OPTIMIZING ASSESSMENT FOR ALL

Classroom-based assessments of 21st century skills in the Democratic Republic of Congo, The Gambia, and Zambia

BROOKINGS

Authors

Helyn Kim was a fellow in the Global Economy and Development Program at the Brookings Institution

Esther Care is a senior fellow in the Global Economy and Development Program at the Brookings Institution

Optimizing Assessment for All (OAA) is a project of the Brookings Institution. The aim of OAA is to support countries to improve the assessment, teaching, and learning of 21st century skills through increasing assessment literacy among regional and national education stakeholders, focusing on the constructive use of assessment in education, and developing new methods for assessing 21st century skills.

Acknowledgements

The authors and the OAA Africa National Teams appreciate the support of UNESCO's Teaching and Learning Educators' Network for Transformation (TALENT) and our thanks are extended in particular to Davide Ruscelli. The authors gratefully acknowledge the Democratic Republic of Congo National Team; The Gambia National Team; and the Zambia National Team for their substantial contributions to their respective country sections of the report. Additionally, the authors thank Joesal Jan A. Marabe of the Assessment Curriculum and Technology Research Centre, University of the Philippines Diliman, for technical contributions, and Aynur Gul Sahin for her creative and editorial support.

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars. Brookings gratefully acknowledges the support provided by Porticus.

Brookings recognizes that the value it provides is in its commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.

Cover photo credit: St. Peter's Lower Basic School, Banjul, The Gambia

TABLE OF CONTENTS

EXECUTIVE STATEMENT

The Optimizing Assessment for All (OAA) project at Brookings explores approaches to educational assessment, specifically through developing assessments of 21st century skills (21CS).

Twenty-first century skills are now firmly ensconced as new learning goals in education systems worldwide, but their implementation in teaching and assessment practices is lagging behind.

We have taken decades to understand how to teach mainstream education subjects like mathematics, history, science, and language. But with these new learning goals, which prioritize how to go about getting answers, rather than just providing a correct response, we are facing new challenges in both assessment and teaching. If we can identify useful approaches to assessment of 21CS in the classroom, then both the assessment tools themselves as well as how students engage with them can provide insights for teaching the skills.

The overall goal of the OAA project is to strengthen systems' capacity to integrate 21CS into their teaching and learning using assessment as the entry point to changing education practices in alignment with changing learning goals. Specifically, OAA is designed to shift mind and practices around the use of assessment; shift perceptions on how assessment relates to the broader education structure—that assessment is not something separate but a critical part of the learning and teaching process; and develop new methods for assessing 21CS in the classroom.

OAA, in collaboration with participating countries from Asia and Africa, has helped identify 21CS valued by these countries, hypothesized what these skills might look like in classroom assessment tasks, and developed these tasks with teachers to ensure that they are usable and valuable in the classrooms. Notably, OAA has worked with established approaches to assessment with which teachers are familiar, and adjusted them to reflect new learning goals. Of course, the work goes far beyond assessment to implications for how we think about education and what we value in the classroom. What we value are the thinking and social processes that individuals use to explore and understand their environment.

More comprehensive information about the complete OAA approach can be found in the "Optimizing Assessment for All: Framework" report, while in this report we focus on the collaborative activities undertaken in Africa by the Democratic Republic of Congo, The Gambia, and Zambia to create 21CS assessment tasks. The mechanics of the activities are described in detail in order to illustrate the methods used in the project and by the countries. For additional examples and guides for task creation, see our forthcoming fourth report, and for discussion about scaling and implementing the OAA approach, see the forthcoming fifth report in the series.

INTRODUCTION

One of the main goals of OAA was for participating countries to develop classroom assessment tasks that can measure 21CS. The project adopted a collaborative approach to develop capacity in assessment design. The project was structured so that national teams had the opportunity to develop assessment tasks together at the regional level, as well as individually at the national level. The objective was to ensure that the national teams were confident in the usability of the developed tasks and in their ability to continue to develop tasks for their particular conditions, needs, and curriculum. The development process was undertaken through a series of workshops, usually convened in one of the participating countries so that the national teams had the opportunity to understand the conditions under which each was working. Between workshops, in-country development work took place, both within the national teams, and with teachers from participating schools in each country. The collaborative approach and task development processes are described in this report—covering the workshops and in-country activities—and culminating in a pilot of the assessments across the countries.

As education systems increasingly emphasize the need for students to apply their learning, focus on 21CS has intensified. With relatively little assessment of 21CS in education, OAA has taken the stance that assessment in the classroom will provide the support needed for teaching in the classroom. Development of classroom assessments will provide grounded examples of what student learning practices in demonstration of 21CS look like, thus informing the development of tools for use at regional and national levels.

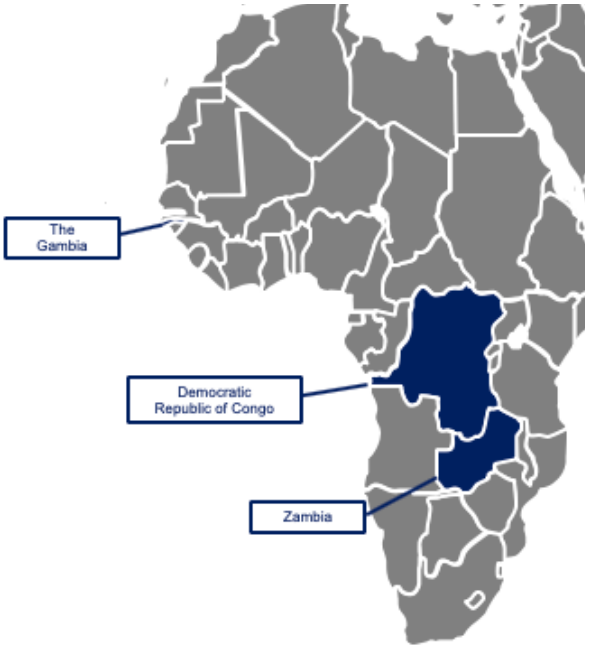
THE OAA APPROACH FOR AFRICA

Similar to the work with three countries in Asia, the Brookings Institution worked intensively with three countries in Africa: two Anglophone countries (The Gambia and Zambia) and one Francophone country (Democratic Republic of Congo [DRC]). The activities were undertaken over a 15-month period through the OAA initiative to improve the assessment, teaching, and learning of 21CS, with support from the Teaching and Learning Educators' Network for Transformation (TALENT) at UNESCO Dakar. Although the overall objectives of the OAA initiative in Africa are identical to those of Asia, a slightly different approach was taken. The decision to modify the approach was both strategic and practical. The Africa approach drew on lessons learned both from the Asia experience (since OAA began in Africa about six months after Asia), as well as from a "mini-study" involving nine African countries that identified current national- and classroom-based tests or test items that would capture 21CS.

Lead National Team members from The Gambia, DRC, and Senegal with OAA Brookings scholars convening in Dakar, Senegal to discuss the project

As countries increasingly adopt learning goals that reflect more holistic perspectives on education, they engage in curriculum review and consideration of what other aspects of education delivery need modification. The nature of learning goals has implications for selecting different forms of assessment, in that some learning goals, such as memorization of facts, for example, can easily be captured through constrained forms of assessment (e.g., multiple choice, short answer questions,or “fill in the gaps” responses). Other learning goals associated with knowledge or skills that are not easily demonstrated through the written word might more coherently be captured through less constrained assessment types. These include tasks that require multiple steps, which might be taken in different sequences or that might require working with different media. In OAA, the focus is on how we might capture student proficiencies in 21CS in the classroom. Of course, to ensure alignment throughout the system, the forms of classroom assessment also need to be consistent with large-scale assessment. The latter typically relies on constrained forms of assessment since these are relatively easy to standardize and efficient to administer. OAA is concerned with optimizing the links between the efficiencies of large-scale assessment forms and the potential richness of classroom-based assessment. The idea of “vertical coherence” (Herman, 2010), where assessment at all levels of the education system is aligned with learning goals, underpins the OAA model.

The approach for Africa: 1) capitalized on the outcomes of the OAA Africa and Asia mini-studies, specifically the potential for adapting some existing tools to assess 21CS; 2) incorporated learnings from the OAA three-country work in Asia; and 3) ensured that the activities and outputs in the two regions were complementary rather than replicated.



OAA-participating countries

This approach enabled the exploration of multiple approaches to designing, developing, and using classroom-based assessments of 21CS.

The National Technical Teams in OAA Asia were most strongly influenced by assessment personnel, while the teams for Africa included stronger teacher and pedagogical expertise representation. Members of each of the Africa teams represented assessment, curriculum, and pedagogical expertise, and all teams included a minimum of one teacher. Therefore, the key differences between OAA Africa and OAA Asia approaches are: greater teacher involvement at the team level in Africa (as distinct from the teacher involvement within participating schools in both regions); and tool adaptation in Africa rather than development from ground up as in Asia. The suite of tools and development of an assessment guide (described in the forthcoming fourth report in the series) from the two regions are complementary. Common to both regions are general principles and processes of item and test development, scoring, and targeting of both constructs and student abilities.

In general, the Africa approach is characterized by the following:

- Participating countries explored as a collaborative group what was already familiar in order to proceed to the next steps in learning. Practically, this meant that some existing assessment tools collected through the mini-study were adapted and extended to target 21CS, as well as developing new tools.
- It takes a range of assessment techniques to measure 21CS: Available assessment examples from the mini-study included short answer items, essays, and set tasks. Using these as starting points, templates were developed, which provided frameworks for developing new items and scoring approaches for 21CS.
- It is collaborative but reflects each country's unique curriculum: The national teams worked together on skill descriptions, assessment item types, and item review but unlike in Asia, the teams were not required to develop items that target identical curricular topics across countries.

THE THREE FOCUS COUNTRIES

The DRC, The Gambia, and Zambia worked collaboratively to design, develop, and use 21CS assessments. These three countries were identified during the Africa mini-study based on their commitment to integrating 21CS into their education systems. For information regarding the National Teams, see Appendix A.

DEMOCRATIC REPUBLIC OF CONGO (DRC)

In its most recent country-wide reform agenda, the DRC identified education as one of the most important drivers in addressing the country's resource management gap. The government's vision for the education sector is "the construction of an inclusive and quality education system that contributes effectively to national development, the promotion of peace and active democratic citizenship" by equipping Congolese students with 21CS, such as creativity, critical thinking, problem solving, and the ability to take initiative.

OAA Africa National Technical Teams from the Democratic Republic of Congo, The Gambia and Zambia at the first workshop in Banjul, The Gambia.

The education system aims at the harmonious formation of the Congolese man, responsible citizen, useful to himself and to the society able to promote the development of the country and the national culture. DRC is focused on training productive, creative, cultured, conscientious, free and responsible citizens, open to social, cultural and aesthetic, spiritual and republican values (Education Charter, CNS, 1992).

THE GAMBIA

The Gambia is committed to developing its human resource base, with priority given to free basic education for all through "accessible, equitable and inclusive quality education for sustainable development." As highlighted in Education Policy 2016-2030, the guiding principles of the education sector are:

- Non-discriminatory and all-inclusive provision of education, in particular with respect to gender equity and targeting of the poor and the disadvantaged groups;
- Respect for the rights of the individual, cultural diversity, indigenous languages, and knowledge;
- Promotion of ethical norms and values and a culture of peace; and
- Development of science and technology competencies for the desired quantum leap.

The education sector aims to ensure that teaching and learning focus on developing the physical and mental skills which will contribute to nation building—economically, socially, and culturally—and develop creativity and the analytical mind. In keeping with the country's commitment to the Sustainable Development Goals, the education sector is dedicated to promoting life skills education to help learners acquire not only knowledge and skills but also adaptive and positive behaviors in a changing social and economic environment.

ZAMBIA

The goal of the education system in Zambia is to "nurture the holistic development of all individuals and to promote the social and economic welfare of society." Zambia's vision focuses on providing "quality and relevant lifelong education and skills training for all," that is accessible, inclusive, and relevant to individual, national, and global value systems. Zambia is committed to providing an education that will meet the needs of Zambia and its people. The aims of the Zambian curriculum are to train:

- Self-motivated, life-long learners;
- Confident and productive individuals; and
- Holistic, independent learners with the values, skills, and knowledge to enable success in school and in life.

Learners acquire a set of values, which encourage them to:

- Strive for personal excellence;
- Build positive relationships with others;
- Become good citizens; and
- Celebrate their faith and respect the diversity of beliefs of others.

In addition, the Education Curriculum Framework 2013 identifies key competencies for learners at each level of school that go beyond literacy and numeracy skills to include critical, analytic, strategic, and creative thinking; problem solving; self management; relationship skills; civic competence; participation; and teamwork.

OAA PROCESS FOR TASK DEVELOPMENT

The OAA Africa approach values the processes and learning that took place as team members engaged in the series of task development activities, rather than the set of tools developed during that process. The latter are relevant only insofar as they are a testament to the capacity of the National Teams and their understanding of test development aligned with local use. Over the course of 15 months, the countries engaged in four workshops that built upon each other, as well as in-country activities between workshops, which were designed to apply the concepts learned in the workshops to schools and classrooms, and with teachers and students. For example, National Team members engaged with schools and policymakers to generate buy-in and build practitioner understanding around 21CS. They visited schools to understand the classroom contexts. They also conducted teacher training sessions to improve teachers' understanding of 21CS; increase their ability to identify the skills demonstrated by students in the classroom; develop new 21CS items that capture the skills that they can use in their classrooms; and troubleshoot scoring issues associated with the assessment of student behaviors.

The purpose of OAA is therefore not about generating a tool that can be used widely but rather to provide a prototype or a model approach that countries can use to integrate 21CS into their teaching and learning. Figure 1 shows the workshop series and regional convenings (in large blue circles), as well as the in-country activities.

Figure 1. Series of activities for the OAA task development process

The steps of the process through the workshops and in-country schedules include:

- Understand the nature of skills and the implications for assessment development;
- Select, define, and deconstruct the target skills;
- Re-vision existing assessment items to target 21CS;
- Conduct “think aloud” sessions to verify the skills and their components being prompted by the items;
- Generate new items that target skills at different levels of difficulty;
- Panel items;
- Pilot items in schools;
- Analyze student responses; and
- Review assessment tasks.

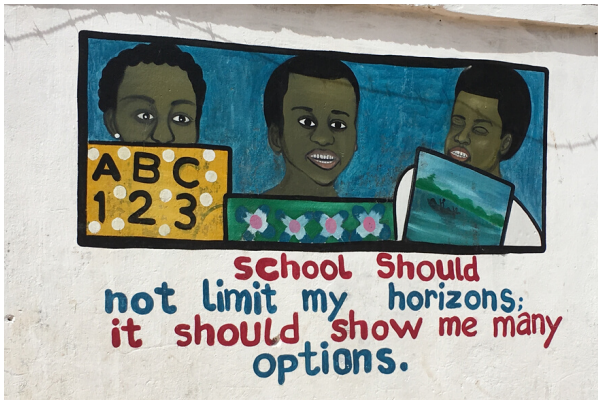
Enthusiasm from Gambian school children

1 UNDERSTANDING THE NATURE OF SKILLS

Education systems around the world acknowledge the importance of 21CS, such as problem solving, collaboration, and critical thinking. However, if the goal is to teach and learn these skills, merely identifying which skills are considered important or even defining the skills is not enough. More in-depth understanding of the nature of the skills is necessary (Care, Kim, Vista, & Anderson, 2018).

The defining characteristics of a 21CS adopted in OAA is that an individual or group of individuals can bring that competency to bear in and across new situations. Skills are different from knowledge in that, although knowledge can be acquired, that in and of itself is not sufficient to put that knowledge into practice. Skills enable one to apply knowledge to different situations and transfer what has been learned in one context to another.

This means that actual skills recognition is important. Traditional methods of information dissemination are not enough to facilitate the application and transfer of skills to new or different situations. Authentic learning tasks (i.e., tasks that are similar to the ones that students will face in the real world) can provide opportunities to apply the skills in different ways.



School walls in The Gambia not limiting horizons

The Gambia: Approach to understanding skills

For the Ministry of Basic and Secondary Education (MoBSE) in the Gambia, providing quality education for all is a core mandate. Therefore, promoting the use of 21CS in classrooms to enhance effective teaching, promote independent learning, and reshape the existing assessment system is a major goal. Although 21CS are highly valued, a closer examination of the curriculum materials revealed that 21CS, specifically problem solving and collaboration, "accidentally" appeared on only a few occasions. In other words, there was no deliberate attempt to integrate the skills within teaching and learning.
Understanding the nature of skills was one of the most critical components of this OAA project because it is the foundation for not only teaching and assessing 21CS but also integrating it into the education system. The first in-country activity that the National Team members conducted was a day-long training session with representatives of four pilot schools to discuss the OAA project, expose them to 21CS, and discuss how the skills can be used to enhance effective teaching and learning in the classroom and beyond. The schools were Mansa Colley Bojang Lower Basic School (LBS), St. Peter's LBS, Abuko LBS, and St. Mary's LBS. Three teachers from each of the four pilot schools (in total, three female and nine male) attended.

The activity was designed to raise teacher awareness as well as develop understanding about the use of the 21CS, and consider how to stimulate learners' thinking. The session began highlighting. The Gambia's current education system and the curriculum materials, which reflected 21CS to a certain degree. The results of the OAA Africa mini study, in which The Gambia participated, illustrated that the skills might have been used in classrooms but only unconsciously rather than explicitly. To prime the teachers, they were asked to list the 21CS that they were aware of and state how they assess those skills in the classroom. Although skills such as problem solving, critical thinking, and effective communication were identified, the teachers were uncertain about how these skills could be assessed in their classrooms. One of the major discussions was around understanding what skills mean and how they are different from knowledge.

Despite recognition of the importance of 21CS, teachers felt that there were real challenges to teaching these skills in the classroom. These included:

- Inadequate or poor planning for collaborative activities that can hinder the progress of a lesson;
- Lack of space that can make it difficult to carry out activities;
- Lack of time for certain activities to be completed in a class period that is usually 30-35 minutes; and
- Lack of availability of materials required for some activities at schools.

THE TARGET SKILLS: PROBLEM SOLVING AND COLLABORATION

Two target skills, problem solving and collaboration, were selected for the purposes of the OAA project to serve as concrete examples to illustrate the task development process. Both skills were explicitly mentioned in the three countries' education policy documentation.

Once the two skills were decided upon, the next step in the process was to define and deconstruct the skills into their components and subcomponents. Taking into account research on the structures of problem solving and collaboration, as well as existing frameworks that identify both the processes and components of the skills, the National Teams worked together through numerous iterations to develop a framework acceptable and relevant for all three countries. For problem solving, three components were identified: information gathering, planning a solution, and managing information. Within these components, sub components were identified, as well. For example, information gathering includes both asking questions related to the problem and organizing information.

Lazarous Kalirani Kays and Beatrice Mbewe sharing their knowledge at Zambia's Stakeholders' Orientation Workshop in October 2019

These subcomponents were further deconstructed to identify more specific processes, such as classifying, analyzing, and describing. For collaboration, four components were identified: participation, communication, negotiation, and decisionmaking. Similar to problem solving, subcomponents were also identified. Tables 1 and 2 show the frameworks for problem solving and collaboration, respectively. These two frameworks set the foundation for the design, development, and pilot of classroom-based assessments of 21CS.

Table 1. Framework for problem solving

Skills components	Subcomponents	Processes
Information gathering (IG)	Ask questions related to the problem (Aq) Organize information (Oi)	Classify (Cla) Analyze (verify, discriminate, compare (Ana) Describe (Des)
Planning a solution (PS)	Generate ideas, options, hypotheses (Ge)	Hypothesize (Hyp) Consider and compare options (ConCom)
	Develop plan (Dp)	Discriminate (Dis) Identify relationships (Rel) Predict (Pre)
Managing information (MI)	Follow a plan (Fp)	
	Compare outcomes with plan (Cf)	Compare evidence with predict (Com) Check logical flow (Clf)
	Justify the process (Ju)	Explain (Exp)
	Synthesize (Sy)	Summarize (Sum)

Table 2. Framework for collaboration

Skills components	Subcomponents
Participation (P)	Take responsibility (Tr) Share (Sh) Take turns (Tt) Engagement (En)
Communication (C)	Receptive (Re) Expressive (Ex)
Negotiation (N)	Compromise (Co) Perspective taking (Pt)
Decision making (D)	Analysis (An) Evaluation (Ev) Plan (Pl)

"Skills are teachable, and to be acquainted with 21CS, one needs to understand the skills and subskills involved in each of the processes... Problem solving is not only limited to mathematics but can also be used in other subjects. A problem arises when one is faced with a situation that has no solution and requires some rigorous processes —ranging from information gathering, analysis, development of a solution, evaluation of options, and decision making. A problem is simply a complex situation that requires a solution."

Mr. Momodou Jeng, Director, Science and Technology Education and of In-service Training Unit, Ministry of Basic & Secondary Education (MoBSE)

The Gambia: Understanding problem solving and collaboration

Describing problem solving as a process, rather than a type of task, helped teachers to better understand the skill itself. Discussions around collaboration also took place to help teachers understand the concept and distinguish it from skills such as cooperation and teamwork.

"While cooperation means working with people and sharing ideas and resources, collaboration means working with people toward the attainment of a shared goal. Collaboration involves working together as a group, assigning roles, and supporting one another toward the successful accomplishment of the task. This means everyone takes responsibility and contributes positively toward the success of the larger group by effectively communicating, actively listening, taking turns, negotiating, and compromising," Mr. Ousmane Senghor, Head of Assessment Unit, MoBSE

RE-VISIONING EXISTING ITEMS TO TARGET 21CS

Rather than starting from scratch, OAA Africa relied on existing items and tasks from national and classroom levels that had been identified through the earlier Africa mini-study (UNESCO, 2020) as having the potential to target 21CS. These existing items and tasks were used as starting points but re-visioned or modified to more clearly target the skills of interest.

The goal was to develop items and tasks that:

- Can capture the skills and their components;
- Can capture these skills at increasing levels of proficiency;
- Are recognizable to teachers as capturing the skills; and
- Have structural features that can be replicated by teachers.

To illustrate this process, countries identified existing tasks with the potential to capture 21CS. For example:

Imagine you are outside your house playing with your friends. Your parent comes home and tells you to go and clean up your room and arrange your toys. You don't want to stop playing. You know your room is messy. What would you do?

The first issue to consider was whether this item could capture problem solving. Using the problem solving framework (Table 1), the teams considered how the item could be modified or expanded, so that the main skills' components and subcomponents could be more explicitly captured. Several questions were added with the identification of the components and subcomponents that were being targeted:

1. What is the problem you are facing?
[**Information gathering** - read the information, gather the relevant pieces, and organize the information.]
2. What additional information do you need before answering the question?
[**Information gathering** - ask questions related to the problem and consider what information might be missing.]
3. What would you do in this case? Name 3 possible solutions.
[**Planning a solution** - generate different ideas and options for how to respond.]
4. Of all possible solutions, what is the best and why?
[**Planning a solution** - consider and compare the different possible solutions, in order to identify the best solution and explain why.]
5. How will you do this? List the steps you would take to implement your solution.
[**Planning a solution** - after identifying the best solution, develop a step by step plan for how the solution will be implemented.]
6. If this solution does not work, what else can you do?
[**Managing information** / **Planning a solution** - compare the solution with the plan, check the logical flow of their plan, and as necessary draft a new solution.]

Another approach to re-visioning tasks is shown in Figure 2.

Figure 2. Example item that was expanded to target problem solving

Q1. Which combination of pots can be used to measure 550 ml?

A. 400 and 500
B. 150 and 400
C. 750 and 1000
D. 150 alone

Q2. If you want to distribute 2200 ml of water evenly to four friends, explain how you would do this.

The original (Q1) item is retained as a straightforward numeracy item, with a second (Q2) item added to target problem solving, and drawing on the components and subcomponents of Information Gathering > Organizing Information, and Planning a Solution > Generating Ideas, Options, and Hypotheses.

Countries initially engaged in this process of re-visioning existing items, and then generalized the process. The generalization enabled the development of several models, or templates, that could be used to structure new assessment tasks for problem solving and collaboration. The templates ranged from selected response items (e.g., matching and multiple choice), to constructed response items (e.g., fill in the blank and short answer), and to performance tasks. Exemplar tasks and associated templates are included in Appendix C.

Think aloud sessions

Although re-visioned items would typically be based on existing items, there is a need to verify whether the targeted skills are actually elicited by the re-visioned or expanded items. It may be clear, for example, that Q1 in Figure 2 targets numeracy as an established construct. However, the extension of the task to a new construct—problem solving—needs to be checked. Therefore, the OAA Africa National Teams conducted think aloud sessions for such items. Think alouds, or what is termed in the academic literature as cognitive laboratories, is a method of studying the cognitive or social processes called upon by tasks as students work through them (Griffin & Care, 2014).

During a think aloud session, students orally report what they are thinking as they complete the task, providing valuable information about the internal, and invisible, processes that are being prompted by the task. These sessions can provide teachers and task developers with information about how students approach a task and any functional issues that need to be addressed in revising the task (Leighton, 2017). See Appendix B for think aloud guidelines.

During the think alouds, teachers or observers are asked to reflect on the following:

- Items' capacity to capture the intended skills, components and subcomponents
- The targeting of the items to the ability levels of the students
- Usability of the items:
  - Did the students have difficulty understanding the instructions?
  - Did the student have everything they needed to respond (e.g., pen, paper, space)?
- Other issues related to item checking:
  - Did students provide evidence of possible misconceptions?
  - Did something unexpected occur?
  - Did the students express interest or frustration?

The Gambia: Think aloud

To support teachers in conducting think aloud sessions with their students, the National Technical Teams in each country held a training session for teachers. For The Gambia, the goals of the training were:

- For the teachers to learn about think alouds as a process that allows students to explore their thoughts in dealing with a task.
- For teachers to be acquainted with the procedure and manner in which think alouds are administered to learners.
- For teachers to begin thinking about how to develop similar test items and administer them to their students appropriately.

It was emphasized to the teachers that the focus should not be on whether the student was able to answer the item correctly, but rather, on the processes that led the student to the answer. For example, when the student is faced with an issue, how does the learner think? What are the processes he/she considers and the skills, components, and subcomponents he/she has applied to arrive at the answer? Understanding the answers to these questions goes well beyond whether the response is right or wrong, which has typically been the focus of our teachers.

After the discussion, teachers were placed in groups and instructed to administer the think alouds to other group members to practice what they learned. They were tasked with identifying a group leader, secretary, teacher, and a student, and to report their observations.

Once the session was over, teachers were asked to administer the items to students in their respective schools and provide feedback. The teachers found the think aloud sessions with their students both challenging and interesting.

They stated that their students had never been exposed to such kinds of tasks, thus making it more challenging for the students. At the same time, students really enjoyed taking part in the think alouds because it was different. The teachers also began recognizing the skills related to problem solving and collaboration and noticing specific behaviors when students were working on the task. For example, in one collaborative task, students discussed the question among themselves to develop a single response where they could agree. During the course of the discussion, they debated and countered each other's opinions before coming to a consensus. However, one of the issues that emerged was that because collaborative tasks were new, they had not been exposed to how to structure the collaborative work. At times, students found it difficult to understand what they were supposed to do as part of the task, especially if the items were not multiple choice or in a format with which they were familiar. As such, in some groups, one person tended to dominate the whole session, while others just observed; in other cases, one or two students would dismiss others' ideas without consideration, thereby not engaging in collaborative processes. When the teachers saw these different levels of competence, the tasks' capacity to capture indications of different student performance became clear.

During the think alouds, teachers made observations and provided feedback about specific items, such as which were too difficult to understand due to language issues, or which were not appropriate in terms of the content. Quite apart from the utility of the think aloud method as a process within task development, the teachers became more aware of the variety of student responses to curricular content.

The data from all think alouds across the three countries were consolidated. Then, based on the data and feedback from teachers and observers, the tasks were revised, along with the template form for each which describes each item structure. Use of these templates is intended to make it easier to create new tasks that can elicit the same skills across different subject areas and grade levels. The templates basically act as a guide for task and item development. Additional guidance describing the rationale and development can be found in the fourth report of this OAA series.

GENERATING NEW 21CS ITEMS BASED ON TEMPLATES

Before generating new items, each National Team decided which subject areas and grade levels to focus on for assessment development (Table 3). Within the subject areas, specific topics were selected as starting points.

The intent was to develop items that: 1) capture the skills and the components of interest; 2) capture these skills at increasing levels of proficiency; 3) are recognizable to teachers as capturing the skills; and 4) have structural features that can be replicated by teachers. Keeping these in mind, three main questions were discussed for generating new items:

- What are the characteristics of high-quality items?
- What makes a good stimulus for an item?
- How do you design a good scoring rubric?

Characteristics of high-quality items include a clear intention; language understood by most students; a simple and authentic context; and strong probability of achieving acceptable answers to the targeted skill. To develop items that can target the skills, creating stimulus material warrants careful attention. A good stimulus is rich and interesting; is optimally challenging (i.e., not too difficult or too easy); does not pose artificial challenges; offers opportunity to pose searching questions; offers opportunity for students to show what they know; and is equally accessible and equitable for students of different abilities. How teachers interpret and record student performance is equally important to consider when developing items. Therefore, to minimize the influence of variation in interpretation and subjectivity, development of a set of scoring criteria or “rubrics” contributes to consistent marking. A rubric can be developed by setting precise guidelines for judging students’ work.

There are several recommendations for writing rubrics which include descriptions of performance across levels of quality. Ensure that:

- each successive description implies a higher level of performance quality;
- behaviors are directly observable;
- inferences can be made about developmental learning—there should be no counts of things right and wrong;
- there is no differential weighting of
- responses - allow the rubrics to account for differences in performance;
- just one central idea is recognized through the increasing levels of quality; comparative
- terms such as more or better, are not used to define quality;
- transparent language is used (no jargon).

Table 3. Grades and subject areas for each country

Country	Grade Level(s)	Subjects
Zambia	6 and 8	Social studies, Math and science
DRC	6	Math, health/environmental/science, and technology
Gambia	4,5, and 6	Social and environmental studies, English language, math, and science

2 TASK AND ITEM DEVELOPMENT APPROACHES

Building on the previous workshops on 21CS and think alouds, the National Teams held additional sessions in their countries to engage teachers in the process of item development. Teams worked across two different approaches to item development (Figure 3).

Through both approaches, National Teams supported teachers to try the items in their classrooms, collect feedback, and refine the items based on the feedback. These processes identified and addressed challenges and issues that arose and helped teachers understand the implications of the way they developed the items.

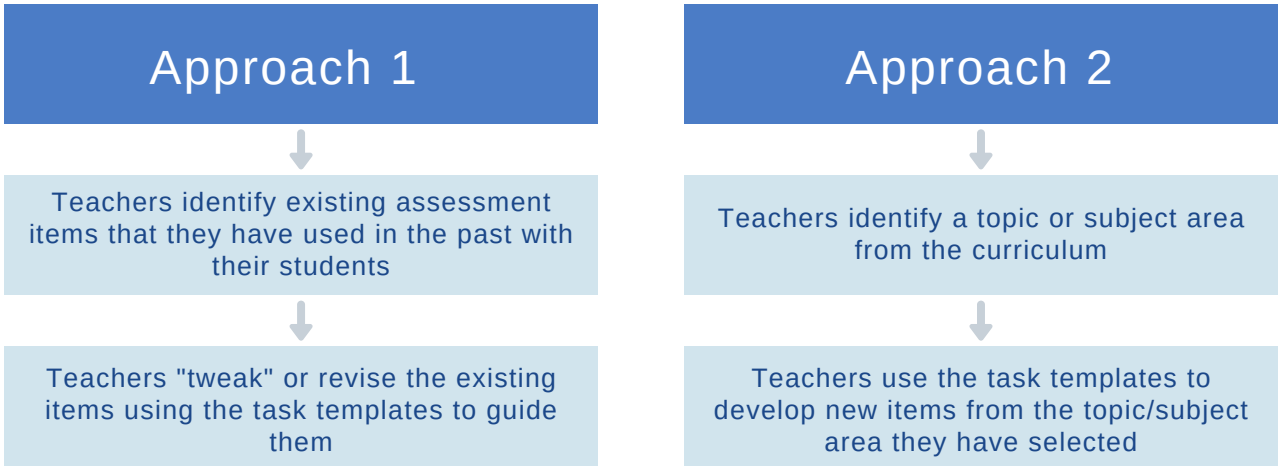
Each National Team worked with teachers in their respective countries to develop eight tasks, for a total of 24 tasks across the three countries, that targeted problem solving and/or collaboration in the subject areas of environment, mathematics, health, English, science, and social studies. There was a mix of task and item types, including dichotomous (correct/incorrect) response, closed constructed response, open constructed response, and performance.

Team members from The Gambia, and Zambia paneling tasks

Paneling of tasks: National and collaborative

Once the tasks were developed, the National Teams met together with their teacher teams to panel the tasks and their associated items. The paneling processes were undertaken in each country in slightly different ways but all with the same goal. Paneling is a process to check and improve items for the purposes of quality assurance, establish content and construct validity, and explore inadequacies in items, and reduce waste in piloting of inadequate items.

Figure 3. Two approaches for item development



Ideally each paneling group would include two or three independent experts, a representative of the item writers, and a couple of teachers at the target grade level and subject area used as the base for the tasks. This distribution would ensure the expertise to examine and revise test items is available.

A chairperson or group leader should facilitate and manage the discussion, and summarize what needs to be done to revise the items. The independent group members and teachers should review the tasks individually and then share perspectives, before requesting clarification or inputs from the item writers. The role of the latter is to respond to these requests rather than defend their items, and explain the rationale for the various features of the items. At the end of the process, a decision is reached concerning whether to discard, amend, or approve the tasks and accompanying items. In addition, the comments and rationale for decisions are noted to ensure the teams have all of the necessary information.

The panel assesses items based on the following criteria:

- whether the item assesses (part of) the construct;
- what students need to know to answer the question and
- whether the curriculum has covered this;
- the authenticity of the item;
- the precision and clarity of the item and its phrasing;
- the amount of time needed to produce an answer;
- adequacy of the scoring rubrics; and
- equity for students of different backgrounds.

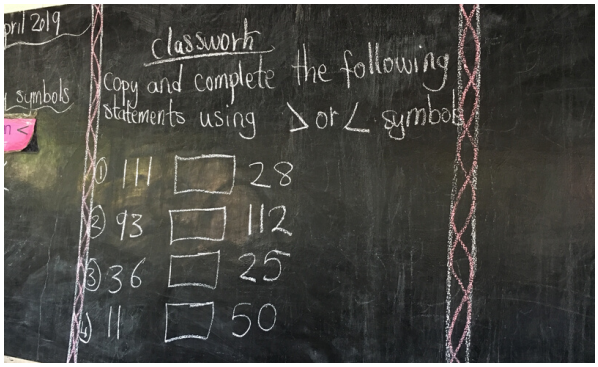
DRC, The Gambia and Zambia team members working across languages

Following the country-based paneling activities, both the newly developed, as well as previous tasks, were again paneled and reviewed in the fourth OAA workshop through a collaborative process. Beyond the review of items, the OAA workshop's focus was on the scoring criteria for items. Review of scoring criteria acts as another stimulus and opportunity to analyze the tasks and items themselves, since it is the specificity of the scoring protocols and coding that tend to highlight previously missed issues. A significant input to the review was documentation of student responses to the items from the think aloud activities in each country. Through interrogation of the written responses of the students, the full richness of their varied ways of thinking sheds light on the strengths and weaknesses of task and item design. In the collaborative workshop, sets of 10-15 student responses to each item were reviewed, categorized at different levels of quality, and then referred back to the scoring criteria. On the basis of the process, the scoring criteria and rubrics were reviewed, and in some cases, the actual items themselves were amended.

3 PILOTING OF TASKS

The purpose of the pilot activity was to check whether tasks such as those that had been developed could be administered at a classroom level, and whether student responses were interpretable in terms of the hypothesized skills. Analysis of the results would enable further finessing of the task templates and provide more information about likely student capabilities in these previously untested 21CS to guide future teaching and assessment.

Based on feedback from the collaborative paneling sessions, seven tasks from across DRC, The Gambia, and Zambia were dropped. The rest of the tasks were adapted across countries as needed for the pilot. Three tasks from the Asia pilot (see "OAA: Focus on Asia" report) were also included, against the possibility that future studies might link the data from the Asia and Africa pilots. The pilot required students in each country to complete the tasks in classroom conditions. Grade 6 students in DRC, Grade 7 students in The Gambia, and Grade 6 and 8 students in Zambia participated in the pilots. Table 4 shows the pilot tasks administered in each country.



Classroom tasks act as stimulus for assessment revis

The piloting process included several steps at the country level:

- Selection of schools;
- Slight revisions and translations of tasks as needed;
- Training sessions around test administration and scoring with teachers, data collectors, and other stakeholders participating in the project;
- Test administration;
- Scoring processes; and
- Data entry.

Each country approached the pilot process slightly differently. To provide a country perspective, DRC describes its process.

Table 4. Pilot items

All Tasks*	Z1	Z2	Z4	Z5	Z6	Z7	G1	G3	G4	D1	D2	D3	D4	D6	D7	D8	A1	A2	A3
Zambia Grade 6			Z4		Z6		G1		G4		D2						A1		
Zambia Grade 8	Z1	Z2		Z5		Z7							D4						A3
DRC										D1	D2	D3	D4	D6	D7	D8	A1		
Gambia			Z4					G2	G3							D8	A1		

Note: Tasks are identified by the country which developed them: Z = Zambia items; G = The Gambia items; D = DRC items; and A = Asia items. G3 and G4 items by The Gambia were re-labeled for the pilot.

The DRC: Pilot of the tasks

The National Team from DRC developed a timeline for the pilot, planning to start the various activities mid-September, two weeks after the start of the 2019-2020 school year. However, the timeline had to be shifted due to unforeseen circumstances, including delays in receiving the funds to implement the pilot; the mandate for experts to implement the free basic education announced by President Félix Tshisekedi; and the preparation of the Mid-Term Review of the Project for the Improvement of the Quality of Education (PAQUE) funded by the Global Partnership for Education. Thus, the piloting activities started in mid-October, rather than earlier.

For the DRC, the items were piloted with Grade 6 students, as the items were appropriate within the context of their curriculum. The items were based in mathematics, science, environment, and health, with five items for problem solving and two for collaboration. From the three items written by Asian countries, the DRC team and the item writers chose one.

Training sessions

To prepare for the training sessions, the National Team members developed materials and guides, including a presentation of the overall study, the concept of 21CS, presentation of the items to be included in the pilot, and scoring processes, as well as logistics associated with students in physical classrooms for test administration. Two training sessions were conducted in two different areas. In the city of Kinshasa, the capacity building workshop took place on October 10, 2019. It involved 11 teachers and four supervisors (the Deputy Inspector General in charge of Assessments, Director of School Guidance, and two item writers).

In Mbanza-Ngungu, the capacity building workshop was organized on October 18, 2019 with 15 participants, including the Provincial Director of Enseignement Primaire, Secondaire et Technique (EPST), the Principal Inspector of EPST, the Sub-Proved, the Senior Provincial Inspector in charge of Evaluations, Teaching Advisers, and three teachers. The training sessions covered how to administer the tasks and how to score student responses.

Selection of schools

Two schools selected at the beginning were replaced. A school in Kinshasa was replaced due to construction on the road leading to the school; a rural school in Mbanza-Ngungu was replaced because there was heavy rain, and schools were cancelled the day before the pilot was scheduled. Thus, in Kinshasa, Primary School (EP) Kengo, urban private school was replaced by EP1 Manyanga; and in Mbanza-Ngungu, EP ZANZA by EP1 KOLA.

In the end, five schools were selected in the city of Kinshasa and three others in the Kongo Central Province, around the city of Mbanza-Ngungu, 150 km from Kinshasa. The selection of schools was made with the assistance of the Provincial Directors of EPST, after communicating to them the desire to have a range of schools that varied on location, size, and socio-economic status. Based on the criteria, the Provincial Directors consulted the Heads of Education Divisions and School Directors.

Task administration

The administration of the tasks involved 120 students, including 75 from the five schools in the city of Kinshasa and 45 from the three schools in Kongo Central. With the delay in starting the activities, the team chose to group the students in one school in Kinshasa and one school in Mbanza-Ngungu.

The Kinshasa site was at BOBOTO College, which received 60 students from four schools. To note, having a large group of 60 students from four Kinshasa schools in the same room made student supervision difficult, especially with the lack of trained teachers. The Collège Des Savoirs opted not to send their students and instead held the pilot within their own school. In Mbanza-Ngungu, the children were grouped together at EP1 KOLA.

The administration of the items revealed one of the characteristics of the level of teaching and learning in the DRC, namely performance disparity between urban and rural areas and between schools in the same environment. Indeed, the students of Mbanza-Ngungu all claimed to have reading and writing difficulties. They mainly answered items that were multiple choice and had great difficulty with items that required writing.

The following are some observations based on the administration:

- The items were given to students at the beginning of Grade 6; however, the items were developed based on the full Grade 6 curriculum, which meant that the students had not yet covered much of the content upon which the problem solving and collaboration skills were to be applied.
- The vocabulary appeared too difficult for many of the students to understand, and needed review prior to further use.
- The teachers understood the importance of the project and wanted to use these types of assessment tasks to encourage students to deeper reflect.
- Teachers mentioned the need for better training for themselves to develop and use assessments of 21CS.

Scoring process
Notwithstanding the training that had taken place, and teachers' practice of the scoring process, it became clear that there had still not been sufficient time for adequate training in scoring. Thus, there were some issues around scoring, especially around understanding which responses aligned with which level of quality described in the rubrics.

Data entry
The data entry was undertaken over a very brief period, and as such it was difficult for the national team to ensure quality control across actual test forms and the database. The process was carried out in accordance with the information provided in the codebook as closely as possible.

KEY POINTS

Although each country followed similar procedures, they each made different decisions on some key points. For example, unlike DRC, The Gambia administered items in their pilot to recently graduated Grade 7 students. This decision was made in the light of the point made by DRC, that students at the targeted Grade 6 level would not, at that time of the school year, have yet covered all the curricular content that was assumed by the items. The decision by Zambia to include both Grade 6 and 8 students provided them with a greater opportunity to explore the impact of curricular knowledge on the application of skills.

The issue of how much training is necessary for pilot implementation is also raised by the slightly differing approaches of the countries. As stated by DRC, factors outside the control of the education team impacted the amount of training provided to participating personnel.

In the meantime, in The Gambia, training was provided over a three-day period, touching on the genesis and background of 21CS, the specific skill areas of problem solving and collaboration, and the practical guidelines for administration of the tasks and scoring protocols. For Zambia, awareness of the large class sizes (averaging 80 learners per class) was factored into the training. In particular, the collaboration tasks posed a challenge for administration, with teachers needing to ensure systematic sampling to select the learners in each class. In each school, 42 students made up 14 groups of three learners. This made it possible for the trained teachers to assess the three-student group tasks.

These differences highlight the significance of the need for systemic action on professional development for teachers as countries update curricular goals. Shifting from a knowledge base to a competency-base requires both pedagogical and philosophical contextualization. These issues are discussed in the fifth report of this series.

Collaborative tasks in Zambia, and problem solving in nutrition classes

4 WHAT DO THE PILOT DATA TELL US?

The data from the pilot were analyzed separately for each country due to the relatively few tasks common across the countries. The objectives of the data analysis were to:

- Identify how the tasks and their items function;
- Determine limitations of the tasks and items and provide suggestions for improvement; and
- Provide user-friendly item maps to demonstrate targeting—that is, how well the tasks and items are matched to student ability.

The information from the pilot data then provides the countries with feedback on functionalities of the tasks. The task administration component also provides invaluable information about student responses as they struggle to understand these different approaches to assessment. And the scoring processes provide scorers and teachers with additional insights into how their students demonstrate the proficiencies of interest. Most importantly, the overall pilot provides evidence to support the validity of the assessment approach and the usability of the templates for future task and item development.

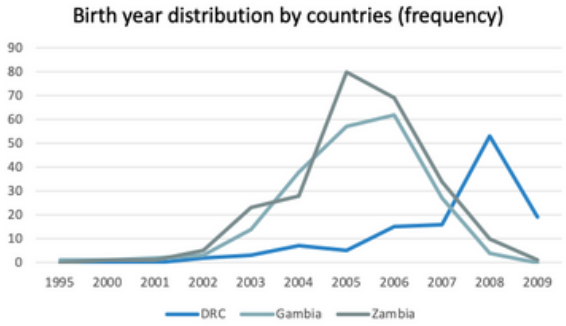
THE PILOT DETAIL

Student responses to the assessment tasks were derived from the student samples in Table 5. Note the differences in numbers of male and female students for DRC and Zambia. Figure 4 shows differences in the ages of the participating students across the three countries. The three countries drew on students across age ranges and grades, and is illustrative of the flexibility of the project in achieving its goal—to explore assessment approaches and specific types of tasks—as opposed to being interested in comparative student performance across countries.

Table 5. Total number of students in the pilot by gender across three countries

Country	Males	Females	Total
DRC	73	73	120
The Gambia	103	103	209
Zambia grade 6 and 8	156	156	348

Figure 4. Age distribution of students across the three countries



DESCRIPTION OF THE TASKS

The 19 tasks and their constituent items were distributed across problem solving and collaboration skills, with some base items in each reflecting content in the relevant subject area alone. Such base items, where they occur, are typically the first items within a task, and establish the context for the application of the 21CS.

The main components and subcomponents of the two skills that are represented across the administered problem solving and collaboration tasks are shown in Table 6. Although tasks and their items were deliberately drafted to target the skills, there was not a more structured plan in which specific numbers of items were to be written to specific components. Hence, the distribution of items across the various components provides a quick image of which components are possibly the most easily targeted using the processes adopted.

Collaborative subcomponents that are not captured by the tasks include turn taking, engagement, receptive communication, evaluation, and planning. Problem solving subcomponents that are not captured by the tasks include following a plan, comparing outcomes with plans, and synthesis. Some others are also infrequently captured directly, although they may be embedded within other actions. For example, although there is no one item that captures “follow a plan” directly, it is clear that the student needs to follow a plan in order to “consider and compare options”; similarly there is no one item that captures just “to ask questions related to the problem”, although the student must have done this mentally to organize or classify information. A process such as “analysis” also clearly underpins many cognitive operations.

ANALYSIS

The reliability of the problem solving and collaboration scales were calculated for each country dataset. The (EAP and WLE) coefficients for The Gambia were all above 0.80, demonstrating acceptable levels. Reliability for scales for DRC and Zambia did not reach acceptable levels based on the cut-offs applied. The reliability analyses must be interpreted with great caution for several reasons. First, the numbers of students are not large. Second, some difficulties in coding and scoring were experienced in the countries. And third, since some tasks included aspects of both problem solving and collaboration, analyzing these together makes the unidimensional assumption that the construct of “collaborative problem solving,” as distinct from two separate constructs of problem solving and collaboration, is being analyzed. The concept of collaborative problem solving is contested (Care & Griffin, 2017).

Table 6. Frequency of skill components and subcomponents represented in the pilot tasks

Component		Subcomponent		Processes	
Problem solving					
Information gathering	14	Organize information	14	Classify Analyze Describe	3 2 8
Planning a solution	19	Generate ideas and options	15	Hypothesize Consider and compare	7 9
Managing information	5	Justify the process	4	Explain	8
Collaboration					
Participation	10	Share	10		
Communication	7	Express	7		
Negotiation	7	Compromize	6		
Decisionmaking	8				

Note. Numbers of subcomponents and process activities do not necessarily sum to totals in the components column, since some of the former are superordinate to the components.

Consequently, for optimal use of pilot results, attention must be focused on interpretation within a task. Analysis of individual assessment tasks and their items generated information useful for rectification of some of the scoring rubrics, as well as recommendations for amendment of task main stimulus, or stem, of the tasks, and their items. The main issues encountered at the item level were: (1) lack of use of some scoring categories; and (2) lack of association of some scoring categories with overall performance (as indicated by zero-level point biserial coefficients for non-zero scoring categories, or negative coefficients associated with non-zero scoring categories). Further analysis is required within countries to verify whether low frequencies of some scoring categories are due to non-appearance of the responses that would populate those codes, or coding difficulties experienced by the scorers. The concerns expressed by the DRC national team about adequacy of scoring training may be germane to this issue.

One of the difficulties associated with assessment of complex skillsets is the accuracy with which particular responses or actions can be associated uniquely, or primarily, with a particular skill or subskill. Some of the items could be seen as indicative of different, although complementary, competencies. This reality could influence the difficulty of coding, particularly for educators who have been familiar with assessment that is focused on correct versus incorrect responses to knowledge-based items. The collaboration tasks and their items required greater rectification than did the problem solving tasks and items. This could well be due to educators' lesser familiarity with collaboration as a competency.

All tasks included multiple items. This approach was taken to make test-taking time efficient. Where just one set of stimulus materials can generate multiple items, the overall reading load for students is minimized. Item difficulty therefore depends not only on the initial stimulus, but also on the characteristics of each item. One assessment task can therefore include both relatively easy as well as average and difficult items. One of the advantages of this style of assessment is that examination of the item difficulties can be used to make decisions about best sets of items to use with students at different competency or grade levels, as well as scoring rubrics which further differentiate ability within items. As an example, Figure 5 shows the item-person map for DRC's student responses to problem solving tasks. The map includes six tasks (D1, D2, D3, D4, D8, A1), with their items and varying difficulty levels of responses to these.

The six tasks were designed to assess problem solving across a variety of science, social sciences, and mathematics topics. Figure 5 depicts what is more and less difficult for students, with items that appear toward the top of the graphic being more difficult than those toward the bottom. (For a description of how to interpret such graphics, please see the "OAA: Focus on Asia" report.)

As an example, Task D1 (a task created by the DRC team) is shown with responses ranging from low level of difficulty at a logit level of about -3.0 with item D1a_Cat1 to D1a_Cat3 at about logit 2.8. (“Cat” or category here refers to the quality level of the response, with Cat1 denoting the most basic quality level.) The task itself deals with issues associated with poliomyelitis, which is a current challenge for the DRC health system and society (World Health Organization, 2020).

The easiest item for D1 requires a student merely to attempt the question posed; the next level of difficulty is achieved when the student can identify either a cause or risk of poliomyelitis reflecting subcomponents of problem solving such as information gathering, organizing, and describing (D1a_Cat2), and then identifying the associations between a cause or effect (D1a_Cat3). A higher level is in principle possible—that of identifying two causes with their associated risks—but no student in this sample provided this level of response. For item D1b, the student is required to generate hypotheses within the context of the problem scenario. As can be seen, three levels of quality response are provided at D1b_Cat1, Cat2, and Cat3 depicting increases in difficulty from providing irrelevant responses to identification of actions, and then linking these with the target outcome through generation of hypotheses.

It is salient to note that the three OAA Africa countries wanted to acknowledge students' effort. This translated into scoring protocols in which an irrelevant response was treated as a higher quality than no response.

Figure 5. Problem solving distribution of DRC students against tasks

Logit	Person ability	Item category
4		D3a_Cat3
		A1a_Cat3 D4a_Cat2
		D1a_Cat3 D4bi_Cat3
		D4c_Cat2 D1b_Cat3 D4a_Cat4 A1e_Cat2 D8c_Cat4 D3b_Cat3
2	XX	D2d_Cat2 A1c_Cat2 D4bi_Cat2 D2c_Cat2
	XXXXXXXXXX	D8a_Cat3
	XXXXXXXXX	A1b_Cat2 D2b_Cat2 D8c_Cat3 A1f_Cat1 A1a_Cat2 D1a_Cat4
	XXXXXXXX	D4a_Cat3 A1f_Cat2 D8b_Cat3 A1d_Cat2 D8c_Cat2 D3a_Cat2
0	XXXXXXXXXXX	D1b_Cat2 D4bii_Cat2 D8b_Cat2 A1a_Cat1 D4c_Cat1
	XXXXX	D8a_Cat2 A1d_Cat1 D3b_Cat2 D8c_Cat1 D2c_Cat1 D2b_Cat1
	XXX	D1a_Cat2 A1b_Cat1 D4bii_Cat1 D8b_Cat1 D8a_Cat1 D2d_Cat1
	XXX	D1b_Cat1 A1e_Cat1 D2a_Cat1 D4bi_Cat1 D3b_Cat1
-2	X	
		A1c_Cat1 D3a_Cat1 D4a_Cat1
	XX	D1a_Cat1
-4		

Key: Example D1a_Cat1 = DRC task number 1, item a, response at category level 1

TASK REVIEW

The value of pilot results such as these lies in the interpretation of the data, in examining what sort of tasks, items, and response categories are easier or more difficult. Knowing what is challenging for students helps the teachers (and of course the developers of assessments) identify what their students are ready to learn next, what might need to be consolidated, and what might be currently beyond the learning readiness of their students.

Based on review of the tasks and their items, there were no tasks recommended for deletion, or in other words, deemed useless. Just four items were recommended for deletion. Amendments to scoring criteria were recommended for many items, with a sizable number constituting amendment or deletion of a scoring category, rather than amendment of the item itself. This brings into question whether the scoring categories, the scoring processes, or both are the source of some anomalies. Happily, where the same tasks were administered across more than one country, the same anomalies were found —indicating that the items themselves rather than the samples need to be explored.

Scoring categories recommended for deletion included those for which there were very few correct responses—indicating that the items were too difficult for the students.

The DRC Team in the final workshop reviewing the scoring criteria

In other cases, the association between the item response and student score was weak, indicating that the item was not functioning in the expected way. Where the mean ability did not increase across increasingly difficult response categories, items were also considered unsatisfactory. Several of these factors sometimes combined to lead to recommendation for deletion of response categories.

For example, the first item in one task asked students to identify two advantages and two disadvantages of the rainy and dry seasons. The scoring categories allowed for no response, irrelevant response, one advantage or disadvantage, and two advantages and disadvantages. Although the item categories functioned appropriately for the first three categories, this was not the case for the final one. In retrospect, this was due to a rubric problem in that it did not follow a logical sequence and was managed differently by countries. At the same time, the low frequencies for the highest level responses were probably accounted for by the expressive literacy levels of students. The DRC OAA team noted that students were less well prepared to cope with questions that required a lot of writing, as opposed to responding to multiple choice or short answer questions. The Zambia OAA team commented post-pilot that there were reading challenges for some of their students.

In this same task for the second item, and its three sub-items, as it shifted to assessment of collaboration components, students were required to agree on the most important advantage and disadvantage of the two seasons, based on their pooled earlier responses. Again the “easiest” levels of responses functioned appropriately, but responses that demonstrated students' negotiation with each other through linking of ideas were very few.

The Zambia OAA team commented that their learners appeared to have few group discussion skills, were unused to conventions of sharing information and then providing feedback, and were shy. In addition, the concept of collaborating as opposed to competing led some students to be reluctant in their sharing of information and ideas.

Analysis of the pilot results allow not only for critical review of the tasks and items themselves, but also serve to highlight cultural learning conventions that have been nurtured in schools for many years. Moving to assessment tasks that value student thinking, discussion, perspective, and interpretation is a philosophical shift —for students, as well as in some cases, for teachers.

The pilot experience provided the national teams with a rich source of information to draw on for their continuing 21CS task development. The training, task administration, and scoring issues provided an equally valuable resource to inform the three education systems about the infrastructure required to effect change.

CONCLUSION

The approach adopted by OAA Africa focus countries is symptomatic of a bottom-up approach to change in educational assessment. It takes the position that assessment forms and practices need to be understood by those education practitioners most closely involved in the actual administration and instructional use of assessment data. Beyond just understanding, the approach takes the position that practitioners need to create and develop their own assessments, and that these assessments should be aligned with technical expertise at central education department or ministry levels. Rather than adopting a model in which teachers and schools are mere receivers of assessment materials, it is a model where assessment becomes a part of instructional design.

Adoption of such a model carries with it risk, of course. The major challenge is ensuring that teachers and school leadership teams receive the professional development inputs that are needed for implementation. Even in the small pilot study reported across the three countries, it is clear that the professional development delivery varied considerably across countries. This is in no way due to fault or inadequacy on the part of the national teams, but is a function of the varied resources and infrastructure in each country, as well as the realities of political and economic issues, health crises, and conflicts.

Long-term engagement in problem solving and collaboration leading to common understandings and friendships – au revoir

REFERENCES

Care, E. & Griffin, P. (2017). Assessment of collaborative problem-solving processes. In B. Csapó, & J. Funke (Eds.), The nature of problem solving. Using research to inspire 21st centurylearning. (pp. 227-243). Paris: OECD Publishing.

Care, E., Kim, H., Vista, A., & Anderson, K. (2018). Education System Alignment for 21st Century Skills: Focus on Assessment. Center for Universal Education at The Brookings Institution.

Education, E. (2013). The Zambia Education Curriculum Framework 2013.

Griffin, P., & Care, E. (Eds.). (2014). Assessment and teaching of 21st century skills: Methods and approach. Springer.

Herman, J. L. (2010). Coherence: Key to next generation assessment success. Assessment and Accountability Comprehensive Center. Retrieved from https://files.eric.ed.gov/fulltext/ED524221.pdf.

Kim, H., & Care, E. (2020). A*ssessment in sub-Saharan Africa: capturing 21st century skills.* UNESCO and Brookings Institution.

Leighton, J. P. (2017). Using think-aloud interviews and cognitive labs in educational research. New York, NY: Oxford University Press.

Ministère de l'Enseignement Primaire Secondaire et Initiation à la Nouvelle Citoyenneté, Ministère de l'Enseignement Technique et Professionnel, Ministère de l'Enseignement Supérieur et Universitaire, & Ministère des Affaires Sociales, Action Humanitaire et Solidarité Nationale. (2015). Stratégie sectorielle de l'éducation et de la formation 2016-2025. Available from: http://www.eduquepsp.education/sgc/wp-content/uploads/2018/06/Strategie-sectorielle.pdf.

Ministries of Basic and Secondary Education and Higher Education Research Science and Technology. (2016). Education Sector Policy 2016-2030 Accessible, equitable and inclusive quality education for sustainable development. Available from: http://www.edugambia.gm/data-area/publications/policy-documents/256-education-policy-2016-2030-web-version.html.

The Ministry of General Education and The Ministry of Higher Education. Education and skills sector plan 2017-2021. Republic of Zambia. Available from: https://www.moge.gov.zm/download/policies/Education-and-Skills-Sector-Plan-2017-2021.pdf.

World Health Organization. (2020). WHO Health Emergency Dashboard. (https://extranet.who.int/publicemergency).

APPENDIX A: NATIONAL TEAMS

Country	Teams	Schools
Democratic Republic of Congo	Mr. Jovin Mukadi Tsangala, Conseiller au cabinet du Ministre de 'Enseignement Primaire, Secondaire et Professionnelle, Cabinet du Ministre Mr. Kasang Nduku, Expert chargé de la formation, Secrétariat Permanent d'Appui et de Coordination du Secteur de l'éducation (SPACE) Mr. Smith Mpaka, Coordonnateur de la Cellule Indépendante d'Évaluation des Acquis Scolaires Mr. Mapasi Mbela Chançard, enseignant au Collège des Savoirs Dr. Jerry Kindomba, Country Director, Giving Back to Africa	**Zone urbaine** Ecole primaire EP1 BOBOTO Ecole primaire CS MANYANGA Ecole primaire EPA 2 GOMBE Ecole primaire EP1 BINZA Ecole primaire COLLEGE DES SAVOIRS (péri-urbain) **Zone rurale** Ecole primaire EP1 BOKO Ecole primaire EP1 KOLA Ecole primaire EP MBAMBA
The Gambia	Mr. Momodou Jeng, Director, Science and Technology Education and In-service Training Unit Mr. Ousmane Senghor, Head of Assessment Unit Mr. Omar Ceesay, Education Officer Mrs. Isatou Ndow, Vice Principal, Gambia College Mrs. Saffie Nyass, Deputy Head Teacher	St. Peter's Lower Basic School Mansa Kolley Bojang Lower Basic School Abuko Lower Basic School St. Mary's Lower Basic School
Zambia	Mr. Victor S. Mkumba Principal Curriculum Specialist Social Sciences; Directorate of Standards and Curriculum Mr. Lazarous B. Y. Kalirani, Principal Education Standards Officer Tertiary Education; Directorate of Standards and Curriculum Mr. Shadreck Nkoya, Assistant Director Research and Test Development; Examinations Council of Zambia Ms. Beatrice B. Mbewe, Teacher Vera Chiluba Primary School; Ministry of General Education	Kabulonga Girls Secondary School Mount Makulu Secondary School Parklands Secondary School Vera Chiluba Primary School Matipula Primary School Chibolya Primary School

APPENDIX B: THINK ALOUD GUIDELINES

A “think aloud” activity (also called a “cognitive laboratory”) is a method of studying the skills and subskills that are used by a student when engaging in assessment tasks. Think aloud activities can provide valuable information about whether target skills and subskills are actually prompted by a task. As students work through the task, they orally report their own mental processes (i.e., explain their thinking and reasoning as they complete the task) so that these can be recorded by the teacher/observer.

These guidelines have been prepared to guide think aloud activities.

For **individual tasks** (i.e., problem solving), a small number students are needed (preferably selected across the range of estimated low, medium, and high ability by their teacher).

For **group tasks** (i.e., collaboration), two groups of the required number of students are needed (ideally including low, medium, and high ability mix of students.) [Note that different collaboration tasks may require different numbers of students].

The same students can be used for all tasks. However, if the tasks take more than 60 minutes, additional students should be selected to participate due to fatigue issues.

INSTRUCTIONS TO TEACHER/OBSERVER

Tell students to work as independently as possible. Students should "think aloud" as they complete the task.

If students need help while completing the tasks, you should first prompt the student to ask his/her group members (if it is a collaborative task), or think some more (if it is an individual task). If this is unfruitful, you should note the question for which help was provided and include a brief description on the think aloud record form.

Record the information from each student or group for each task on a single form. The forms are provided specific to each of the tasks.

READ THIS TO STUDENTS

“We are asking you to help us as we create new tasks for students. You will not be marked on the task–we just want you to help us. So, don’t worry if you do not understand anything or are unsure—we need to know how you go about working out what to do.

The tasks you will be doing are designed to assess [collaboration/problem solving]. This means you will have to work with partners/alone to find out what you have to do and solve the problems you are given. If you get stuck, you should try and work out what you have to do rather than ask me.

We asked for your help today because we want to know what you think about when you work on these tasks. In order to do this, I am going to ask you to THINK ALOUD as you work on the different problems. What I mean by think aloud is that I want you to tell me EVERYTHING as you work on the task. I would like you to talk aloud CONSTANTLY from the time you start, until you finish. I don’t want you to try to plan out what you say. Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking so that I know what’s going on.

If you stop talking at any point, I'll remind you to KEEP TALKING. If you feel uncomfortable or don't wish to continue, you can stop the tasks at any time.

Do you understand what I want you to do?"

Remember

- Sit near the student but not in their personal space.
- If the student is silent for more than a few seconds, prompt with **Keep talking or What's happening?**
- If the student is talking too quietly, say **Please speak louder."**
- Be sure that the student is also providing her/his responses or taking action to the task, not only talking—say: **Please provide/write down your response.** If the student is having trouble providing responses and talking simultaneously, have the student talk first, then provide her/his responses.
- If the student asks you what to do because s/he does not understand a question, tell her/him: **Maybe think about it another way**, or if a collaborative task: **Ask your group members or do whatever you think makes sense.** You should not help them solve the task.
- Be attentive with body language by head nodding or smiling in response to students.
- Do NOT tell the student if s/he got an answer right or wrong.
- Do NOT tell the student if s/he did well/poorly on the activity.
- Do NOT show bias for certain questions or item formats (e.g., Do not say anything like, "This is not a very good problem." OR "Problems like these don't test many skills.")

Facilitation

Provide the task. After a couple of minutes of the student having started, give the student quick feedback. Tell the student if they need to speak more, or are doing what you need. You may need to model thinking aloud in order to help the student understand what to do.

Note taking

1. As the student completes each task, make notes on the "Think Aloud Record Form."
2. Many of the tasks may be unfamiliar to the students. We do want to know if there are issues with the way the tasks are presented. You should note these issues in the comments section of the form.

    - Usability issues:
        - Did you see evidence that the student had difficulty understanding the instructions
        - Did the student have everything needed to respond (e.g., pen, paper, space, or other)?
        - Briefly summarize any usability issues encountered in the comments.
    - Comments:
        - Summarize any issues not mentioned above. Did students provide evidence of possible misconceptions? Did something unexpected occur? Did students express interest or frustration?
        - Keep a record of when you did not understand what the student was doing.

"Think Aloud" Record Form

Teacher / observer name			
Grade level of student/s			
School name			
Student name			
Task number: 1	**Circle YES or NO to confirm statements below; otherwise leave blank.** **For time needed, please write number of minutes**		
Target skill:	Task appears to test the targeted skill	YES	NO
	Task appears to test the targeted skill	YES	NO
	Task appears to test the targeted subskill/s		
	• Info gather – Organize info – describe	YES	NO
	• Plan solution – Generate hypotheses	YES	NO
	• Plan solution – Develop plan	YES	NO
	Task was ☐ easy, ☐ appropriate, ☐ difficult for student/s		
	Students knew how to complete the task without help	YES	NO
	How many minutes are needed for this task?	mins	
Comment here if there are other kinds of answers than are currently catered for in the marking instructions for this item	**Enter observation notes and comments here; include student think-aloud commentary and answers.**		

APPENDIX C: TEMPLATES AND EXAMPLES

Each template presented is followed by a task example. More detail about how such tasks capture specific components and subcomponents, and their scoring codes, can be found in the fourth report in this series.

Template 1	**a)** Describe a problem (e.g., health, social, science, environmental) **b)** Suggest what can be done to solve the problem	• Info gather – Organize info – describe • Plan solution – Generate hypotheses
Task 1	**a)** Describe one environmental problem that you know about. What causes the problem, and what is the result? **b)** Suggest what can be done to solve the problem.	• Info gather – Organize info – describe • Plan solution – Generate hypotheses
Template 2	• Pose an arithmetic calculation with one number missing and with at least two mathematical operators, for the student to identify the missing number. • Provide at least one inaccurate answer, and ask how this could have been calculated. • Request that the student create a similar task.	• Plan solution – Generate hypotheses – Consider and compare • Plan solution – Develop plan – Predict
Task 2	I think of a number. I multiply it by 3. The answer is 18. **a)** What is the number? i. 6 ii. 9 iii. 21 iv. 54 **b)** Your friend Fatima chose answer (iii). How do you think she got her answer? **c)** Your friend John chose answer (iv). How do you think he got his answer? **d)** Create a similar numeric task to a), with four response options, one of which is correct, and two of which resemble the reasoning shown by Fatima and John. (e.g. "I think of a number. I divide it by "x". The answer is "x". etc.)	• Plan solution – Generate hypotheses – Consider and compare • Plan solution – Develop plan – predict
Template 3	Item that requires the student to generate [x] hypotheses in order to explain a situation in terms of cause and effect.	• Generate hypotheses • Manage info – Justify – Explain
Task 3	Fertilizer can poison people. Fertilizer was spread in the fields by the farmers. People in the village became sick due to the fertilizer. How did this happen? Name one reason if: **a)** a drought followed straight after the fertilizer was spread; **b)** heavy rain occurred straight after the fertilizer was spread.	• Generate hypotheses • Manage info – Justify – Explain

<table>
<tr>
<td>Template 4</td>
<td>Provide information.

a) Show examples of misinterpretation of the information and
b) Request explanations for how the misinterpretations were reached.
c) What can you do so that these misinterpretations do not recur?</td>
<td>• Generate ideas – Consider and compare
• Justify – Explain
• Plan solution – Generate hypotheses</td>
</tr>
<tr>
<td>Task 4</td>
<td>Seven boys went to play and collected some small stones - shown in the table. Three students calculated the following:

a) Zane found the mean as 37+25+20+16+23+11+25 = 154
Do you agree with Zane? Explain your answer.
b) Davide found the median as the middle data value 37, 25, 20, 16, 23, 11, 25 = 16
Do you agree with Davide? Explain your answer.
c) How many more small stones should each collect for the mean to be 40? Show these in the table.

<table>
<tr><th>Students</th><th>Small stones</th><th>Additional small stones</th></tr>
<tr><td>Zane</td><td>37</td><td></td></tr>
<tr><td>Ousmane</td><td>25</td><td></td></tr>
<tr><td>Davide</td><td>20</td><td></td></tr>
<tr><td>Paul</td><td>16</td><td></td></tr>
<tr><td>Christian</td><td>23</td><td></td></tr>
<tr><td>Jovin</td><td>11</td><td></td></tr>
<tr><td>Jose</td><td>25</td><td></td></tr>
</table>
</td>
<td>• Generate hypotheses
• Manage info – Justify – Explain</td>
</tr>
<tr>
<td>Template 5</td>
<td>Provide problem that includes a finite amount of a resource.

Provide information about different people having different needs for the resource and the amount of the resource needed for each activity.

In a group, each student takes on a role and identify:

a) What is best outcome for self? (individually)
b) What is the best outcome for all? Each member shares what is best outcome for self before deciding as a group what is best for all. Explain reasoning (as a group).</td>
<td>• Generate ideas – Consider and compare
• Justify – Explain
• Participation – Sharing
• Communication – Expressive
• Decisionmaking</td>
</tr>
<tr>
<td>Task 5</td>
<td>Serrekunda is an over-crowded town. Large families including grandparents, and their adult children with their wives and husbands and children, all live in small two- or three-room houses. One issue that these families experience is that there is not enough coal for everyone to use as needed.

Different members of the family have different daily needs and uses for coal.
• The mother needs coal to make food.
• The grandmother needs coal for heat during the day to stay warm and watch the baby grand-daughter.
• The young student needs coal for light to study.

Each of these activities require different daily amounts of coal.
• 7 units of coal to make food
• 8 units of coal for light
• 8 units of coal for heat

The family has only 15 units of coal total for each day.

In a group of three students, identify who will take on which family role (mother, grandmother, or young student).

Question Set 1:
Individually, state what is the best outcome for yourself.

Question Set 2:
As a group, discuss and come up with the best outcome for all members of the family. Explain your answer.</td>
<td>• Generate ideas – Consider and compare
• Justify – Explain
• Participation – Sharing
• Communication – Expressive
• Decisionmaking</td>
</tr>
</table>

Template 6

Provide a problem or prompt. Then in a group of three students, go through the following process:

	Student 1	Student 2	Student 3	Subskills
Round 1: Think of factors and proposal of points by each individual member	Factor and Proposal	Factor and Proposal	Factor and Proposal	• Generate ideas – Consider and compare • Participation – Sharing • Participation – Turn taking • Communication – Expressive
Round 2: Justification for your perspective from each member	Justification for proposal	Justification for proposal	Justification for proposal	• Justify – Explain • Participation – Sharing • Participation – Turn taking • Communication – Expressive
Round 3: Consensus #1 and reason	Agree with Member 2	Own proposal	Own proposal	• Justify – Explain • Participation– Turn taking • Communication – Expressive • Negotiation – Compromise
Round 4: Justification	Justification for proposal #2	Justification for proposal	Justification for proposal	• Justify – Explain • Negotiation – Perspective taking
Round 5: Consensus #2 and reason	Agree with Member 2	Own proposal	Agree with Member 2	• Justify – Explain • Decision making
Task 6	Koffi has breathing problems and coughs a lot. The doctor says his illness is not related to a virus or bacteria, but rather to the air he breathes. Work as a group of three students. **a)** Each student in the group must suggest at least on factor that can contribute to poor air quality. **b)** Next, each member of the group must give at least one reason why their factor is the most important. **c)** Next, as a group, the students must agree on the factor that is most important.			