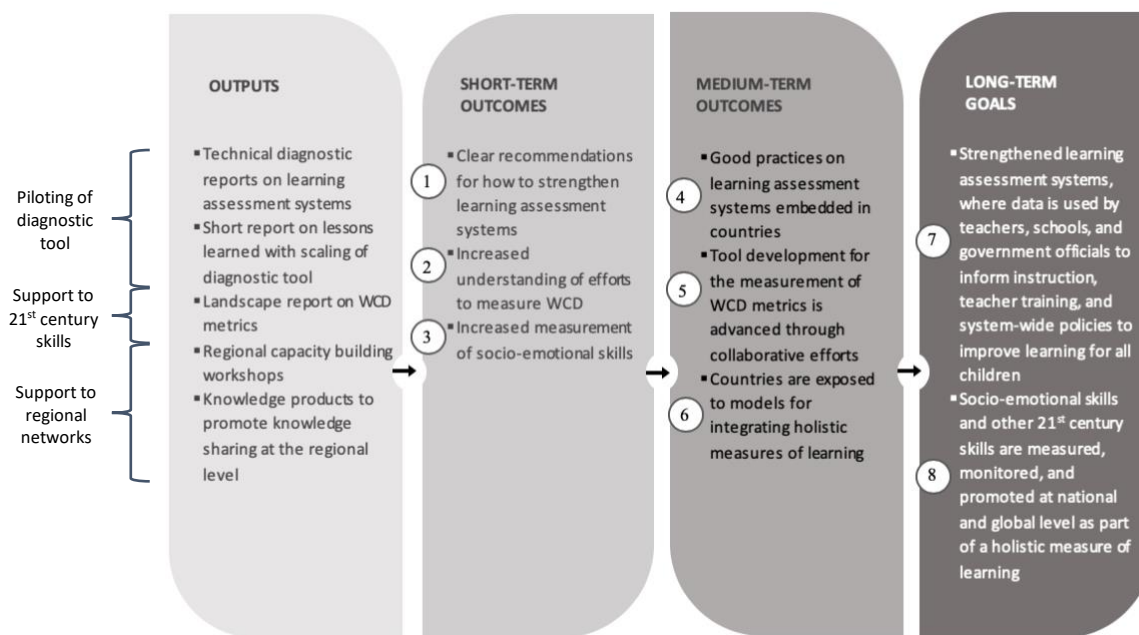
The findings of this evaluation are organized in two sections: first, a deep exploration of the extent to which A4L contributed to the achievement of its goals and objectives and second, an assessment of the operational efficiency of the initiative and its perceived value. In a subsequent section, the evaluators explain the aspects of the project that contributed to its success and/or shortcomings.

The first section addresses the initiative’s effectiveness, relevance and sustainability as they relate to each contribution claim found in the initiative’s theory of change.

A4L’s Theory of Change

To systematically collect and assess evidence on the extent and nature of GPE’s contributions to the initiative’s goals and objectives, the evaluation team has elaborated on A4L’s detailed logic model in table 2 to identify the primary pathways of change for the project, shown in Figure 2. Through this exercise, we have identified eight ‘contribution claims’, moving from the initiative’s outputs to its short-term outcomes, medium-term outcomes, and long-term goals. The links underlying each claim were tested during key stakeholder interviews and the document review process.

Figure 2. Contribution claims assessed in the evaluation

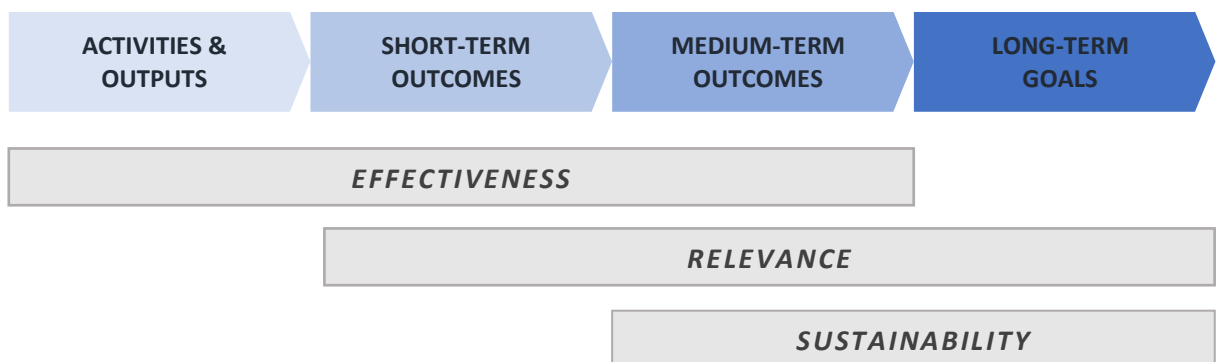


- Contribution claim #1:** A4L tools, knowledge products, and regional technical capacity workshops contributed to providing clear recommendations for how to strengthen learning assessment systems
- Contribution claim #2:** A4L tools, knowledge products, and regional technical capacity workshops contributed to an increased understanding of efforts to measure whole child development, such as 21st century and socio-emotional skills

3. **Contribution claim #3:** A4L tools, knowledge products, and regional technical capacity workshops contributed to an increased understanding of socio-emotional skills
4. **Contribution claim #4:** A4L contributed to embedding good practices on learning assessment systems in countries
5. **Contribution claim #5:** A4L contributed to advancing tool development for the measurement of WCD metrics, such as 21st century and socio-emotional skills
6. **Contribution claim #6:** A4L contributed to exposing countries to models for integrating holistic measures for learning
7. **Contribution claim #7:** A4L contributed toward its long-term goal of strengthening learning assessment systems, where data is used by teachers, schools, and government officials to inform instruction, teacher training, and system-wide policies to improve learning for all children
8. **Contribution claim #8:** A4L contributed toward its long-term goals of the measurement, monitoring, and promotion of socio-emotional and other 21st century skills at the national and global levels as part of a holistic measure of learning

In line with the overall objective of the evaluation to assess the effectiveness, relevance, and sustainability of the project, the evaluators assume these elements as building upon each other when moving along the results chain from outputs to long-term goals. Throughout the report, effectiveness is understood as a critical precondition to achieve short-term outcomes (i.e., the extent to which the objectives were achieved); effectiveness and relevance are seen as critical preconditions for achieving medium-term outcomes (i.e., that activities and short-term outcomes are relevant to country and global priorities); and relevance and sustainability are critical preconditions for achieving long-term goals (i.e., the extent to which benefits will continue following the project completion) (see figure 3). These assumptions are guided by the recognition that project implementers have less control over the project’s success as you move from outputs to long-term impact, which requires country and global actors to take on and sustain results. The following sections thus build upon each other in exploring the effectiveness, relevance, and sustainability of each of A4L’s components.

Figure 3. How questions of effectiveness, relevance, and sustainability align to the theory of change



A4L'S Activities and Outputs

To assess the extent to which GPE's activities contributed to short-, medium-, and long-term goals, it is first necessary to understand the extent to which A4L achieved its intended outputs (Evaluation Question 1 [EQ1])¹¹ and whether there were critical variances between planned and actual project activities (EQ4).

According to project documents, there were a number of delays in A4L's activities. First, the program specialist in charge of coordinating the project joined the GPE Secretariat in September 2017, two months after the initiative's formal start.¹² Understandably, this led to some delays in the implementation of project components given the time it takes to get acquainted with the project and commence work.

In relation to the first component of A4L, the ANLAS toolkit, the World Bank procurement process also took longer than anticipated, resulting in a delay in contracting ACER to commence work. In addition, the selection of pilot countries took more time than expected due to lengthy discussions among Secretariat Country Leads, Coordinating Agencies, and GPE Focal Points within Ministries of Education from short-listed candidate countries.¹³ As one interviewee remarked, governance structures are "thick," meaning it can take a long time to communicate and build consensus across and within both education ministries and within the GPE Partnership. Moreover, ACER required additional time to revise the pre-pilot version of the toolkit given the "substantive volume and nature of the feedback" received on the first draft.¹⁴ The piloting of the ANLAS toolkit also took longer than planned, given delays at the country level due to exogenous factors such as holidays, strikes, and elections, as well as factors such as the length of time needed for data collection, language translation, and in-depth processes for broad stakeholder engagement.

The second component of A4L – support to regional assessment networks – was similarly delayed due to the length of time it took to finalize grant agreements with UNESCO Bangkok and UNESCO Dakar. The agreement with UNESCO Dakar was particularly delayed and the network commenced work in April 2018.

However, despite these delays, implementation remained relatively on track and all outputs from both of these components were achieved as expected. A report summarizing lessons from three ANLAS pilot countries was released in October 2019 and the two regional networks each held four capacity-building workshops, produced a number of knowledge products related to learning assessments and facilitated the sharing of knowledge and expertise, listed in Table 3.

¹¹ See Table 1 for full list of evaluation questions

¹² Porticus 2017 Annual Report.

¹³ Porticus 2018 Annual Report.

¹⁴ Ibid.

Table 3. Component 2 of A4L initiative - List of outputs emerging from support to regional networks from January 2018 to May 2020

	Regional Capacity Workshops	Knowledge Products	Knowledge-sharing Activities
NEQMAP	<ol style="list-style-type: none"> 1. Bangkok, Thailand: “Data Quality and Accuracy of Large-Scale Learning Assessments (March, 2018) 2. Penang, Malaysia: “Conceptualization, Measurement and Use of Contextual Data” (September, 2018) 3. Bandung, Indonesia: “Assessment for Learning” (June, 2019) 4. Manila, Philippines: “Promoting Transversal Competencies Across Curriculum, Pedagogy and Assessment” (September, 2019) 	<ul style="list-style-type: none"> • Thematic reviews on the topics of assessment for students with disabilities, gender and assessment, and language barriers and assessment¹⁵ • Study on assessment of transversal competencies, coordinated by the Brookings Institution¹⁶ • 2nd edition of topical case studies on the national assessment system in Nepal¹⁷ • 3rd edition of topical case studies on SEA-PLM¹⁸ • Mapping study of learning assessments in Asia-Pacific region (forthcoming) 	<ul style="list-style-type: none"> • 5th, 6th and 7th annual NEQMAP meeting and Steering Group meeting. • 5th, 7th, and 8th editions of biannual newsletter • Three regional meetings for technical team experts of OAA • Expert meeting on inclusion and equity in learning assessments • Three webinars on: <ul style="list-style-type: none"> ◦ Computer-based testing¹⁹ ◦ The culture of testing²⁰ ◦ Learning disabilities²¹ • Special seminar on lessons for education development in East Asia in collaboration with the World Bank²²
TALENT	<ol style="list-style-type: none"> 1. Dar es Salaam, Tanzania: “Alignment between Curriculum, Teacher Training and Learning Assessments (July, 2018) 2. Dakar, Senegal: “Effective Reporting, Dissemination and Use of Large-Scale Learning Assessment Systems” (November, 2018) 3. Addis Ababa, Ethiopia: “Classroom-Based Formative Assessment” (July, 2019) 4. Yaounde, Cameroon: “Development and Strategic Use of National Large-Scale Learning Assessments” (December, 2019) 	<ul style="list-style-type: none"> • Draft of mapping study on learning assessment systems • Draft of policy brief and cross-country review on the use and impact of learning assessment systems to improve learning outcomes • Regional study on the assessment of 21st century skills with the Brookings Institution • Guidelines on formative assessment in the context of distance learning • Brief on measures adopted to assess learning during times of school closures (forthcoming) 	<ul style="list-style-type: none"> • Discussion forum on teacher training and professional development • Online training course on alignment of curriculum, teacher training and learning assessment • Regional meetings for technical team experts of OAA • Webinar on the use of learning assessment data, in collaboration with IIEP²³ • Webinar on conducting assessments during school closures²⁴

¹⁵ <https://bangkok.unesco.org/content/neqmap-thematic-review>

¹⁶ <https://bangkok.unesco.org/content/assessment-transversal-competencies-current-tools-asian-region>

¹⁷ <https://bangkok.unesco.org/content/integrating-findings-national-assessment-student-achievement-policy-process-experience>

¹⁸ <https://bangkok.unesco.org/content/southeast-asia-primary-learning-metrics-program-thinking-globally-regional-context>

¹⁹ <http://bangkok.unesco.org/content/computer-based-assessments-open-new-pathways-self-directed-learning-neqmap-webinar>.

²⁰ <https://unesdoc.unesco.org/ark:/48223/pf0000261955>

²¹ <https://bangkok.unesco.org/content/how-can-we-help-learners-hidden-disability-searching-answers-upcoming-seminar>

²² <https://bit.ly/2CvNzaQ>

²³ <http://www.iiep.unesco.org/en/webinar-use-learning-assessment-data-what-have-we-learnt-so-far-5006>

²⁴ <http://www.education2030-africa.org/index.php/en/lien-talent-blog-en/1317-webinar-talent-iicba-on-assessment-during-times-of-schools-closure>

In contrast, activities conducted under the third component of the A4L project – aimed at supporting broader measurements of learning (i.e., 21st century skills and whole child development) – deviated substantially from initial plans. The A4L project was initially meant to focus on the uptake and systematization (referred to as ‘scaling’ in some documents) of classroom assessment tools focused on 21st century skills, building on work undertaken by the Brookings Institution’s OAA initiative.

The partnership between GPE and Brookings followed three phases of work within the Porticus Measuring What Matters Learning Partnership. In its first phase of work, GPE provided funds through UNESCO Bangkok/NEQMAP and UNESCO Dakar/TALENT for Brookings to coordinate mini studies on the measurement of 21st century skills in eight countries in the Asia-Pacific region and nine in the Africa region. Three countries in Asia (Cambodia, Mongolia and Nepal) and three countries in Africa (Democratic Republic of the Congo, The Gambia, and Zambia) were then selected by Brookings for the second phase of work under its Optimizing Assessment for All (OAA) initiative to “design, develop, administer, and pilot classroom-based assessment of 21st century skills.”²⁵ In these phases, GPE provided additional support via UNESCO Bangkok/NEQMAP and UNESCO Dakar/TALENT for supplemental convenings of the countries participating in the OAA work in the sidelines of regional workshops.

For the third phase, the initial plan was for the A4L initiative to “support the uptake of 21st century skills measurement at a broader, system level,” in one country from each of the regions, commencing July 2019.²⁶ However, this uptake and systematization did not occur. Consequently, the last tranche of the Porticus grant was reduced by \$300,000.²⁷

According to key stakeholder interviews, the reason for the activity being removed from the scope of work stems from a misalignment in the project’s conception. The primary funder of A4L, Porticus, required that GPE and the Brookings Institution work together as they had worked with the two organizations to submit grant proposals to Porticus with complementary objectives and scope. However, respondents described the linkages between the two initiatives as being “not well-conceived” and “artificial.” Based on discussions with stakeholders from Brookings, GPE, and Porticus, there appear to be two broad limitations of the collaboration, which made it difficult to agree on timely and effective adjustments:

- The program design depended on a handoff of assets between the two organizations rather than joint deliverables;
- Expectations were not fully aligned in terms of what the work would look like in practice, as what appeared in the MOU was interpreted differently by the different parties;

This suggests that the primary issue was with the activity’s design rather than faults in its implementation.

Another activity within the third component of work – a landscape review on 21st century skills – was also adjusted slightly from its original conception. First conceived as a review on the measurement of 21st century skills, it was reconceptualized as an *overall* global landscape review on 21st century skills with an updated goal to consider the role GPE could take in supporting its partner countries in

²⁵ Care, E., A. Vista & H. Kim. *Optimizing Assessment for All: Measuring 21st century skills*. Washington, DC: Brookings Institution. <https://www.brookings.edu/product/optimizing-assessment-for-all/>

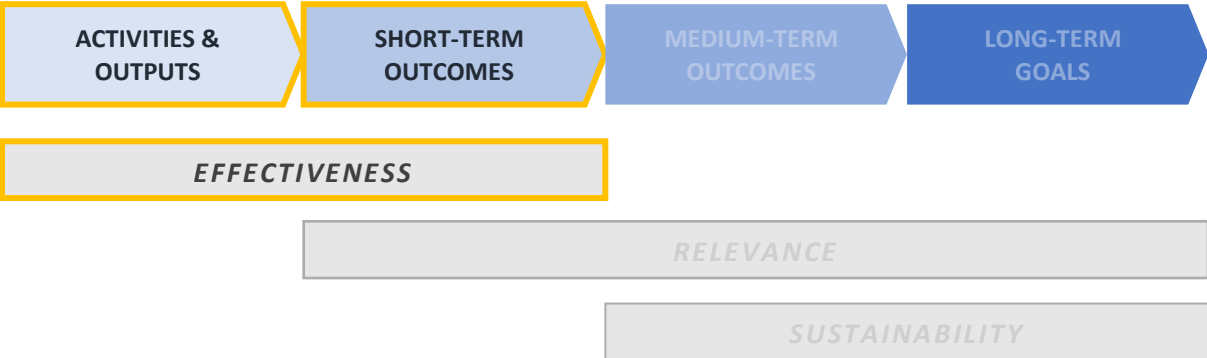
²⁶ GPE (2018). Meeting of the Finance and Risk Committee October 10-11, 2018. Washington, DC.

²⁷ Ibid.

embedding 21st century skills within their systems. This adjustment was understood as being better aligned with the value-add of the GPE partnership model and more responsive to countries’ interest in the issue. The review was based on a scan of key definitions, frameworks and initiatives, a stocktake of Education Sector Plans and GPE grant documents in a sample of 15 partner countries, a deep dive on assessment in the same countries based on the mini studies conducted by Brookings and a scan of the work of GPE partners in this space. Given the report’s link to GPE’s ongoing strategy development as well as a longer-than-anticipated timeline for completing the mini-studies and interviews with GPE partners, the timing of the report was delayed by seven months and was released in January 2020.

A4L’S Contributions to Short-Term Outcomes

With an understanding of the achievement of A4L’s outputs, we can then assess the extent to which A4L achieved its intended short-term outcomes (EQ2). As described above, this section will assess the effectiveness of each of A4L’s components, in turn.



Despite delays and shifts in the quantity of activities, the short- and medium-term outcomes, and the long-term goals of the A4L initiative remained consistent during its duration. In the short-term, intended outcomes were to provide clear recommendations to countries for how to strengthen learning assessment systems, provide an increased understanding of efforts to measure whole child development (WCD) or 21st century skills, and to facilitate increased measurement of socio-emotional skills. These efforts were meant to be mutually reinforcing, with some overlap between each of the project’s outputs and the three short-term outcomes.

Effectiveness of ANLAS

At the activity level, the ANLAS toolkit was the primary means of achieving the short-term outcome of providing clear recommendations to countries for how to strengthen learning assessment systems. According to project documents, the expected outcome of ANLAS was “a set of recommendations to inform the development and implementation of strategies to build and improve such systems as part of the wider sector planning process” and to “identify areas for improvement and make recommendations to support the development and implementation of education sector planning strategies.” ANLAS is a country-led, participatory process and recommendations were meant to fit the local context.

Box 1. Tools within the ANLAS toolkit²⁸

- Manual
- Analytical tools (analytical tables; synthesis tables)
- Process tools (stakeholder database, national team training agenda, national team training presentation, stakeholder and document mapping tables, implementation plan, risks and mitigation strategies template, budget template, stakeholder briefing presentation)
- Reporting and dissemination tools (dissemination strategy template, country report template, key findings template, key findings presentation template)

According to project documents, **ANLAS did successfully support the three pilot countries to undertake a comprehensive analysis of their national learning assessment systems and to develop recommendations that can be used to inform strategies to improve their assessment systems.** This was confirmed in key stakeholder interviews, where respondents agreed that the ANLAS tool provided recommendations for action that were unique to their context. As stated by one respondent, “the analysis made it possible to identify the weak points in the various learning assessment programs in our education system and to propose the necessary corrective measures.” A subsample of recommendations related to the context and coherence of the assessment system of each pilot country is available in Appendix 11.

Effectiveness of regional workshops

Support to the TALENT and NEQMAP regional networks was intended to contribute to two short-term outcomes in A4L’s logic model: to provide recommendations for how to strengthen learning assessment systems and to increase understanding of efforts to measure WCD.

Related to the first objective to provide clear recommendations on how to strengthen learning assessment systems, **many country-level stakeholders noted the value of regional workshops and knowledge exchange activities in facilitating peer learning as well as an awareness of best practices and intervention strategies from countries facing similar hurdles in strengthening learning assessment systems.** National government representatives spoke particularly highly of activities focused on strengthening the alignment and coordination across agencies working on curricula, teacher training, and assessment.

In terms of the second objective, one of the regional workshops (organized by NEQMAP) specifically targeted the theme of 21st century skills – in Manila on “Promoting Transversal Competencies Across Curriculum, Pedagogy and Assessment.” In addition, a number of knowledge products and knowledge-sharing activities, including webinars and study papers, addressed this topic (see Table 3).

Interviews with regional and country stakeholders suggest that these activities were effective for workshop participants to gain familiarity with key terminology and concepts around 21st century skills, as well as to build an awareness of efforts in other countries to integrate these skills in assessments and curricula. This is echoed in the findings of an independent evaluation of NEQMAP, which found that participating in workshops and research by individual NEQMAP members led to “a) an increase in understanding of the concept of transversal competencies, b) a better understanding of how

²⁸ See: <https://www.globalpartnership.org/content/toolkit-analysis-national-learning-assessment-systems-anlas>

assessment may be used for policy change, c) a better understanding of the issues of one's own country, and d) a desire to bring about change."²⁹

Effectiveness of efforts to integrate 21st century skills

As the third component of the A4L initiative was scaled back and no longer includes activities related to the uptake of classroom assessment tools focused on 21st century skills, **the third short-term outcome of increasing the measurement of socio-emotional skills has not been achieved as intended through A4L activities.**

There is some indication from key stakeholder interviews that the OAA project, conducted by the Brookings Institution with travel and convening support from GPE, did have some direct effects on countries in integrating 21st century skills into the design of their assessments, as well as advocating for the incorporation of 21st century skills into curricula and pedagogy. OAA activities, for instance, directly informed Senegal's rewriting of their curriculum, allowing policymakers to understand the types of skills they want to prioritize and allowed for more specificity in assessment frameworks. Similarly, Nepal was able to add items to the national assessment and advocate for incorporating 21st century skills into curricula and pedagogy.

In relation to the successes of the OAA project, GPE's level of contribution to an increased measurement of socio-emotional skills is likely small, though not insignificant. There is evidence that A4L slightly magnified the efforts of the OAA project through the convening of side meetings in the margins of regional network workshops as well as contributing knowledge and time to discussions on country selection and project design. But stakeholders noted that the OAA would have progressed in a similar vein without GPE's direct support, suggesting that GPE's level of contribution is limited.

It is important to note that interviews with key stakeholders revealed a secondary short-term goal for GPE – to signal the increasing and defined role GPE could play in supporting country partners around incorporating 21st century skills in their systems. Respondents agreed that **the landscape review was very successful in presenting a potential position for GPE and stimulating discussions about the strategic direction of GPE's priorities in assessment practices.** This effort is in line with increasing interest and demand from GPE country partners and also serves to enhance GPE's relevance and profile in the global space.

Contribution rating

The evaluation team used a contribution rubric to categorize the results of the analysis along three lines of inquiry: 1) Level of contribution; 2) Significance of outcome; and 3) Strength of the evidence underlying the claims. Each component is rated on a scale from low (red), medium (yellow), to high (green). The level of contribution refers to the extent the outcome would or would not have happened in the absence of A4L; significance is the importance of the effect (e.g., how many people benefited from A4L or the size or scale of the benefit); and the strength of the evidence refers to the decisiveness of the available evidence in relation to the outcome. A full description of the rating rubric is contained in annex 1.

²⁹ Independent Evaluation of NEQMAP

Contribution claim #1: A4L tools, knowledge products, and regional technical capacity workshops contributed to providing clear recommendations for how to strengthen learning assessment systems

Level of Contribution	Significance of Outcome	Strength of Evidence
The ANLAS exercise made a substantial contribution for the three pilot countries in providing clear recommendations based on their unique contexts. Regional workshops also provided opportunities for peer learning and knowledge sharing related to recommendations on how to strengthen learning systems, which magnified this contribution. It is unlikely that these countries would have been afforded the same opportunity to generate these recommendations, in the same way, without A4L.	ANLAS recommendations greatly benefited the three pilot countries. In addition, the TALENT and NEQMAP workshops brought together multiple agencies from many other countries in the Africa and Asia-Pacific regions, who benefitted from capacity-building efforts to strengthen learning assessment systems.	ANLAS documents provide a clear and definitive list of recommendations developed by each of the ANLAS pilot countries that can be used to inform strategies to improve their assessment systems. The value and relevance of these recommendations was confirmed in key stakeholder interviews, as well as the value of recommendations shared during regional workshops.

Contribution claim #2: A4L tools, knowledge products, and regional technical capacity workshops contributed to an increased understanding of efforts to measure WCD, such as 21st century and socio-emotional skills

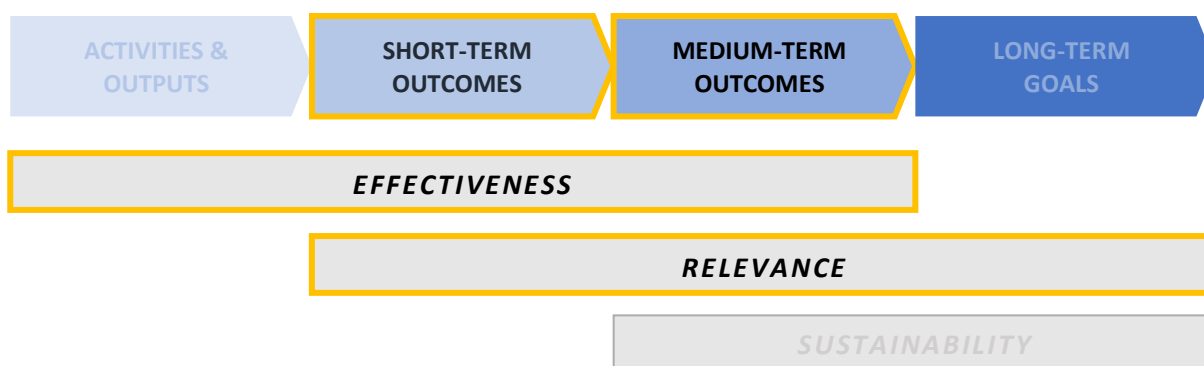
Level of Contribution	Significance of Outcome	Strength of Evidence
The combined effect of the three components of A4L – regional workshops, the ANLAS exercise, and the landscape review – contributed to facilitating awareness and an increased understanding of definitions and means of measuring 21 st century skills. However, since other organizations are also making significant contributions in this field, the evaluators believe that these efforts would have had similar effects in the absence of A4L.	The component on 21 st century skills in the ANLAS toolkit had a distinct effect on the knowledge and awareness of the three pilot countries. Participants of the regional workshops and knowledge-sharing activities that focused primary on 21 st century skills also noted that these efforts facilitated increased awareness, though these effects were more limited as only a small number of outputs focused on that theme.	An increased awareness of efforts to define and measure 21 st century skills was captured in stakeholder interviews. As ‘awareness’ and ‘understanding’ are difficult to measure by other means, the evaluators feel this evidence is best fit for this purpose.

Contribution claim #3: A4L tools, knowledge products, and regional technical capacity workshops contributed to an increased measurement of socio-emotional skills

Level of Contribution	Significance of Outcome	Strength of Evidence
<p>As the A4L initiative did not complete activities related to the uptake of classroom assessment tools focused on 21st century skills as originally planned, the project did not significantly contribute to the increased measurement of socio-emotional skills. Some activities under the guise of the Brookings OAA project were partially supported by GPE, but these efforts likely would have happened in a similar way in the absence of A4L.</p>	<p>While A4L’s contribution is likely small, the outcomes are not insignificant. For instance, activities were undertaken in Senegal and Nepal to incorporate 21st century skills into curricula and pedagogy.</p>	<p>The evaluators were not able to interview all members of the regional networks or participants in the OAA work to confirm the extent of the project’s reach, but all stakeholders that were interviewed were in agreement about the level of A4L’s contribution.</p>

A4L’S Contributions to Medium-Term Outcomes

As described above, this section will assess both the effectiveness and relevance of each of A4L’s components, in turn. In terms of effectiveness, the evaluators assess whether the project achieved its intended medium-term outcomes and its level of contribution to the goals of embedding good practices on learning assessment systems in countries, advancing tool development for the measurement of WCD metrics, and exposing countries to models for integrating holistic measures for learning (EQ2) as well as the extent to which A4L activities address the needs and priorities of country, regional and global actors (EQ6).



Effectiveness of ANLAS

In the A4L logic model, a short-term outcome produced by ANLAS activities – clear recommendations for how to strengthen learning assessment systems – was one of means of achieving the first medium-term objective to embed good practices on learning assessment systems in countries. As the objective of the

ANLAS diagnostic tool was to encourage countries to facilitate thinking and discussions about the ‘system’ of learning assessments and build a coherent sector strategy, the evaluation team interprets this goal of ‘strategic’ or ‘systems-level’ thinking as the primary “good practice” to be embedded at the country level as a result of the ANLAS piloting.

According to project documents, the three pilot countries aimed to use ANLAS as an input to the following activities as part of the sector planning process:

- Ethiopia: To develop the country’s next Education Sector Development Program (ESDP) VI plan (planned for 2020/21-2024/25); and to be used as a source document for curriculum revision, teacher training institutions, and teacher continuous professional development
- Vietnam: To develop the new Education Development Strategy for 2021-2030; and support the development of assessment regulations based on a new general education curriculum, to be piloted during the 2020-21 school year
- Mauritania: To inform the design of the Plan d’actions triennal (Triennial Action Plan) (PAT) 2019-2021, the Programme National de Développement du Secteur Éducatif (PDNES) III (2021-2030) and the country status report expected in 2020

Interviews with country focal points confirm that these efforts have been taking place as planned, which suggests that good practices have been embedded in each pilot country’s strategic thinking, understood in this sense as being preserved in sector plans and other strategic documentation. This success is partly owed to the thoughtful selection of pilot countries in alignment with the timing of sector planning processes. In all cases, the timing of the ANLAS pilot meant discussions and findings from the diagnostic exercise were able to feed directly into the upcoming sector plans. Extensive discussions among education ministries and the GPE Partnership within short-listed candidate countries also ensured sufficient buy-in from selected pilot countries to fulfil the exercise and integrate recommendations into national strategies.

Relevance of ANLAS

The extent to which these practices can be embedded in countries beyond the three pilot countries depends on the relevance of ANLAS to country’s needs and priorities. **Key findings from the document review and stakeholder interviews suggest that while the comprehensive nature of ANLAS is useful to countries, demand for the tool may be limited as noted in interviews with the GPE Secretariat and some government stakeholders.** While initial consultations around A4L indicated demand for such a tool, there were several similar tools developed by other organizations in the time between the initial consultations in 2013 and the publication of ANLAS in 2019. These efforts are described below.

First, some respondents felt that the ANLAS tool was “supply-driven,” rather than responsive to country interests and demands. While there may be a recognized need to promote system-level thinking, it is less clear if there is demand from countries for these diagnostic activities. Global, regional, and country-level stakeholders suggested that many governments are already aware of major shortcomings in their learning assessment systems and most, particularly **countries with more nascent systems, may not require the level of detail that ANLAS offers and limited funds may be better spent on more pressing issues.** This point is particularly relevant for Mauritania, whose large-scale learning assessment system was classified as nascent in 2018 according to data in GPE’s Results Framework (see Table 4). The system’s weaknesses may be behind the reason that the country required additional

financial support to hire the assistance of an external coordinator to assist the national team in undertaking ANLAS.

Table 4. Classification of international and national large-scale learning assessment systems in ANLAS pilot countries

	2015	2018
ETHIOPIA	Established	Established
MAURITANIA	Under Development	Nascent
VIETNAM	Established	Established

Second, some stakeholders felt that the ANLAS toolkit duplicates, rather than complements, other efforts in the sector. For instance, the ANLAS toolkit adds to a list of existing frameworks, tools and rubrics available to support the analysis of learning assessment systems, such as the World Bank’s SABER-Student Assessment initiative, the OECD’s PISA for Development Capacity Needs Analysis and the Network for African Learning Assessments (NALA)’s Learning Assessment Systems Evaluation Framework (LASEF). While SABER pre-dated the initial consultations on A4L, the other two tools were developed since 2013. On the other hand, some respondents saw value in the comprehensiveness of the tool in that it integrates 21st century skills and focuses on the alignment among curriculum, pedagogy and assessment systems.

Third, respondents suggested that the diagnostic tool is burdensome to national country-teams in that it is both time- and resource-intensive. Even in cases where there is interest and goodwill from countries, assessment teams are often stretched very thin and have limited time and bandwidth to undertake a months’ long exercise as required by ANLAS. This echoes sentiments in country documents where, for example, Mauritania mentioned that the process tested the “patience” of participants as multiple interviews with the same stakeholder are necessary to complete the exercise, and Vietnam detailed the “significant amount of time” it took to find and read documents carefully for the data collection process.

Effectiveness of support to regional networks

Via two short-term outcomes in A4L’s logic model – to provide recommendations for how to strengthen learning assessment systems and to increase understanding of efforts to measure WCD – support to regional networks aimed to contribute to two medium-term objectives – to embed good practices at the country level and to expose countries to models for integrating holistic measures of learning.

Through the platform of the NEQMAP and TALENT regional networks, countries have had multiple opportunities to gain technical knowledge, share examples of best practices, and learn from the experiences of peers. Support to these regional workshops have created a structure that allows for systematic knowledge exchange between countries rather than ad hoc engagements, ensuring that participants gain a full picture of country issues and strategies.

Significantly, some country stakeholders mentioned that regional workshops were the only means for national agencies that oversee different components of the education system – curriculum, assessment, and pedagogy – to meet and discuss cross-cutting issues such as 21st century skills and

alignment. These connections made outside of their countries have helped the participating countries connect better when they are back in their offices. They noted that bringing together curriculum, assessment and teacher training officials to develop a roadmap for assessment, for example at the TALENT workshop in Dar es Salaam, created a sustained structure for coordinating across these agencies once the participants returned to their home countries.

Countries from both the Africa and Asia-Pacific regions shared during key stakeholder interviews that they found the workshops to be incredibly valuable for gaining insight into different models and frameworks to strengthen their learning assessment systems, both countries with nascent systems and those with more well-established sector strategies. There is also some evidence that countries have taken this knowledge and embedded it at the country-level. For instance, Bhutan is now developing its first national education assessment framework using knowledge gained from the technical workshops about item development, sampling, test implementation, and dissemination of assessment results. Similarly, Cambodia has also made efforts to strengthen its assessment framework and improve communications across departments within the Ministry of Education.

Some stakeholders, however, mentioned that the fast pace of A4L deliverables made embedding work within countries more difficult since all available time had to be dedicated to planning the next workshop. This was particularly the case for the TALENT network, due to the added difficulty of coordinating a larger number of countries across a wider area compared to the Asia-Pacific region.

Relevance of support to regional networks

National, regional, and global stakeholders all agreed that GPE support to the regional networks filled a significant gap, particularly as this is an area of work where not many other donors show interest. Regional workshops are resource intensive, and financial support from GPE allowed networks to greatly expand the scope and scale of work.

Regional work is also very relevant and responsive to demand from countries. Topic areas of the regional workshops aligned very well to the interests and needs of countries, particularly interest in strengthening alignment among curricula, teaching training and assessment, strengthening classroom assessments, and gaining technical skills for administering large-scale assessments. Demand from countries is also evidenced in that both TALENT and NEQMAP have received requests for additional support, and several countries are mobilizing their own funding in order to participate. Cabo Verde, South Africa and several other countries sponsored national delegates to participate in the TALENT capacity-building workshop on effective reporting, dissemination and use of large-scale learning assessments.

However, some felt that funding provided to the TALENT network was not commensurate with need or the ambition of the project, and support could have been made more relevant to the African context. Both NEQMAP and TALENT received the same amount of financial support despite higher operational costs in Africa for coordinating meetings and activities, and that NEQMAP was well-established prior to receiving support from GPE, while TALENT was a new network that would require additional financial support to commence operations. For instance, when the first tranche of GPE funding was delayed, the

NEQMAP Secretariat was able to mobilize and advance funds from other sources to support the implementation of planned activities.³⁰

Effectiveness and relevance of efforts to integrate 21st century skills

While it is clear that the landscape review of 21st century skills was insufficient to achieve the third medium-term outcome – advancing tool development for the measurement of WCD metrics – it was shown earlier that **the landscape review of 21st century skills has been effective in its goal as a signal to country partners and the global community about GPE’s interest in this space and as an input into GPE’s strategic thinking.** Though discussions are ongoing, stakeholders suggest that 21st century skills may be featured more prominently in GPE’s upcoming strategic plan, due in part to the landscape review exercise, which raised awareness and stimulated internal discussions. The new 2025 GPE goal will be “To accelerate access, learning outcomes and gender equality through equitable, inclusive and resilient education systems fit for the 21st century.”

However, the relevance of the document beyond internal GPE processes has been questioned by different stakeholders. Some felt that the report drew heavily on existing analysis (for instance, the Brookings OAA work) and its value to the sector was unclear, while others suggested that the working definition of 21st century skills is not clear enough to be relevant for country-level work. **This suggests that the impact of the landscape report is limited outside the GPE Secretariat.**

Contribution rating

The evaluation team used a contribution rubric to categorize the results of the analysis along three lines of inquiry: 1) Level of contribution; 2) Significance of outcome; and 3) Strength of the evidence underlying the claims. Each component is rated on a scale from low (red), medium (yellow), to high (green). The level of contribution refers to the extent the outcome would or would not have happened in the absence of A4L; significance is the importance of the effect (e.g., how many people benefited from A4L or the size or scale of the benefit); and the strength of the evidence refers to the decisiveness of the available evidence in relation to the outcome. A full description of the rating rubric is contained in annex 1.

Contribution claim #4: A4L contributed to embedding good practices on learning assessment systems in countries

Level of Contribution	Significance of Outcome	Strength of Evidence
A4L’s ANLAS exercise, regional workshops, and regional knowledge-sharing activities have had a significant contribution to the ability of countries to embed good practices, including the timely analysis of system weaknesses, facilitation of cross-agency coordination, and skills-building exercises. A4L’s support to regional networks, in particular, filled a distinct gap and	Many countries have benefited from A4L support, but impact could be deepened further. In the case of ANLAS, countries with more mature systems may benefit more than more nascent systems and regional networks are not equipped to provide targeted capacity building.	The evaluators are basing this assessment on key informant interviews, which is the best evidence currently available, but subject to bias and other limitations. Stronger evidence of the claim could be seen in official planning documents submitted by the three ANLAS pilot countries when they become available, or other official documentation from

³⁰ Assessment for Learning (A4L) Grant Implementation Progress Report for Regional Assessment Networks. NEQMAP (January 2018-June 2018)

responded to country need and demand. While a number of organizations are working with countries to embed good practices on learning assessments (for instance, in designing diagnostic tools) outcomes would likely be different in the absence of A4L.		participants of regional workshops.
--	--	-------------------------------------

Contribution claim #5: A4L contributed to advancing tool development for the measurement of WCD metrics, such as 21st century and socio-emotional skills

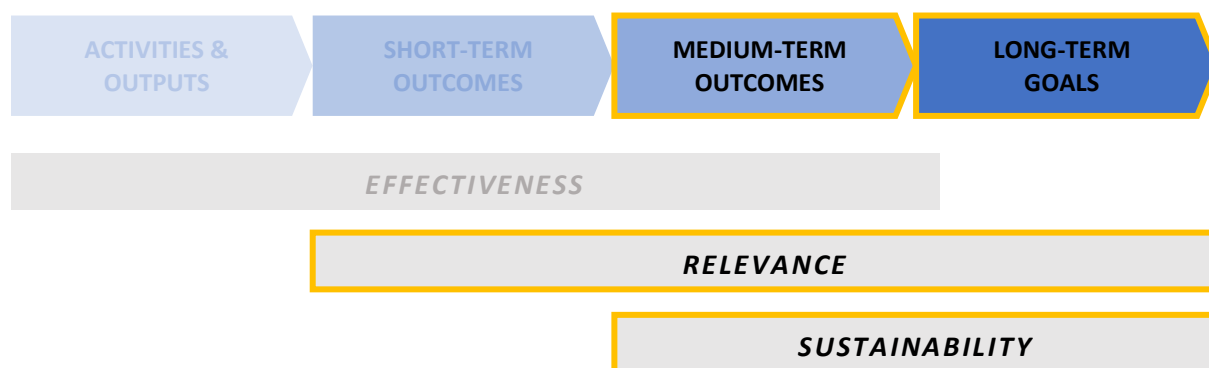
Level of Contribution	Significance of Outcome	Strength of Evidence
As the A4L initiative did not include activities related to the uptake of classroom assessment tools focused on 21st century skills as originally planned, A4L did not contribute to the advancing tool development for the measurement of WCD metrics.	N/A	N/A

Contribution claim #6: A4L contributed to exposing countries to models for integrating holistic measures for learning

Level of Contribution	Significance of Outcome	Strength of Evidence
A4L made a limited contribution to exposing countries to models for integrating holistic measures for learning, primarily through support provided to regional networks for technical workshops, knowledge products, and knowledge sharing activities. The application of 21 st century skills was also a cross-dimensional feature of ANLAS.	Several countries benefited from participating in additional OAA convenings on the sidelines of the regional workshops as well as content on OAA work within the main agenda of regional workshops (supported financially by GPE), which provided exposure to models for integrating holistic measures of learning. The inclusion of 21 st century skills as a cross-dimensional feature in ANLAS further exposed the three ANLAS pilot countries to strategies for integrating holistic measures for learning.	The extent to which countries were exposed to models was captured primarily through stakeholder interviews but could not be confirmed or refuted by other means.

A4L'S Contributions to Long-Term Goals

This section assesses the extent to which A4L contributed to its intended long-term goals of strengthening learning assessment systems and promoting the measurement of socio-emotional skills at the national and global levels (EQ3), which, as described above, depends in part on the relevance and sustainability of A4L's medium-term outcomes. In terms of relevance, the evaluators assess the extent to which A4L complements GPE's support at the country level (EQ7) and ongoing global efforts (EQ8). In terms of sustainability, the evaluators assess whether A4L allowed for new or strengthened collaboration among actors at the global, regional, and country levels (EQ10), and whether the initiative catalyzed support and investment in the areas of learning assessment systems (EQ9). While it is too early in the project's lifecycle to concretely determine success toward A4L's long-term goals, the evaluators believe that these conditions lay the foundation for strengthening learning assessment systems in the longer term.



Overall relevance of A4L

All country-level stakeholders cited lack of technical capacity, human resource capabilities, and financing as primary barriers to reform in their learning assessment systems, suggesting targeted support to one or all of these challenges will be required moving forward. Some others also mentioned the absence of a national institution in their countries with the mandate to carry out assessments supported by political will to share results.

Country, regional, and global stakeholders all agreed that A4L laid a good foundation for the strengthening of learning assessment systems worldwide, but further capacity building is needed to truly make a long-term impact at the country level. As put by one interviewee, "A project ... has a lifetime, but capacity building helps to own and sustain a project's objective." Early successes – such as increased awareness of challenges related to learning assessment, the compilation of good practices, and giving confidence to countries that reform is possible – will need to be deepened through additional capacity building efforts and support for reforms. Some stakeholders warned that the impact of these types of smaller-scale projects are limited without being paired with future support, particularly at the country level.

For instance, the long-term success of support to regional workshops could be limited without additional efforts from GPE to strengthen national capacity. Regional workshops (by design) tend to be one-off events, and as such this limits their effectiveness as a vehicle to support long-term impact at the

country-level. While the regional networks are valuable for sharing knowledge and building awareness, only 2-3 participants from each country can attend workshops, and often different representatives attend different workshop, limiting opportunities to deepen knowledge. Similarly, the duration of the workshops is short and there are few opportunities for hands-on experience or follow-up engagement. For example, participants of a workshop in the Philippines highlighted the value of an exercise to develop a strategic plan but were disappointed that there was no follow-up activities or accountability regarding the planned actions. These aspects limit the extent to which an assessment agency in a country can build the necessary skills to implement sustainable reforms. Additionally, technical workshops, understandably, target technical audiences, but the participants are often not in the position to make critical decisions about reform priorities or strategies.

Relevance of A4L to GPE country-level support

For some stakeholders, the pathway for integrating A4L at the country-level has not been specified and it is unclear whether A4L complements GPE's support at the country level. Many stakeholders see a distinct disconnect between A4L's project-based work and GPE's financing model that takes a system-wide approach. Representatives from the GPE Secretariat's country teams shared confusion for what their role should be in relation to A4L's products – whether as a neutral arbiter of discussions at the local level to assist countries in making their own decisions on priorities and sequencing, or to promote the use of a specific tool or methodology. Some stakeholders viewed promoting a specific tool or methodology as not aligned with GPE's mission.

For some country partners, it is also unclear what mechanisms exist to request additional support for A4L work at the national level if there is continued demand. For instance, some countries that expressed interest in using the ANLAS tool were unsure of possible next steps or how to ask for assistance in integrating the use of the diagnostic into their sector planning processes. This could have been due to lack of internal communication and coordination within the Secretariat, or misalignment between the grant cycles and the period in which the ANLAS tool was developed. Representatives from the TALENT and NEQMAP regional networks also saw a disconnect between technical workshops and partners working in GPE countries. They noted that it could be valuable if ESPIG grant agents and representatives from LEGs participated in workshops.

As paraphrased from one respondent, GPE needs to map out its intention or goals around lifting the results and learning from A4L in the future – through the potential of ANLAS being a core part of the ESP process; in considering how the Secretariat can support DCPs around their efforts to embed 21st century skills that emerged from the landscape report; and in supporting UNESCO Bangkok and Dakar in their capacity building efforts around assessment in their respective region. They noted that these strands should not be one-off investments but rather part of a bigger picture. This aligns with GPE's goals of this evaluation, to incorporate lessons learned from A4L into the strategic planning process.

Relevance of A4L to global efforts

In terms of whether A4L complements ongoing global efforts to strengthen learning assessment systems, project documents note that **the senior education specialist has been actively involved in a number of global engagements related to learning assessments and metrics**, including the Global Alliance to Monitor Learning (GAML), coordinated by the UNESCO Institute for Statistics (UIS) to support national strategies for measuring learning and international reporting against learning- and skills-related

indicators of SDG4.³¹ The GPE Secretariat serves on GAML’s Strategic Planning Committee as well as several of its task forces and participates regularly in meetings. In fact, the A4L project was designed initially to complement the mandates of UIS and GAML. Some initial A4L objectives were downsized in response to the extent of activities undertaken by GAML, and the scope of the initiative was narrowed to prevent duplication.

The senior education specialist and one consultant also engage with a number of other global partners, including the OECD, IEA, USAID, UNICEF, UNESCO-IIEP, the World Bank, and the different regional assessment programs and the organizations that serve as their Secretariats (e.g., PASEC/CONFEMEN, PILNA/EQAP).³² The GPE Secretariat holds regular check-in discussions with all of these partners and actively socializes A4L activities and outputs amongst them.

Sustainability of A4L

The extent to which A4L contributes to the long-term goals depends not only on the relevance and relevance of the initiative’s components, but also on their sustainability, otherwise understood as the extent to which A4L catalyzed support and investment in the areas of learning assessment systems (EQ9) and whether A4L allowed for new or strengthened collaboration among actors at global, regional, and country levels (EQ10).

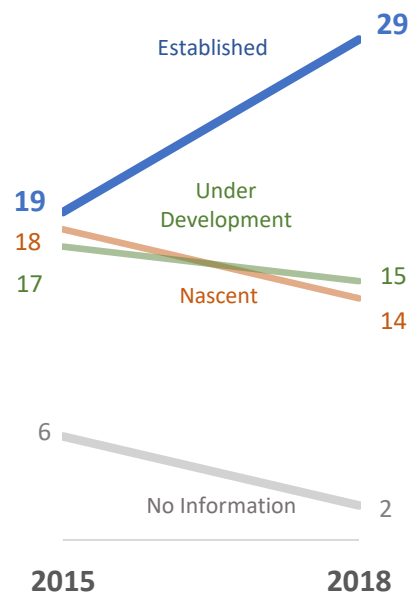
Positively, many countries have noted that in the years previous to the launch of A4L, learning assessments gained prominence in education sector plans as well as the recognition of the need for assessments to address learning crisis. This is evidenced by the growth in the number of DCPs characterized as having an “established” learning assessment system between 2015 and 2018 (Figure 4). However, further research is needed to determine the extent to which the plans for learning assessment systems translate into implementation, and ultimately into improved learning outcomes.

But given the relatively small scale of A4L, respondents from the GPE Secretariat and global stakeholders questioned the extent to which A4L can catalyze additional support for learning assessment systems beyond the GPE DCPs that directly engaged with the initiative. Some stakeholders also saw the design of A4L as a missed opportunity to advocate for more financing for learning assessments to amplify the efforts of others in the space. However, ANLAS has the potential to influence ESPs in the three pilot countries and beyond, which have the potential to effectively increase ESPIG funding for learning assessments.

³¹ Porticus 2019 Annual Report

³² Porticus 2018 Annual report

Figure 4, Classification of learning assessment systems of GPE partner countries



Source: Results framework data from 2019 GPE Results Report

There is some indication that A4L has catalyzed support for learning assessment systems within the GPE Secretariat. As noted earlier, A4L activities, particularly the landscape review of 21st century skills, have stimulated discussions about the strategic direction of GPE’s priorities and have been included in negotiations regarding GPE’s next strategic plan. There is also now a permanent position for a thematic lead to sustain momentum and broaden the portfolio of work. Learning assessment is one of the six thematic areas in the Knowledge and Innovation Exchange (KIX), and GPE recently announced a US \$2 million grant to People’s Action for Learning (PAL) Network, Pratham Education Foundation and the Australian Council for Education Research (ACER) to develop a common-scale assessment of early and foundational math learning across the Global South. Several members of the Secretariat said that the support by foundations and the program specialist position afforded by A4L helped create the space for learning assessment systems to be featured as a GPE thematic area.

According to project documents, there are plans to use this permanent staff position to further disseminate and embed A4L’s work. Now that the ANLAS toolkit is available as a global public good, the A4L team plans to sensitize country-level stakeholders about the possibility of using GPE funds to conduct the ANLAS diagnostic exercise. However, the degree to which the GPE Secretariat will further integrate or expand on A4L activities is unclear. The sustainability of the regional networks is particularly uncertain, as GPE has “been more or less the sole funder of the networks’ activities.”³³

In terms of strengthening collaboration at the national, regional and country levels, stakeholders noted that there is a need for better synergies and coordination between the different initiatives and actors. Some global stakeholders felt that engagement across different global agencies came too late in

³³ Porticus 2019 Annual report

the process and appeared to be more of a “courtesy” rather than truly engaging them to design initiatives and products that complement each other’s purview, scope, and level of access. Respondents felt that communication could be further formalized and established at all stages of project development. They noted that the area of 21st century skills is a space that could particularly benefit from expanded and strengthened partnerships, as it is a comparatively nascent and contested space.

The need for better coordination is particularly true at the country level where national stakeholders are sometimes uncertain how different efforts by multiple country partners (ex: GPE, World Bank, regional and international assessment networks) interact and complement each other. Some country stakeholders noted that they were able to, for instance, leverage trainings and technical support from international and regional assessment organizations (ex: PISA-D, SEA-PLM, PASEC/CONFEMEN) to strengthen their ability to engage in the TALENT and NEQMAP regional workshops, but this degree of complementarity is ad-hoc and dependent on individual country efforts.

Contribution rating

The evaluation team used a contribution rubric to categorize the results of the analysis along three lines of inquiry: 1) Level of contribution; 2) Significance of outcome; and 3) Strength of the evidence underlying the claims. Each component is rated on a scale from low (red), medium (yellow), to high (green). The level of contribution refers to the extent the outcome would or would not have happened in the absence of A4L; significance is the importance of the effect (e.g., how many people benefited from A4L or the size or scale of the benefit); and the strength of the evidence refers to the decisiveness of the available evidence in relation to the outcome. A full description of the rating rubric is contained in annex 1.

Contribution claim #7: A4L contributed toward its long-term goal of strengthening learning assessment systems, where data is used by teachers, schools, and government officials to inform instruction, teacher training, and system-wide policies to improve learning for all children

Level of Contribution	Significance of Outcome	Strength of Evidence
<p>Given the project’s influence on GPE’s continued work in the area, the evaluators believe that A4L laid the groundwork to contribute to the long-term goal of strengthening learning assessment systems, particularly because of GPE’s value-add in relation to sector planning and financing. These efforts could be strengthened even further with increased efforts to enhance coordination at the country level.</p>	<p>Many GPE partner countries, as well as non-GPE partner countries that participate in regional networks and are served by global public goods, have benefited from A4L’s activities. The reach of the project was substantial given the comparably small amount of funds but could be strengthened through targeted capacity-building efforts and increased clarity on alignment with GPE’s country-level processes.</p>	<p>Evidence of A4L’s effect was confirmed through stakeholder interviews, and triangulated through other means, including a questionnaire, document review, and web scan.</p>

Contribution claim #8: A4L contributed toward its long-term goals of the measurement, monitoring, and promotion of socio-emotional and other 21st century skills at the national and global levels as part of a holistic measure of learning

Level of Contribution	Significance of Outcome	Strength of Evidence
<p>Given that 21st century skills is a space in which many organizations are currently operating, the extent of A4L's direct contribution toward the long-term goals of the measurement and promotion of 21st century skills is likely limited. The inclusion of 21st century skills in ANLAS provides an opportunity for future growth in this area</p>	<p>The evaluators believe that A4L had a small but critical effect on the promotion of 21st century skills, particularly in signalling GPE's future priorities and in informing the development of KIX. This has the potential to affect a large number of partner countries and beneficiaries in the longer term.</p>	<p>The evaluators are primarily basing this assessment on key informant interviews, which is the best evidence currently available, but subject to bias and other limitations.</p>

PART 2 - EFFICIENCY ANALYSIS

The evaluation examined the efficiency of A4L, in particular the extent to which the management and coordination of A4L among the GPE Secretariat, implementing agencies, and funders been well-organized, consistent and efficient in terms of staff-time and resources (EQ11). We were also interested in the extent to which the Secretariat has been efficient in terms of communication and dissemination around A4L activities and achievements (EQ12). Finally, we examine the extent to which the collaboration with foundations has added value for GPE in supporting work in specific thematic areas (EQ13).

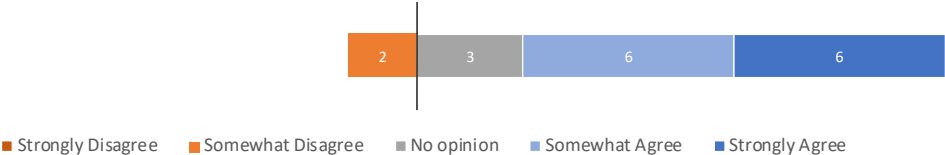
Management

Overall, A4L stakeholders felt the day-to-day management of the initiative was well-organized. The A4L team at the Secretariat, especially the senior program specialist, was well-regarded overall as efficient and having good leadership. The majority of survey respondents (15 of 17, or 71%) felt that the day-to-day management was well-organized, and only one respondent thought that the goals and objectives of A4L were not consistent during its implementation.

The goals and objectives of the A4L initiative remained consistent during its implementation



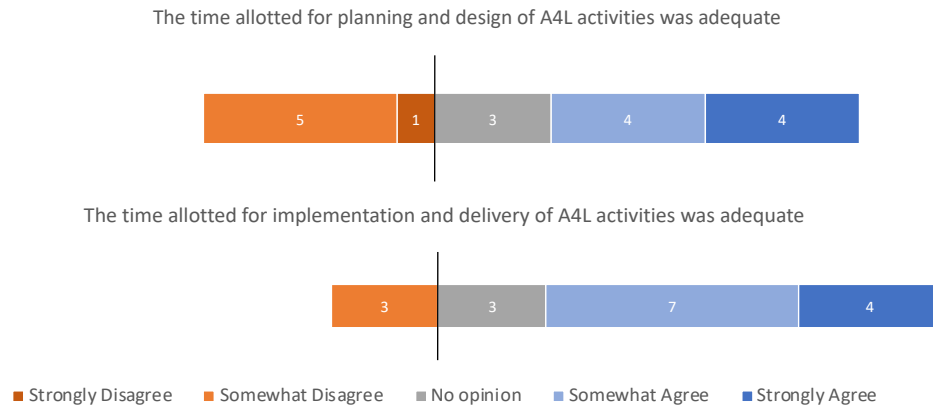
The day-to-day management of the A4L initiative was well-organized



One theme that arose in multiple interviews was the high transaction costs in terms of time spent managing the grants and relationships. This included time spent negotiating the grants between GPE and the funders, and time spent negotiating and managing the sub-grants to the implementing agencies.

While the GPE was flexible in terms of accommodating delays from partners and granting no-cost extensions to the regional networks, **implementing partners said they would have liked to see even more flexibility in their agreements with GPE**, for example to adjust the timing of certain activities or hire additional staff so they could better focus on implementation.

Survey respondents were split in their perceptions of whether the time allocated for planning and design of A4L was adequate, with 6 of 17 saying it was inadequate, 8 saying it was adequate, and 3 having no opinion. More respondents felt that the time allotted for implementation and delivery was adequate (11 of 17).

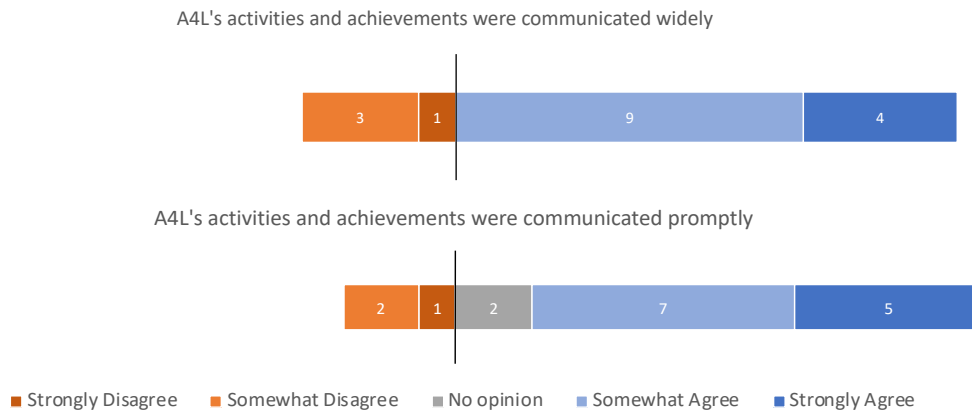


One of the reasons why the transaction costs were perceived to be high on the GPE Secretariat side are that the detailed monitoring and reporting required by one of the foundations was not complementary to GPE’s operating model. The other foundation had very little engagement with the GPE Secretariat and the transaction costs were viewed as manageable. The A4L team at the Secretariat made progress on this issue midway through the project by scheduling monthly check-ins with funders, to reduce the number of ad hoc requests they were responding to.

Some of the issues with management stem from a misalignment of expectations early in the initiative. One of the funders considered themselves to be an engaged partner within the A4L initiative but members of the Secretariat noted that the GPE model was not set up to accommodate that level of involvement.

Communications

The Secretariat produced 14 blogs, 2 publications (ANLAS toolkit and landscape review), 12 presentations and 3 webinars on A4L. The initiative was also mentioned by other organizations in at least 20 reports, blogs and articles. Of the 17 A4L stakeholders who participated in the survey, 13 (76%) felt that A4L’s activities were communicated widely and 12 (71%) felt they were communicated promptly, as shown below.



The A4L initiative was promoted at a number of conferences, including the 2018 CIES conference in Mexico City in March 2018, where GPE organized a panel including colleagues from UNESCO Bangkok, Brookings and the GPE Secretariat. The GPE Secretariat also presented A4L at a South Asia-focused conference on “Using Large-Scale Assessments to Improve Teaching and Learning”, hosted by the World Bank in New Delhi in April 2019. The A4L team also made presentations directly to the government of Djibouti, and country leads for seven additional countries have made requests for information about learning assessment systems on behalf of their DCPs to the A4L team.

The ANLAS toolkit was published in October 2020 on the GPE website, with an accompanying blog. Several partners also published blogs announcing the toolkit. A webinar was hosted by GPE on November 18, 2019 which was attended by 65 participants. The landscape review on 21st century skills was published on the GPE website in January 2020, along with an accompanying blog. The review was further disseminated via a brown bag lunch for GPE staff members and Washington DC-based colleagues, as well as an external webinar organized in May 2020 which was attended by 82 participants.

The A4L team wrote a chapter in the UIS’s SDG 4 Data Digest 2018 and was also closely involved in the writing of the 2019 GPE Results Report, which had a thematic focus on learning. This included co-authoring of the first chapter (on learning), which included a substantive section on GPE’s support in the area of learning assessment systems, citing A4L.

Table 5 shows the GPE communications activities for A4L.

Table 5. GPE Communication Activities for A4L

Communications product/event	Date	Audience
Blogs (14), plus mentions in others	Throughout initiative	Global education community
Presentations at regional workshops (6)	1. Bangkok, December 2017 2. Bangkok, March 2018 3. Dar es Salaam, July 2018 4. Penang, September 2019 5. Addis Ababa, July 2018 6. Bangkok, November 2019	Government officials
Presentations at international conferences	March 2018 (CIES), April 2019 (World Bank New Delhi)	Global education community
Brown Bag Lunch (2)	April 2019, January 2020	GPE Secretariat
ANLAS Toolkit	October 2019	Government officials
ANLAS Webinar	November 2019	DCP representatives, global education community
Presentations to DCP governments (1)	2019	Government officials
Webpage on A4L	2020	Public
Landscape Review of 21 st century skills	January 2020	GPE Secretariat and Board

Landscape Review Webinar	May 2020	DCP representatives, global education community
--------------------------	----------	---

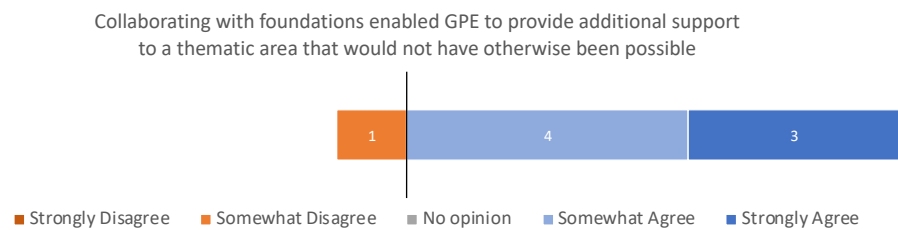
In addition to the GPE-published documents, there were at least 21 external mentions of A4L beginning in 2015 through 2017, with many in the global education community expecting it to be the implementation arm of GAML (See Appendix 12 for the web scan results). In 2018-2020, the web references were mostly generated by GPE and the implementing agencies, and consisted mostly of descriptions of workshops and meetings. The Ethiopian National Educational Assessments and Examinations Agency (NEAEA) published a brief article describing the ANLAS stakeholder briefing held in Addis Ababa in January 2019 and linked to their ANLAS report.

Several interviewees requested further dissemination around ANLAS, with materials that help countries understand how to get started in using the tool. The Secretariat is finalizing the French and Spanish translations of ANLAS in July 2020, which is intended to help with dissemination and uptake of ANLAS in additional countries and will publish these along with a blog.

Collaboration with foundations

Since working with philanthropic donors on a project like A4L is a new approach for GPE, this evaluation also examined to what extent the collaboration with foundations has added value for GPE in supporting work in a specific thematic area. As one funder noted, “the challenge as a philanthropic funder is to avoid forcing GPE into a project level reporting, since that defeats the very purpose of GPE.”

Overall, collaborating with foundations was considered to be an important enabler of additional resources to the learning assessment thematic area, with all but one survey participant agreeing that the activities would not have been possible without foundation support.



Several interviewees mentioned that the foundation funding on the theme of learning assessment was instrumental in allowing GPE to hire a program specialist and build a thematic area on learning assessment. It signaled the importance of a theme that is increasingly gaining traction both globally and at the country-level, and is fundamental to the current GPE funding model. Strengthening Learning Assessment Systems is one of the priority thematic areas of KIX, a new initiative that supports global goods, research, innovation and exchange of experience across the Partnership.

Notwithstanding these benefits of collaborating with foundations, many stakeholders felt that the transaction costs were simply too high, and the GPE model does not allow for in-depth engagement with donors on a project. Several suggestions of how foundation donors could have an impact while supporting the GPE model include supporting organizations in the LEGs to help the government

implement the parts of their ESPs related to learning assessment, and directly supporting the regional hubs like TALENT and NEQMAP.

LESSONS LEARNED AND RECOMMENDATIONS

A key line of enquiry for this evaluation is to explain aspects of the project that contributed to its success and/or shortcomings (EQ5) as a means to shape future GPE programming related to learning assessment systems. This section shares these key lessons learned as well as a series of recommendations for the GPE Secretariat.

What worked well

Key drivers of A4L's success included its leadership and management, particularly the project lead. Despite significant delays at the beginning of the project, implementation remained relatively on track and all outputs from components 1 and 2 were achieved as expected. Providing funding for a permanent senior education specialist and thematic lead for learning assessment has also allowed GPE to meaningfully participate in global and regional discussions related to learning assessment. The specialist also serves as the thematic resource within the GPE Secretariat for all issues and inquiries related to learning assessment systems and is currently working to support Djibouti but expects much more demand for advising on learning assessment issues by country leads on behalf of the DCPs they work with in 2020 and beyond.

Another driver of success was the thoughtful selection of ANLAS pilot countries based on needs and alignment to timing with sector planning processes. All of the government stakeholders who participated in A4L activities reported significant needs related to strengthening learning assessment, and these needs aligned with the features of an assessment system put forth in ANLAS. Country stakeholders also welcomed the timely and relevant support to regional networks as a structured mechanism for peer learning and knowledge exchange related to joint challenges faced in reforming learning assessment systems.

The technical quality of partners such as ACER and Brookings, as well as the convening power of UNESCO Bangkok and UNESCO Dakar also strengthened the impact of A4L's activities. However, there was a tension between the technical strength of activities such as ANLAS and OAA and the accessibility and simplicity of tools requested by DCPs.

What hindered impact

Participants in the evaluation noted that while there were extensive negotiations with the funders prior to the start of the initiative, there was the limited time offered during the grant period for project development with implementing partners and funders, including necessary discussions to align expectations. Similarly, the equal allocation of funding across regional implementing agencies was not commensurate with the different needs of the organizations or the ambition of the project.

There was also a misalignment between the perceived promotion of one thematic area (i.e., strengthening learning assessment systems) over others, and how that related to the neutral role GPE is expected to play in facilitating the development of the education sector planning process.

While there was overall demand for A4L activities as noted by the GPE board members approving the initiative, the degree to which there was demand for the activities in the specific countries where they were implemented was questioned.

For one of the A4L funders, there were high expectations on the level of the funder's engagement in the initiative and the amount of communications and other requests that were expected of the A4L team. This was in contrast to another funder, whose limited requests to the GPE Secretariat helped ensure that staff time was spent on project implementation and not donor relations and reporting.

Recommendations

Building from these lessons, the evaluation team makes the following recommendations to the GPE Secretariat:

Leverage the A4L activities and tools to provide direct support to the education sector planning process, potentially using a BELDS-like model. The evaluators note a significant contrast in the two initiatives of GPE's KGPE strategy. While BELDS was focused on harnessing global momentum, knowledge, and tools to strengthen processes for mainstreaming ECCE into ESAs and ESPs, A4L focused on building the tools, awareness and political will as outcomes in themselves, partly because of A4L's long legacy as an action agenda prior to GPE's launch of the project. Now that the tools are developed and the dialogue is underway, there is an opportunity to make a more direct link between this thematic area and the ESP process. This should include increased sensitization of ongoing activities and available tools for Secretariat Country Leads.

Use KIX to fill a global gap for knowledge and innovation around 21st century skills. Several A4L stakeholders noted that the area of 21st century skills is a space that could particularly benefit from expanded and strengthened partnerships, as it is a comparatively nascent and contested space. This could cut across the learning assessment and teaching and learning thematic areas. KIX can be a venue for further dissemination of resources produced under A4L, including research products produced by the regional networks, and offer an opportunity for learning exchange and linkages with national education sector planning processes.

Consider adapting a "lite" or "screening" version of ANLAS as an entry point for countries with nascent learning assessment systems. This would provide a way for DCPs to engage with the toolkit to identify high-level needs to incorporate into their ESPs, with the expectation that the full administration of ANLAS would come at a later time.

Support deepened regional capacity building efforts on learning assessment systems. While the workshops were considered useful by government participants, linking these workshops with ongoing, just-in-time technical support when they are designing, administering, and analyzing and reporting results from assessments could help ensure that the knowledge shared is embedded in practice. This is another activity that could be supported through the KIX learning exchanges.

Amplify research and knowledge products generated by regional networks on GPE's website and other external and internal communications mediums. A wealth of knowledge has been created, and continues to be created, that is varied and comprehensive enough to be relevant across many country contexts. GPE should take the opportunity to expand the audience of this work.

Set guidelines for how GPE engages with private foundations. These could include the minimum contribution level (currently US \$2 million as noted by the GPE Secretariat), minimal reporting requirements, expectations around the degree of engagement foundations will have with GPE, and boundaries on when and how foundation staff engage with GPE staff (for example, by nominating a single representative to act on behalf of all private funders engaged on a particular issue area). Within the guidelines, we recommend building in a sufficient inception phase for planning and negotiating donor-funded initiatives and setting expectations about the amount of detail and planning that will be provided for proposal vs. the inception report.

APPENDICES

Appendix 1. Additional details about contribution analysis and matrix key

As A4L is working in a global space where a number of complementary initiatives are operating, the evaluation adopted a theory-based approach using Contribution Analysis (CA). This approach, which is a blend of various theory-based methods such as process tracing and outcome harvesting, examines the complementarity or relevance of the A4L initiative alongside other influencing factors.

CA examines the theory of change against results observed and builds a credible contribution story by demonstrating plausible associations between inputs and subsequent changes.³⁴ The overall aim is to understand why results did or did not occur and the role the project played in the larger landscape of influencing factors.

The evaluators took the following steps to infer causality from the A4L activities:

- 1) Scoped the causal relationships found in the theory of change and assess the plausibility of the expected contribution given the intervention size and reach.
- 2) Collected evidence against the following 4 dimensions:
 - a. A4L supported activities (to assess whether programs were implemented as planned);
 - b. Intended outcomes (whether they were observed or not observed);
 - c. How and why change happened; and
 - d. Other factors that might have influenced the outcomes, and their relative importance
- 3) Assembled and assessed the contribution claim, and challenges to it. The 8 claims emerged from the project’s detailed logic model and were confirmed with a review of the project documents and through key stakeholder interviews

The evaluation team used a contribution rubric to categorize the results of the analysis along three lines of inquiry: 1) Level of contribution ;2) Significance of the outcome; and 3) Strength of the evidence underlying the claims. Each component is rated on a scale from low, medium, to high.

Contribution Matrix Key.

	Contribution Rating	Significance Rating	Evidence Rating
High	This outcome could not have happened without your actions	It’s newsworthy. Many people are believed to have benefit from the change. And/or a large number of people are believed to have benefit a lot.	The evidence is an excellent fit for the description of the component and shows a strong connection with the outcome. It’s very unique to your explanation, and it would be very hard to explain the evidence unless your explanation is correct.

³⁴ Befani, B. and O’Donnell, M. (2016): Choosing Appropriate Evaluation Methods: A Tool for Assessment and Selection. Bond. Available at: https://www.bond.org.uk/sites/default/files/caem_narrative_final_14oct16.pdf

<p style="text-align: center;">Medium</p>	<p>You made a substantial contribution to a key part of the outcome, and you believe it would not have happened in the same way without your efforts. Other actors also played a substantial contribution to the outcome.</p>	<p>It's important for a reasonable number of people. Or some people may have benefited a great deal</p>	<p>The evidence is an okay fit for the description of the component. You need this evidence but it's not decisive. It shows some connection with the outcome, but it is quite possible other actors' efforts are also a reasonable fit for the component and have a comparably good connection to the outcome.</p>
<p style="text-align: center;">Low</p>	<p>The outcome would have probably happened anyway</p>	<p>Few people are believed to benefit from the change. Those that do benefit will benefit very little, and various people may also lost out or be negatively affected.</p>	<p>The evidence is not a good description of the component and shows no clear connection between actions and the outcomes.</p>

Appendix 2. List of stakeholders interviewed

ORGANIZATION	NAME	TITLE
GPE Secretariat		
GPE Secretariat	Jo Bourne	Chief Technical Officer/Manager of Education Policy & Performance (EPP) Team
	Raphaelle Martinez	Senior Education Specialist/Team Lead, Education, Policy & Learning (EPL) sub-team within EPP
	Jean-Marc Bernard	Senior Education Specialist
	Talia Miranda De Chaisemartin	Senior Education Specialist/Country Lead
	Ian Macpherson	Senior Education Specialist/Lead for KIX (situated in EPL/EPP)
	Stijn De Lameillieure	Lead for Engagement with Foundations/External Relations (EXR) Team
	Medjy Pierre-Louis	Consultant
	Ed Lamot	Senior Education Specialist/Country Lead for Mauritania (ANLAS pilot) and Senegal (TALENT)
	Adrien Boucher	Country Lead
Implementing Agencies		
Australian Council for Educational Research (ACER)	Ursula Schwantner	Senior Research Fellow
	Claire Scoular	Senior Research Fellow, Education Policy and Practice
	Jeaniene Spink	Research Director, Education and Development
Brookings Institution	Esther Care	Senior Fellow, Global Economy and Development
	Helyn Kim	Postdoctoral fellow (former)
Conference of the Ministers of Education of French speaking countries (CONFEMEN)	Hilaire Hounkpodoté	PASEC Coordinator
UNESCO	Mark Manns	Programme Officer, Section for Inclusive Quality Education, UNESCO Bangkok
	Jun Morohashi	Regional Programme Coordinator, Executive Office, UNESCO Bangkok
	Gwang-Chol Chang	Chief of Section at UNESCO HQ (Former Director of UNESCO Dakar)
	Davide Ruscelli	Programme Officer, UNESCO Dakar
	Valérie Djioze-Gallet	Education Programme Specialist, UNESCO HQ (Former Lead for Teaching and Learning at UNESCO Dakar)
	Claude Akpabie	Chief of Education, UNESCO Dakar
UNESCO IIEP	Hugues Moussy	Team Lead, Research and Development

	Ieva Raudonyte	Associate Research Officer, Research and Development
UNESCO IIEP-Pôle de Dakar	Patrick Nkengne	Education Policy Analyst
	Marcelo Souto Simao	Program specialist (former)
UNESCO Institute of Statistics	Silvia Montoya	Director
UNICEF	Tserennadmid Nyamkhuu	Programme Officer, Section for Inclusive Quality Education at UNESCO Bangkok (former)
Country Partners		
National Educational Assessment and Examinations Agency (NEAEA), Ethiopia	Mesaye Demessie	Deputy Director General, National Learning Assessment Directorate
	Yilikal Wondimeneh	Director, National Learning Assessment Directorate
Ecole Normale Supérieure, Mauritania	Mohamed Salem Tfeil Amar	Research Teacher
Ministry of National Education, Senegal	Cheikhna Lam	Director of Education Planning and Research
Ministry of National Education and Technical Education, Cote d'Ivoire	Joseph François Désiré Kauphy	Director of La Vieille et Suivi
Ministry for Basic and Secondary Education, The Gambia	Ousmane Senghor	Head of the Learning Assessment Unit
Bhutan Council for School Examinations and Assessments (BCSEA)	Kinley Dema	Education Monitoring Officer, Assessment and Monitoring Division
Ministry of Education, Youth and Sport, Cambodia	Ung Chinna	Director, Education Quality Assurance Department
	Sarin Sar	Chief of Office, Education Quality Assurance Department
	Hav Khou	Vice-Chief of Office, Education Quality Assurance Department
Ministry of Education, Nepal	Deviram Acharya	Section Officer, Education Review Office
External Stakeholders		
OECD	Michael Ward	Senior Policy Analyst
IEA	Dirk Hastedt	Executive Director
World Bank	Marguerite Clarke	Senior Education Specialist/Global Lead for Learning Assessment
	Julia Liberman	Operations Officer, Education Global Practice
PAL Network	Sara Ruto	CEO
	Ketan Verma	Lead on Assessments
Funders		
Porticus	Ryan Burgess	Global lead for Primary, Secondary and TVET Education
	Eileen O'Malley	Senior Analyst
Dubai Cares ³⁵	Anna Bertmar Khan	Director of Programs

³⁵ The representative from Dubai Cares was unavailable for a direct interview but provided responses to the emailed questionnaire.

Appendix 3. Generic interview protocol for GPE Secretariat

Introduction

1. Please briefly introduce yourself: What are your current roles and responsibilities at GPE? For how long have you been in this position?

GPE's involvement with learning assessments

2. In broad terms can you describe how well GPE is doing in the area of strengthening learning assessment systems?
3. What do you think are the key challenges for partner countries to implement reforms to strengthen learning assessment systems?
 - a. Where have you observed most progress? Are there specific factors that you think explain positive changes?

A4L's design

4. Please briefly describe your role(s) in relationship to the A4L initiative. How long were you in this role?
5. Can you describe how A4L fits within the larger portfolio of GPE support to strengthening learning assessment systems and/or promoting the measurement of social emotional skills?
6. What strengths and/or limitations in GPE support was A4L designed to respond to?

Effectiveness & relevance

7. What are the main objectives of A4L as you see them?
8. Did the objectives of the project evolve over time? If yes, what contributed to this evolution?
9. Do you think A4L has been successful in meeting its objectives as you've outlined? Why or why not?
10. Describe the management of the initiative. What worked well? What could have been improved?
11. Have A4L activities had any noticeable effects on GPE's global-level influence? Please describe.
 - a. If yes, which components of the project have had the most influence?
12. What about effects on country-level financing and programming? Please describe the effects.
 - a. If yes, which components of the project have had the most influence?

13. What barriers did you experience/were you aware of that hindered the effectiveness of A4L?
Think about these barriers in terms of:
 - a. At the global level?
 - b. Regional level?
 - c. National level?

14. Was there anything you think A4L could have done to be more effective or relevant?
 - a. At the global level?
 - b. Regional level?
 - c. National level?

15. What are GPE's future priorities in supporting countries' efforts to strengthen learning assessment systems?
 - a. Did A4L play a role in informing these priorities? Can you provide specific examples?

16. Do you have any other comments?

17. Can you suggest anybody else that I should talk to for the purposes of this evaluation?

Appendix 4. Generic interview protocol for implementing partners

Introduction

1. Please briefly introduce yourself: What are your current roles and responsibilities? For how long have you been in this position?
2. What was your role in relation to the A4L project? How long were you in this role?
3. Who were your primary contacts at GPE related to the A4L work? Did this change over time?
4. What resources did GPE contribute? Money? Staff time? Other resources (e.g., networks and knowledge transfer)? How did this compare to the relationships with other partners?
5. Was the partnership different than the way you normally operate? If so, how?
6. Did support from GPE allow your organization to undertake activities that you would not have otherwise been able to do? If so, what made this possible?
7. What are the objectives of the A4L project overall as you see them? How do those objectives relate to the objectives of your project?
8. Did the design or objectives of the project evolve over time? If so, why? Who determined the goals?

Effectiveness

9. Please describe your project's activities according to your initial project proposal/document.
 - a. Which ones have you accomplished?
 - b. How did the final deliverables deviate from what you originally proposed? Think about when they were delivered, the content, etc.
 - c. What was the reason for the variation?
10. According to the A4L documentation, there are three short-term objectives of the initiative. Has your work contributed to these? In what ways?
 - a. Providing clear recommendations for how to strengthen learning assessment systems
 - b. Increased understanding of efforts to measure whole child development, such as 21st century and socio-emotional skills
 - c. Increased measurement of socio-emotional skills
11. What about the following medium-term objectives? Please provide specific examples where possible.
 - a. Embedding good practices at the country level
 - b. Advancing tool development for measurement of WCD metrics
 - c. Exposing countries to models of holistic measures for learning

12. Do you think the project has been successful at setting the right conditions to strengthen learning assessment systems in the longer term? What about promoting the measurement of socio-emotional skills at the national and global levels?
13. What barriers did you experience/were you aware of that hindered the effectiveness of A4L? Think about these barriers in terms of:
 - a. At the global level?
 - b. Regional level?
 - c. National level?

Relevance

14. Do you feel A4L activities complement other efforts in the sector? Which A4L activities, in particular, and how do they complement other activities? Ex:
 - a. ANLAS toolkit
 - b. Support to regional networks & technical workshops
 - c. Landscape report on 21st century skills & convenings for the development of new assessment tools
 - d. General knowledge products (blogs, papers, events, etc...)
15. Did your organization collaborate with any new organizations while participating in this project? Please describe.
16. Was there anything that GPE and A4L could have done to be more relevant?
 - a. At the global level?
 - b. Regional level?
 - c. National level?

Efficiency

17. Describe the management of the initiative. What worked well? What could have been improved?
18. Based on your experience with A4L and this project, what would you do differently next time?
19. How was your work disseminated? Did GPE assist in any way?
20. How did the GPE Secretariat communicate the project's achievements? Was there anything else you wish would have been communicated?
21. What are your organization's future priorities in supporting countries' efforts to strengthen learning assessment systems?
 - a. Did A4L play a role in informing these priorities? Can you provide specific examples?
22. Do you have any other comments?
23. Can you suggest anybody else that I should talk to for the purposes of this evaluation?

Appendix 5. Generic interview protocol for country partners

Introduction

1. Please briefly introduce yourself. What are your current roles and responsibilities? How long have you been in this role?
2. What was your role in relation to the A4L project? How long were you in this role?

Relevance & Effectiveness

3. According to your Education Sector Plan/ESA/latest documentation, strengthening learning assessment systems is included in the following way (or not included):
 - a. Describe or share directly from the ESA/ESP
4. Is that still generally what your country is hoping to achieve?
5. In broad terms, can you describe how well your country has been doing in strengthening your learning assessment system?
 - a. What do you think are the key challenges for your country to implement reforms to strengthen learning systems?
 - b. Where have you observed most progress? Are there specific factors that can explain positive changes?
6. What A4L activities did your country participate in?
7. Do you think the A4L activities you mentioned have helped meet your country's needs in the area of strengthening learning assessment systems?
 - a. What resources were most useful? *Ex: Money? Staff time? Networks? Knowledge transfer?*
 - b. Are there any additional needs that you wish A4L would have addressed?
8. What other partners contribute to strengthening your learning assessment system?
 - a. How does GPE's support compare to contributions from other partners?
9. What are the objectives of the A4L project overall as you see them? How do those objectives relate to the objectives of your country as identified in your ESA/ESP?
10. Did the design or objectives of the project evolve over time? If so, why? Who determined the goals?
11. Did support from A4L allow your agency to undertake activities that you would not have otherwise been able to do? If so, what made this possible?
12. Did your agency collaborate with any new organizations while participating in this project? Please describe your relationships with the other organizations.
13. How would you rate your satisfaction with A4L's activities with respect to:

- a. Relevance to your country's context
- b. Usefulness of the activities
- c. Technical expertise
- d. Planning and implementation of activities

14. How do you think A4L activities could have been more relevant? More effective?

Sustainability

15. Can you share examples of ways that you or your country have used knowledge gained through A4L's activities?
16. Have any follow-up activities been undertaken since your participation with A4L? If yes, can you describe these activities?
17. Will strengthening learning assessment be part of your next ESP? If so, is that informed by your experience in A4L?
18. What are some challenges you face in taking this work forward?
19. Apart from the A4L initiative, has your country used GPE funding for other activities related to strengthening learning assessment systems?
 - a. Would you characterize these efforts as being aligned, overlapping, or unrelated?
20. Broadly, how satisfied are you with the support you receive from GPE and other partners in the A4L project?
21. Do you have any other comments?
22. Can you suggest anybody else that I should talk to for the purposes of this evaluation?

Appendix 6. Generic interview protocol for funders

Introduction

1. Please briefly introduce yourself: What are your current roles and responsibilities? For how long have you been in this position?
2. What was your role in relation to the A4L project? How long were you in this role?
3. Who were your primary contacts at GPE? Did this change over time?
4. Was the partnership with GPE different than the way you normally operate? If so, how?
5. What are the objectives of the A4L project overall as you see them? How do those objectives relate to the goals of your organization?
6. Did the design or objectives of the project evolve over time? If so, why? How was your organization involved in determining the goals?
 - a. For Porticus: Probe about decision to drop the Phase 3 work on country implementation. Was this always a potential piece and not a guarantee? How was it communicated to the grantees?

Effectiveness

7. According to the A4L documentation, there are three short-term objectives. Do you think they were achieved? Please provide specific examples where possible.
 - a. Providing clear recommendations for how to strengthen learning assessment systems
 - b. Increased understanding of efforts to measure whole child development, such as 21st century and socio-emotional skills
 - c. Increased measurement of socio-emotional skills
8. What about the following medium-term objectives? Please provide specific examples where possible.
 - d. Embedding good practices at the country level
 - e. Advancing tool development for measurement of WCD metrics
 - f. Exposing countries to models of holistic measures for learning
9. Do you think the project has been successful at setting the right conditions to strengthen learning assessment systems in the longer term? What about promoting the measurement of socio-emotional skills at the national and global levels?
10. What barriers did you experience/were you aware of that hindered the effectiveness of A4L? Think about these barriers in terms of:
 - a. At the global level?
 - b. Regional level?
 - c. National level?

Relevance

11. Do you feel A4L activities complement other efforts in the sector? Which A4L activities, in particular, and how do they complement other activities? Ex:
 - a. ANLAS toolkit
 - b. Support to regional networks & technical workshops
 - c. Landscape report on 21st century skills & convenings for the development of new assessment tools
 - d. General knowledge products (blogs, papers, events, etc...)
12. Was there anything that GPE and A4L could have done to be more relevant?
 - a. At the global level?
 - b. Regional level?
 - c. National level?

Efficiency

13. Describe the management of the initiative. What worked well? What could have been improved?
14. Based on your experience with A4L and this project, what would you do differently next time?
15. How did the GPE Secretariat communicate the project's achievements? Was there anything else you wish would have been communicated?
16. How was the project work disseminated? Do you think the dissemination was effective in reaching the target audiences?
17. What are your organization's future priorities in supporting countries' efforts to strengthen learning assessment systems?
 - a. Did A4L play a role in informing these priorities? Can you provide specific examples?
18. Do you have any other comments?
19. Can you suggest anybody else that I should talk to for the purposes of this evaluation?

Appendix 7. Generic interview protocol for external stakeholders

Introduction

1. Please briefly introduce yourself. What is your current role and responsibilities? How long have you been in this role?
2. In what ways have you been engaged with the A4L project? With GPE in general?

Relevance & Efficiency

3. Do you feel the following A4L activities complement other efforts in the sector? Why or why not?
 - a. A diagnostic tool for countries to conduct a comprehensive analysis of their learning assessment systems
 - b. Support to regional networks to organize technical workshops and coordinate research and knowledge-sharing activities
 - c. Landscape review on 21st century skills to promote the integration of 21st century skills in education systems in partner countries
4. Was there anything that GPE and A4L could have done to be more relevant?
 - a. At the global level?
 - b. Regional level?
 - c. National level?
5. Do you think GPE has been effective at communicating about the A4L project? Are you familiar with the following recent publications?
 - a. ANLAS toolkit
 - b. Landscape review on 21st century skills
 - c. KIX paper on strengthening learning assessments
6. What are your organization's future priorities in supporting countries' efforts to strengthen learning assessment systems?
 - a. Has GPE played a role in informing these priorities? Can you provide specific examples?
7. Do you have any other comments?
8. Can you suggest anybody else that I should talk to for the purposes of this evaluation?

Appendix 8. List of reviewed project documents

- “A4L Workplan March 2018”. (Excel spreadsheet)
- “A4L Workplan November 2018”. (Excel spreadsheet)
- “A4L Workplan March 2019”. (Excel spreadsheet)
- “A4L Grant implementation progress report for NEQMAP (July-Dec 2018)”
- “A4L Grant implementation progress report for NEQMAP (Jan-Jun 2019)”
- “A4L Grant implementation progress report for NEQMAP (July-Dec 2019)”
- “A4L Grant implementation progress report for regional assessment networks: NEQMAP Final”
- “A4L Grant implementation progress report for TALENT (Jun-Nov 2018)”
- “A4L Grant implementation progress report for TALENT (Dec 2018-May 2019)”
- “A4L Grant implementation progress report for TALENT (Jun-Dec 2019)”
- “Annex 3 – Assessment for learning concept note.”
- “Assessment for Learning (A4L) initiative: Progress Report for Dubai Cares (November 2018-December 2019)”
- “Assessment for Learning (A4L) initiative: Progress Report for Dubai Cares (July 2017-October 2018)”
- “Assessment for Learning (A4L): Porticus Annual report 2017”
- “Assessment for Learning (A4L): Porticus Annual report 2018”
- “Assessment for Learning (A4L): Porticus Annual report 2019”
- “BELDS Annual Update for Dubai Cares: Reporting Period: 2/17-2/18”
- “GPE Results Framework: Baselines, milestones, and targets”. No date.
- “Logic Model: Assessment for Learning (A4L)” (December 2017)
- “Logic Model: Assessment for Learning (A4L)” (January 2019)
- “Logic Model: Assessment for Learning (A4L)” (August 2019)
- “NEQMAP Programme Evaluation Report: Final”
- 2015. “Assessment for Learning (A4L): An international platform to support national learning assessment systems”
Discussion document prepared for Learning Metrics Task Force 2.0
- 2016. “GPE Grant proposal application: Porticus”
- 2016. “Grant Request Draft Proposal: Dubai Cares”

2017. "NEQMAP Narrative proposal"

2017. "NEQMAP Detailed work plan"

2018. "TALENT Proposal: Narrative"

2018. "TALENT Proposal: Workplan"

ACER. "Analysis of National Learning Assessment Systems (ANLAS): A GPE initiative to strengthen learning assessment systems". (Brief)

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Summary Report."

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Manual."

Anderson, Kate. 2019. "Strengthening learning assessment systems: A Knowledge and Innovation Exchange (KIX) discussion paper." GPE: Washington, DC.

GPE. 2020. "21st century skills: What potential role for the Global Partnership for Education. A landscape review."

---. "Results Report 2019"

---. "Results Report 2018"

---. 2018. "Update on the Assessment for Learning (A4L) Initiative." Meeting of the Finance and Risk Committee October 10-11, 2018. Washington, DC.

---. 2016. "Targeted financing arrangements for Assessment for Learning – Strengthening Learning Assessment Systems initiative." Meeting of the Board of Directors December 1-2, 2016. Siem Reap, Cambodia (BOD/2016/12 DOC07).

L'équipe National Mauritania. 2019. "Analyse des systèmes d'évaluation nationaux des apprentissages (ANLAS): Rapport Mauritanie".

-- "Analyse des systèmes d'évaluation nationaux d'apprentissage (ANLAS): Mauritanie Résumé

-- "Analyse des systèmes d'évaluation nationaux d'apprentissage (ANLAS): Mauritanie". (PowerPoint presentation)

National Educational Assessment and Examinations Agency (NEAEA). 2019. "Analysis of National Learning Assessment Systems (ANLAS): ANLAS Ethiopia Country Report".

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Ethiopia Executive Summary".

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Pilot report in Ethiopia". (PowerPoint presentation)

Vietnam Technical Team. "Analysis of National Learning Assessment Systems (ANLAS): Country report (Pilot Vietnam)".

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Vietnam Executive Summary".

---. 2019. "Analysis of National Learning Assessment Systems (ANLAS): Vietnam". (PowerPoint presentation)

Appendix 9. Web scan protocol

The web scan was conducted by searching for several combinations of terms:

“Assessment for Learning” + “GPE”

“Assessment for Learning” + “Global Partnership for Education”

“A4L” + “GPE”

“A4L” + “Global Partnership for Education”

"Partenariat mondial pour l'éducation" + "A4L"

"Partenariat mondial pour l'éducation" + "L'évaluation au service des apprentissages"

"PME" + "A4L"

"PME" + "L'évaluation au service des apprentissages"

The overwhelming majority of documents were in English, with several French translations. One article in Amharic was posted on the Ethiopian NEAEA website.

Appendix 10. Evaluation matrix

Evaluation Questions	Evaluation Questions	Data Sources and Collection Techniques
Effectiveness	To what extent has A4L achieved its intended outputs?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews
	To what extent has A4L achieved its intended short-term and medium-term outcomes?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews
	To what extent has A4L contributed towards its long-term goals?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews ● Sub-sector data analysis
	What were the most critical variances between planned and actual project activities? What contributed to these variances?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews
	What aspects of project design and organization most contributed to its success and/or shortcomings?	<ul style="list-style-type: none"> ● Key stakeholder interviews
Relevance & Sustainability	How relevant have A4L activities been at country, regional and global levels in terms of strengthening learning assessment systems?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews ● Web scan ● Sub-sector data analysis
	How do A4L activities complement and/or add value to GPE's support to learning assessment systems at the country level?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews
	How do A4L activities complement and/or add value to ongoing global efforts to strengthen learning assessment systems?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews ● Web scan
	To what extent has A4L catalyzed support and investment in the areas of learning assessment systems (at both global and country level)?	<ul style="list-style-type: none"> ● Key stakeholder interviews ● Web scan ● Sub-sector data analysis
	To what extent has A4L allowed for collaboration with existing and new actors at global, regional, and country levels in regard to strengthening learning assessment systems?	<ul style="list-style-type: none"> ● Document review ● Key stakeholder interviews ● Web scan
Efficiency	To what extent has the management and coordination of A4L by the GPE Secretariat been efficient?	<ul style="list-style-type: none"> ● Key stakeholder interviews ● Questionnaire
	To what extent has the Secretariat been efficient in terms of communication and dissemination around A4L activities and achievements?	<ul style="list-style-type: none"> ● Key stakeholder interviews ● Questionnaire ● Web scan
	To what extent has the collaboration with foundations added value for GPE in supporting work in specific thematic areas?	<ul style="list-style-type: none"> ● Key stakeholder interviews ● Questionnaire

Appendix 11. Subsample of recommendations to strengthen the context and coherence of the assessment systems in ANLAS pilot countries

	Vietnam	Ethiopia	Mauritania
<u>Focus area 1: Context of the assessment system</u>			
<p>1. Legislation or Policy</p> <p><i>The assessment system is guided by legislation or policy.</i></p>	<ul style="list-style-type: none"> • Need to amend policies, circular of instructing how to conduct classroom assessment towards developing student’s capacity from primary, lower secondary and upper secondary level in accordance with curriculum, new textbook. • Improve the system of books that requires teachers to monitor, evaluate, and comment on students to reduce administrative works for teachers. 	<ul style="list-style-type: none"> ▪ Federal Education law should be developed to strengthen, resolve disputes and protect the assessment system ▪ 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. 	<ul style="list-style-type: none"> ▪ Need official documents to guide various programs of assessment ▪ Put in place an orientation law for the education system where evaluation plays an important role ▪ Integrate specific guidelines in curricula for assessing academic achievement, including test models ▪ Explicitly include the assessment of 21st century skills in the assessment framework
<p>2. Leadership</p> <p><i>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.</i></p>	(Achieved)	<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 merit base system of hiring education officials so that they can give technical and professional support in the area. ▪ The Ministry of Education should establish strong monitoring and support system for assessment Programs on funding, 	<ul style="list-style-type: none"> ▪ Integrate the objectives of the evaluations into the matrix of sectoral action plans ▪ Put in place adequate communication tools to inform stakeholders in the education system ▪ Strengthen the capacities of decision-makers, teacher trainers, school leaders and teachers in the use and implementation of the results of prior learning assessments

		disseminating, communicating & utilization of assessment results.	<ul style="list-style-type: none"> ▪ Ensure sharing of results from large-scale assessments
<p>3. Funding</p> <p><i>The government provides sufficient and stable funding for the assessment system.</i></p>	<ul style="list-style-type: none"> • Need to develop strategic long-term plans, in next 20 years, or; • Add budget of the regular assessment activities into the regular government budget. • Supplement the budget for education assessment experts, especially sampling and data analysis experts. 	<ul style="list-style-type: none"> ▪ The government should allocate appropriate budget for effective assessment practices, capacity building & technology infrastructure for the assessment system 	<ul style="list-style-type: none"> ▪ Establish a stable source of funding, allowing for long-term organization of large-scale national evaluations ▪ Strengthen the skills of those in charge of central, regional and local structures in the preparation of funding requests
<p>4. Institutional arrangements</p> <p><i>The government has institutional arrangements in place for designing, implementing, analyzing and using data from various learning assessments.</i></p>	<ul style="list-style-type: none"> • Need to invest for the long-term assessment program designation, be inherited and developed information to use any resources. 	(Achieved)	<ul style="list-style-type: none"> ▪ Strengthen the skills of the staff of various structures in terms of assessment ▪ Create a central committee responsible for monitoring, sharing, use and communication of results from assessments
<p>5. Capacity to use assessment data</p> <p><i>The government has the capacity to use data from learning assessments for evidence-based decision making in education policy and practice.</i></p>	<ul style="list-style-type: none"> • MOET had a directive dispatch, guidelines for Units to study data and recommendations from the report to use them properly, making it effective in improving the classroom assessment development policy, test examination and large-scale assessment. 	<ul style="list-style-type: none"> • Continuous training on utilization of software applications must be given for, experts in NEAEA. • Intensive awareness and capacity building programs must be given for education officials and/or experts of curriculum department on the importance and utilizing the assessment results for formative evaluation. • The Ministry of education need to develop a system that enhance the capacity of EMIS in a way, information is efficient, effective and accessible to all key stakeholders 	<ul style="list-style-type: none"> • Reinforce system management capabilities (management and planning) • Conduct a diagnostic study on the need for training • Implement a training program that meets those needs • Include training on the use of data at the budget level

<p>6. Public engagement</p> <p><i>The public understands, is engaged in, and supports the assessment system.</i></p>	<ul style="list-style-type: none"> • Strengthening the propaganda to make the public better understand about the quality of examinations, and test exams • Make the public positively participate in, and not evaluate wrongly or incorrectly about the education quality and examinations, assessments. 	<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 printing and non-printing medias 	<ul style="list-style-type: none"> ▪ Ensure the sharing of results from large-scale assessments through: <ul style="list-style-type: none"> • Feedback workshops • Discussion in media • Publication on MENFP site ▪ Develop an assessment culture through the participation of media in sharing results and the organization of awareness days
<p><u>Focus area 2: Coherence of the assessment system</u></p>			
<p>1. Structure of the education system</p> <p><i>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.</i></p>	<p>(Achieved)</p>	<p>(Achieved)</p>	<ul style="list-style-type: none"> • Regularly organize large-scale national assessments to encourage decision-makers to take relevant and appropriate decisions • Participate in regional assessments (ex: PASEC) • Assess the quality of initial training • Integrate 21st century skills in primary and secondary assessments • Provide sufficiently aggregated performance and contextual data • Ensure comparability between assessments
<p>2. Education policy priorities</p> <p><i>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.</i></p>	<p>(Achieved)</p>	<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 a constructed response test items in the future. • EMIS should device a mechanism to collect and organize classroom assessment data to align with NLSAs to inform the system for evidence-based decision. 	<ul style="list-style-type: none"> • Take assessment data into account in developing strategies, planning and steering the system • Design and implement, on a regular basis, large-scale assessments to identify determinants of quality learning • Provide data in useable formats to a commission in charge of implementing results of the assessments at the institutional level

<p>3. Learning standards and curriculum</p> <p><i>The learning domains/subjects in the assessment system are aligned with official learning standards and/or curriculum.</i></p>	<ul style="list-style-type: none"> • Currently general education program needs to update knowledge, modern assessment skills with the aim to supporting teachers for enhancing classroom assessment capacity. • MOET needs a policy for training how to design test, use method of learners' capacity assessment for teachers in basic way so that they can develop the test in standard format, be able to measure knowledge, skills, practical capacity to reach the purpose which is needed to assess. • New general education program needs to focus on developing the outcome standard system so that teachers can refer to it and design suitable methods, classroom assessment techniques. 	<ul style="list-style-type: none"> • Intensive capacity building should be provided for teachers and school leaders on classroom assessment techniques and its alignment with curriculum standards (competencies) • Instructional materials should be developed in such way that they incorporate more guidance for continuous classroom assessment • NEAEA and Curriculum Development and Implementation Department should device a way to assess 21st century skills. 	<ul style="list-style-type: none"> • Strengthen the articulation between teaching programs and evaluation programs through the strict use of orientation documents • Create an evaluation framework for all levels of the system
<p>4. School education workforce development</p> <p><i>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</i></p>	<p>MOET needs to:</p> <ul style="list-style-type: none"> • Develop the guideline text and documents of testing method, classroom assessment in details, suitable with current context so that education managers (from DOET, Division of Education and Training, Principals) and teachers can easily use it; • Develop more guidelines documents of classroom assessment techniques in details, focusing on modern 	<ul style="list-style-type: none"> • Teacher training institutions and MOE should prepare classroom assessments training materials that aligned with the official documents. • Capacity building materials should be 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 	<ul style="list-style-type: none"> • Develop training programs for developing skills of teaching staff and school leaders in assessment that are in line with the assessment framework, once developed • Integrate modules for assessments in the continuous and initial training programs of the ENIS and ENS • Make systematic the organization of classroom assessments • Improve monitoring and supervision of teachers

<p><i>guide the assessment system.</i></p>	<p>education method guidelines so that teachers can know how to conduct it;</p> <ul style="list-style-type: none"> • Frequently organize training workshop and develop the capacity for school staffs: (1) rationale (for example, clarifying concepts, terminology, differences between capacity assessment and assessment according to knowledge and skills standards); (2) Technical methods for classroom assessment; (3) Techniques for developing student’s capacity assessment tests include: developing test matrix, Multiple choice and open ended question writing techniques, essay and objective test development technique. • Develop more guidelines materials, share more experiences online for teachers to study and develop their self-competency about above issues. 	<p>school leader on assessment programs.</p> <ul style="list-style-type: none"> • MoE should device mechanism for monitoring and evaluating the impact of training on assessment programs. 	<ul style="list-style-type: none"> •
<p>5. Use of data</p> <p><i>Assessment data is used by various key stakeholders for evidence-based decision making in education policy and practice.</i></p>	<ul style="list-style-type: none"> • Need to invest more budget to conduct the international assessment result analysis, develop national report; • Relevant Units need to use data and recommendations from data to: (1) develop better education policies; (2), reduce teacher’s stress (Ex: reduce some administration work so that teachers can focus on innovating teaching methods and checking the assessment); (3) improve the living quality standard for teachers and education 	<ul style="list-style-type: none"> • NEAEA should improve accessibility of data that address the needs of various stakeholders timely. • MoE should establish policies and accompanying rules and regulation for accountability; • MoE should provide training/awareness creation on using data for various stakeholders. 	<ul style="list-style-type: none"> • Operationalize the EMIS • Strengthen the skills of MENFP structures to use data from assessments • Encourage stakeholders to use assessment results

	managers (Ex: promote teacher's salary, have a policy supporting for teachers who do not have a house can stay in the public house; rotate teachers within areas so that teachers in remote areas will not be disadvantaged).		
--	---	--	--

Appendix 12. Web scan results

Title of media	Year	Publishing organization	Description	Type
Assessment for Learning: An international platform to support national learning assessment systems	2015	Brookings	Blog introducing A4L concept note	Blog
Learning Data for Better Policy	2016	Center for Global Development	Brief description of what A4L proposed to do	Mention in report
Global Education Monitoring Report	2016	UNESCO GEMR	Brief mention of the A4L proposal	Mention in report
Welcoming two new initiatives to accelerate quality teaching and learning in Africa	2016	Brookings	Blog mentions A4L	Blog
The GPE Board meets in Oslo	2016	ASPBAE	Mentions approval of A4L at Board meeting	Meeting summary
Rethinking the Financing and Architecture of Global Education	2016	Education Commission	Brief mention of A4L in the context of LMTF and GAML	Paper
Aligning key initiatives in monitoring learning	2016	NORRAG	Discussion of GAML, A4L, Ed Commission and CGD's paper	Blog
Five ways to empower teachers and drive learning progress	2017	Education International	Brief mention of A4L	Mention in blog
DFID Better Education Statistics and global Action to improve learning (BESTA) Business Case	2017	DFID	Describes A4L in the context of other initiatives to support learning assessment	Donor documents
IICBA Participated in the Regional Workshop on National Learning Assessment Systems in Sub-Saharan Africa	2017	UNESCO IICBA	Describes December 2017 TALENT workshop in Dakar	Blog
Regional Workshop on National Learning Assessment Systems in Sub-Saharan Africa	2017	UN Senegal	Describes December 2017 TALENT	Blog

			workshop in Dakar	
NEQMAP builds capacity for quality learning assessment data across the Asia-Pacific	2018	Education Innovations	Describes March 2018 NEQMAP workshop	Blog
Five years on, NEQMAP redoubles efforts to improve quality learning assessment data in Asia-Pacific	2018	UNESCO Bangkok	Description of the March 2018 workshop	Blog
World Development Report	2018	World Bank	Mentioned as an initiative to get better comparable data on learning	Report mention
Mobiliser les parties prenantes africaines pour améliorer les résultats d'apprentissage	2019	CONFEMEN	Description of A4L and TALENT (GPE cross-post)	Blog
The Promise of Large-Scale Learning Assessments	2019	UNESCO	Brief description of ANLAS, A4L, and KIX	Mention in report
Workshop on education assessment system conducted	2019	Ethiopian NEAEA	Description of ANLAS workshop in Ethiopia and ANLAS report	Web article
Champions for Learning: The Legacy of the Learning Metrics Task Force	2019	Brookings	Article on A4L by Jean-Marc Bernard	Report chapter
Analysis of National Learning Assessment Systems (ANLAS) Manual	2019	USAID	ANLAS Manual posted on USAID's edu-links.org site	Document posting
Assessment for Learning – bringing the attention back to learning	2019	UNESCO Bangkok	Describes a June 2019 workshop (does not mention A4L)	Blog
Educational Planning in the age of Digitisation	2020	Journal of Educational Planning	Mentioned as an initiative to generate better data	Research article