

Event Summary

AI & Assessments

AI for Education

FAB INC.

FAB DATA.IO

On 3rd October 2023, Fab Inc. and the **Bill and Melinda Gates Foundation** held the first in a series of convenings.

This was the first step in our programme to build a community to work collaboratively to harness the potential of artificial intelligence for education.

This document provides notes from the event.



Contents

One	Overview
Two	Aims
Three	Speakers
Four	Discussion topics
Five	Presentations
Six	Summary of discussions
Seven	Next steps

One

Overview

The AI & Assessments event, held virtually on 3rd October 2023, brought together nearly 100 participants from across 17 low-income countries, mostly across sub-Saharan Africa, alongside participants from North America and Europe. These included:

1. Educators from government
2. International and local NGOs
3. Implementers
4. Tech-startup networks
5. Technology developers
6. Large assessment companies

Two

Aims

- 1. To showcase promising uses of AI tools**
that support summative and formative assessment in low and middle-income countries (LMICs).
- 2. To highlight**
areas where further research and development are needed.
- 3. To facilitate learning**
among key players in the field.

Three

Speakers

Co-hosted by Fab Inc. and BMGF, the convening kicked off with high-level overviews from the organisers. This was followed by discussion of innovations in high-income countries by **Prof Rebecca Allen** and then three presentations outlining three current innovations using AI for assessment in sub-Saharan Africa:

1. Angaza Elimu

Using generative AI to develop and evaluate test items for formative and summative assessment

2. Rising Academies

Optical character recognition for marking

3. Binding Constraints Lab*

Voice recognition to facilitate assessment of reading level

Prof Matthias von Davier, an expert in assessment, then provided reflections on the three presentations from a high-income perspective, situating them within the global progress of the field.

** A consortium led by the Binding Constraints Lab with Neurabuild, Western Sydney University and the University of Cape Town.*

Four

Discussion topics

Participants were then divided into breakout rooms to explore two topics:

1. How can AI help plan and develop tests that are fair, valid and useful for teachers, learners and other stakeholders?
2. What are the technical challenges of using AI for marking, analysing, and improving learning and potential solutions?

These discussions highlighted areas where further research and development is needed to facilitate widespread implementation of AI tools for assessment in LMICs.

This note summarises the presentations, key takeaways from the reflections and discussions, and proposes next steps.

Five

Presentations

1. Framing the Discussion

María José Ogando Portela, Fab Inc.

2. Using generative AI to develop and evaluate test items for formative and summative assessment

Kiko Muuo, Angaza Elimu

3. Using Optical Character Recognition to Evaluate Learners' Work

Alexandra Fallon, Rising Academies

4. Using AI to automate Early Grade Reading Assessments

Sipumelele Lucwaba, Binding Constraints Lab

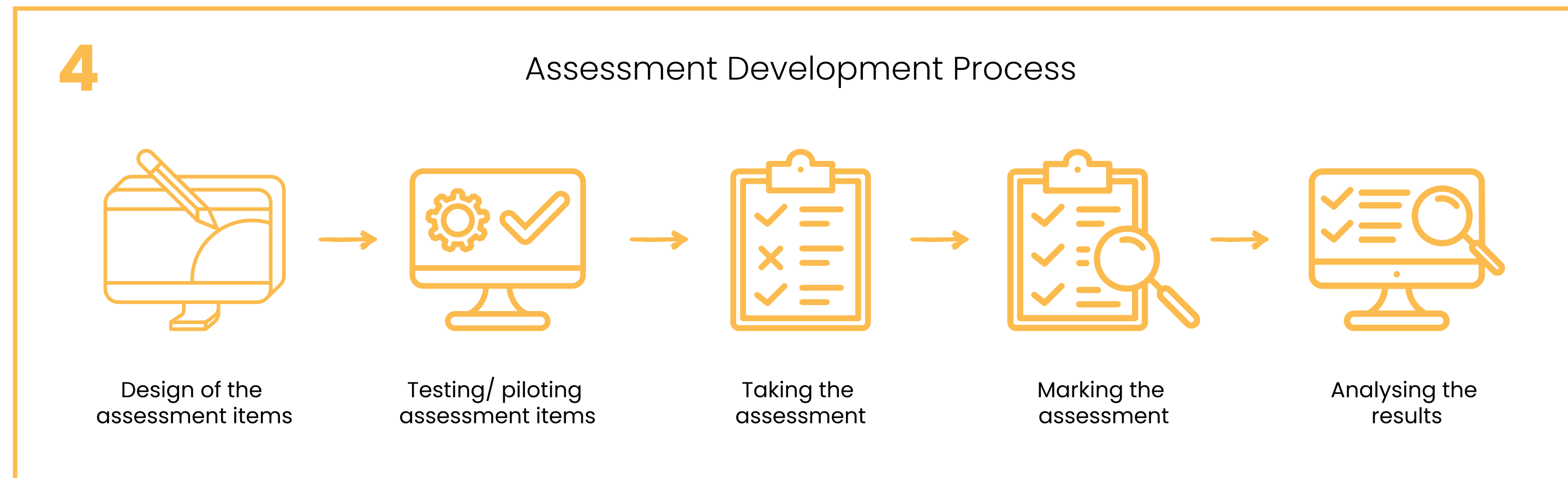
Five

Presentation One: María José Ogando Portela, Fab Inc.
Framing the Discussion

Planning

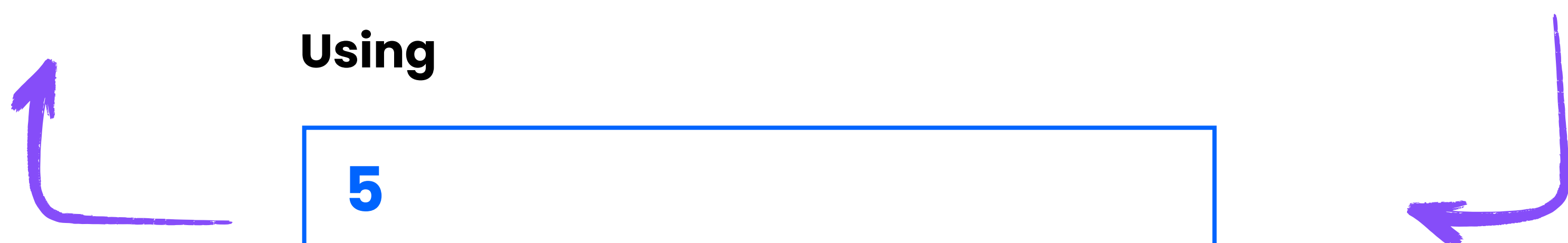
- 1**
Why assess?
- 2**
What to assess?
- 3**
How to assess?

Preparing, adapting, implementing and analysing



Using

- 5**
Changes at the system, school and classroom level.



Five

Presentation One Framing the Discussion

Learners in LMICs are not acquiring the skills and knowledge they need. In part this is because formative and summative assessments have not adapted and evolved to meet the needs of a modern education system and curriculum. However, AI has the potential to completely change how we conduct and use assessments.

The diagram helps us understand the focus of emerging innovations in assessment, what works and what does not, and where there are still gaps that need further exploration.

In conversations with EdTech innovators, assessment experts, and academics, we found innovations focusing on the design of test items (test questions). We also came across uses of AI that can improve the method of taking the test, such as using AI in oral reading fluency tests. These AI tools also improve the scoring and marking of such tests. Finally, there are applications for AI to bridge the gap between the analysis of the test results and their use by stakeholders, from class teachers to policy makers.

Although the examples presented during this event explore individual components of AI enhanced assessment, there is great potential to combine them to provide a comprehensive suite of assessment tools for key aspects of FLN.

However, there are some risks of 'hallucination' with AI generation of test items and if they are introduced without proper checks, it could undermine confidence in the use of AI in assessments more generally. Therefore, starting at a small scale and with low-stakes assessments is the best strategy.

“

One crucial takeaway is the importance of keeping the assessment objectives and the intended use in mind when discussing the role of AI tools in assessment.

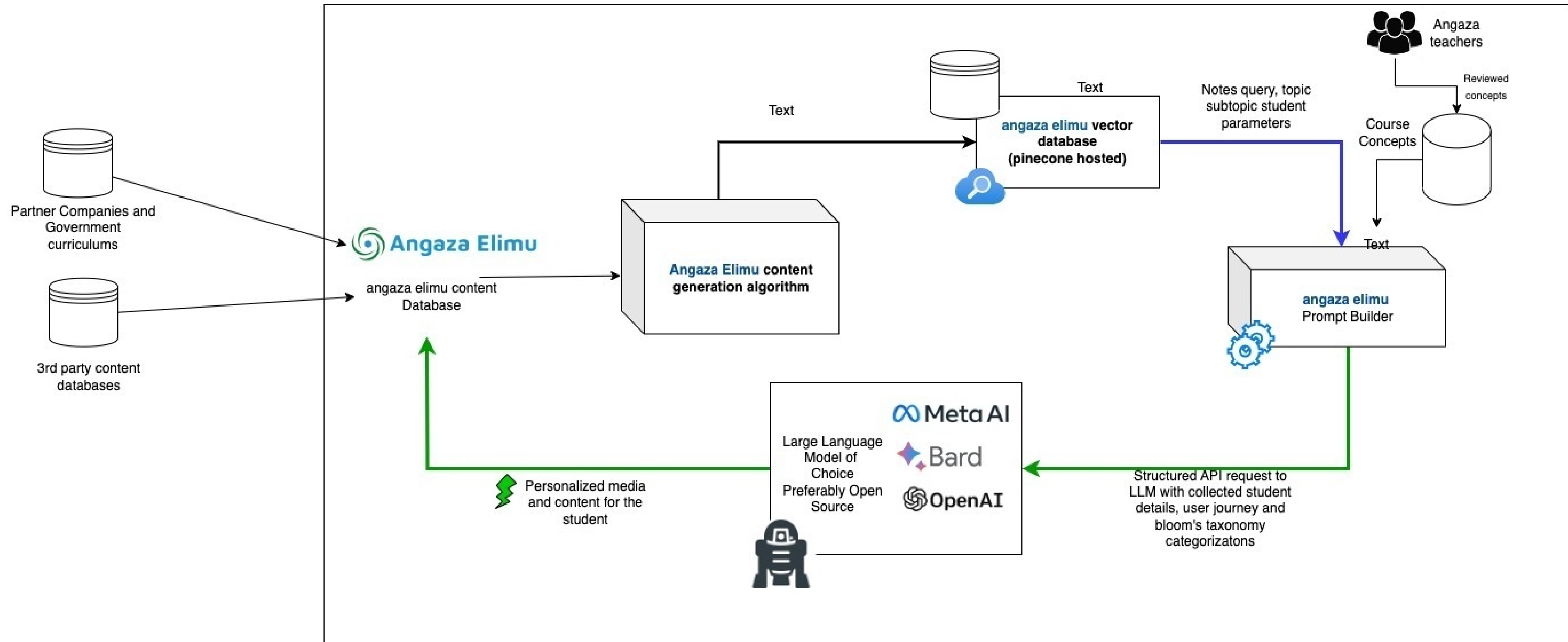
”

**María José
Ogando Portela**
Fab Inc.

Five

Presentation Two: Kiko Muuo, Angaza Elimu

Using Generative AI to Develop and Evaluate Test Items for Formative and Summative Assessment



Five

Presentation Two

Using Generative AI to Develop and Evaluate Test Items for Formative and Summative Assessment

There is a lack of personalised learning in the Kenyan education system and as a result, teachers have little information on the progress made by individual students.

Angaza Elimu is an adaptive and interactive platform that uses a mobile app or web interface to provide a personalised learning experience for students. An important component of this personalisation is formative assessment items developed using Large Language Models (a subset of Generative AI that deals specifically with the generation of text, using generative models trained on previous examples). Assessment allows the app to build a profile of students' abilities and share this data with teachers.

To use Generative AI effectively for test item generation, rich sources of data have to be identified (e.g Kenyan publishers and educational institutions). A lot of time is spent cleaning this data to ensure consistency. Once ready, the data sets are fed into the Angaza Elimu AI models and test items are generated. These were then reviewed by human expert teachers who contextualised and evaluated them. This sped up the process of item generation.

However, striking the correct balance between AI-driven assessment and human oversight is an ongoing challenge. Human expertise plays a pivotal role in validating the assessment items produced by generative AI to ensure accuracy and relevance. Furthermore, human expert teachers also provide additional context and guidance on specific learning objectives in order to improve the models and produce better assessments.

“

Robust personalisation is achieved through assessments generated by AI tools that are better tailored to learners' needs. These provide better data for targeted teaching interventions.

”

Kiko Muuo
Angaza Elimu

Five

Presentation Three: Alexandra Fallon, Rising Academies Using Optical Character Recognition to Evaluate Learners' Work

The image displays a document analysis interface. On the left, a sample document is shown with handwritten student work. The document is titled "G1 - Math Version C" and includes fields for Name, Age, Gender, and ID (14690). The work is organized into sections: "S 1 - Write" (problems 1a-1d), "S 2 - Write" (problems 2a-2d), and "S 3 - Solve" (problems 3a-3f). Handwritten answers are visible in boxes: 1a) 6, 1b) 13, 1c) 5, 1d) 27; 2a) 11, 12, 13, 14; 2b) 45, 46, 47, 48; 2c) 10, 15, 20, 25; 2d) 10, 12, 14, 16; 3a) 4 + 3 = 7; 3b) 6 + 2 = 8; 3c) 7 + 0 = 7; 3d) 9 - 6 = 3; 3e) 5 - 4 = 4; 3f) 12 + 4 = 4 + 15.

On the right, the "Results" panel shows a search bar and a "Segment by line" dropdown. Below, the segmented content is displayed as a grid of buttons, each containing a snippet of the original document's content, such as "G1 - Math Version C", "14690", "Name:", "Age:", "Gender:", "S 1 - Write", "S 2 - Write", "S3 - Solve", "3a) 4 + 3 = 7", "1a)", "6", "2a) 11 12 13 14", "3b) 6 + 2 = 8", "1b) 13", "2b)", "45 46 47", "48", "3c) 7 + 0 = 7", "1c) 5", "2c) 10 15", "20 25", "3d) 9 - 6 = 3", "3e) 5 - = 4", "1d) 27", "2d)", "10 12 14 16", "3f) 12 + 4 = 4 + 15".

At the bottom of the interface, there are two buttons: "Choose a sample document" and "Upload document".

Five

Presentation Three:

Using Optical Character Recognition to Evaluate Learners' Work

Hand-marking assessments is time consuming. Furthermore, the assessments tend not to be well analysed – often resulting in students receiving an overall percentage for the test with no granularity as to performance on each question. This leads to wasted time and wasted insight.

Optical Character Recognition (OCR) can help address these issues by turning pictures of handwritten text into machine readable and editable text.

Rising Academies has experimented with Google's AI tool (Document AI) and Amazon's AI tool (Textract) to mark Grade 1 mathematics scripts. The team's pilot shows that Document AI achieved 62% accuracy compared to human markers. Textract gave 86% marking accuracy. However, due to a difference in the setup, using Textract requires Rising to develop additional code to identify the key text (the answers). A deep dive of inaccurate responses shows a clear pattern with specific numbers not being read correctly (number 15 and 5). Spotting and correcting common mistakes resulted in 92% marking accuracy.

Rising envisions their next step to be making this data available and actionable for teachers (via WhatsApp message highlighting students' strengths and areas to improve).

“

There's a lot of value you can unlock by looking at question by question data, but doing those analytics by hand is really time consuming.

”

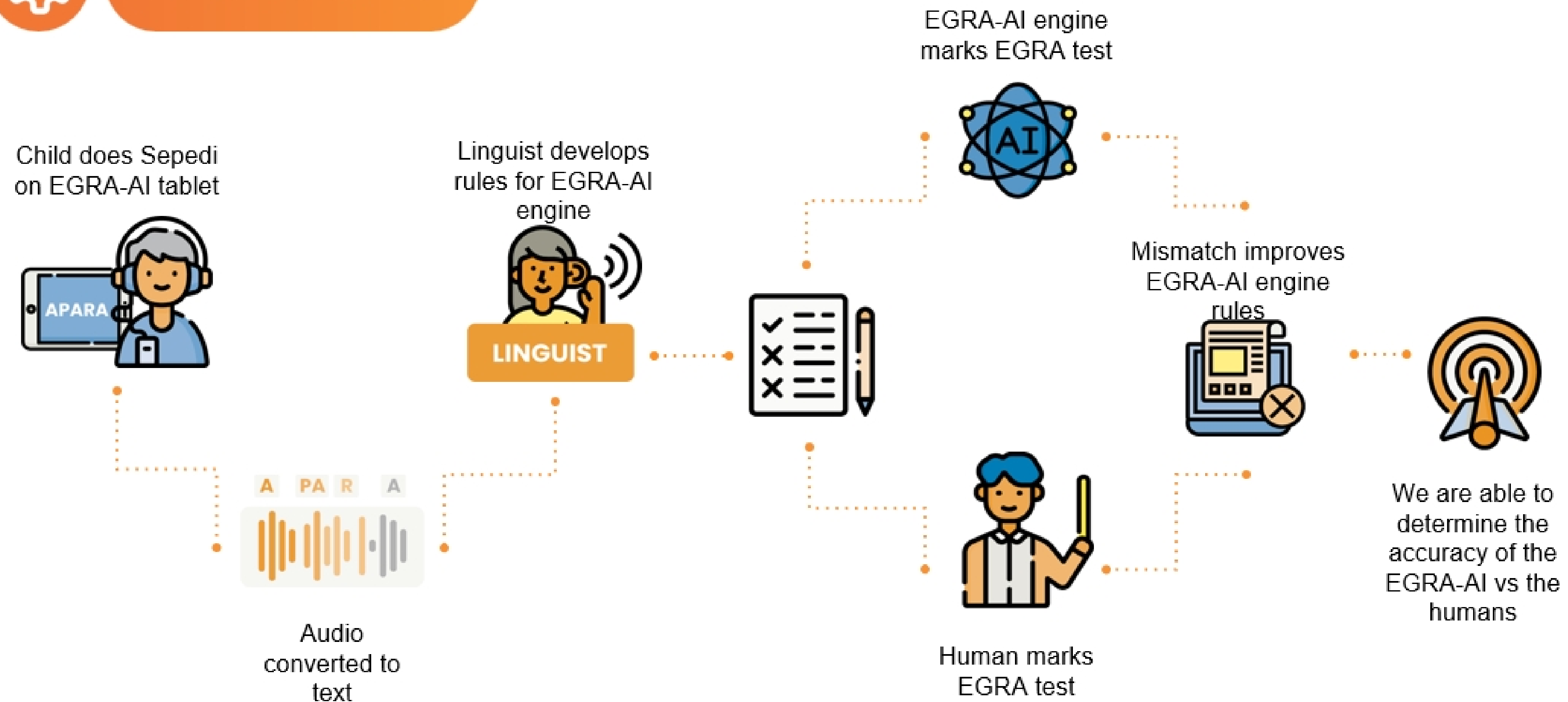
Alexandra Fallon
Rising Academies

Five

Presentation Four: Sipumelele Lucwaba, Binding Constraints Lab Using AI to automate Early Grade Reading Assessments

How it works

October 2023



Five

Presentation Four:

Using AI to automate Early Grade Reading Assessments

Early Grade Reading Assessments (EGRA) are reliable but are expensive and time consuming because they are done one-to-one and require trained facilitators. This cost makes it challenging to get data at a national level. The Binding Constraints consortium is focusing on automating EGRA reading assessments using a voice recognition AI tool [Wav2Vec2Phoneme](#) to record assessments taken in the Sepedi language and turning them into machine readable text.

The app was developed from Read Up, an AI tool first deployed in Australia. It was a good fit for several reasons. First, it requires little training data - very little was available in Sepedi. Second, it had already been tested in other African languages and had performed well. Finally, it is able to be deployed offline which is very important in areas with very limited or no connectivity (such as rural schools in South Africa).

To test the accuracy of the model, Wav2Vec2Phoneme converts children's reading into text. The accuracy of the reading is then graded separately by human markers and by a generative AI model trained by the group. The two gradings are then compared. Currently the reading model is being trialled in 50 schools in Limpopo province. Rules developed to adapt the Read Up model for the context include accepting rolling 'Rs' and substituting 'Ps' for 'Bs'.

Current work on the project is focused on: i) Refining the training model; ii) Exploring biases, e.g. gender, rural or urban, and; iii) Identifying the efficiencies created by using this model.

The opportunity offered by the development of this tool is that the cost of EGRA will be greatly reduced and it will become far more widely adopted in many more languages. This will allow a much more detailed picture of children's early reading abilities and needs to be developed.

“

The problem with reading is not only having actual interventions to address the problem, but also understanding what the challenges are at a classroom level.

”

**Sipumelele
Lucwaba**
Binding
Constraints Lab

Six

Summary of discussions

Following the speakers and reflection from noted academic assessment experts Prof. Rebecca Allen and Prof Matthias von Davier, the attendees split into 8 breakout rooms to discuss two broad questions relating to the speakers and to the framing diagram.

Six

Discussion 1:

How can AI help plan and develop tests that are fair, valid and useful for teachers, learners, and other stakeholders?

The main recurrent themes in this discussion centred around the triangle of accuracy, transparency (or explainability), and scalability.

1. Accuracy

How well a model performs in correctly predicting or classifying data points is paramount in the test being useful to stakeholders.

2. Explainability

Ability to understand and interpret the decisions or predictions made by AI models was seen as essential to acceptance. Not being able to explain why the model gives certain answers breed suspicions over bias and inaccuracy. Nonetheless, in the context where current, human-driven, practices were perceived as corrupt, AI tools can be seen more favourably.

3. Scalability

Ability of the AI system to efficiently expand its operations and adapt to the increased scope and complexity that comes with scaling up is also crucial. As such, we need to ensure transferability of AI tools, making sure that they are contextually and culturally appropriate. This translates directly to requiring more data and training, which can be achieved with collaboration among actors.

The importance of having humans in the loop was stressed many times by the participants. It is not the case that AI tools will replace humans completely. A more crucial question is on what the optimal roles of humans should be.

Six

Discussion 2:

What are the technical challenges of using AI for marking, analysing, and improving learning, and potential solutions?

There are a number of opportunities for using AI tools to mark and analyse student work. AI marking of student work will free teachers from a time consuming task. However, thought needs to be given to how this time is best used to improve the quality of learning. Formative assessment feedback needs to be better understood and valued both by students and teachers.

Participants also discussed how AI analysis of assessment data could highlight multi-country or multi-context issues which can then be addressed collectively e.g. challenging letter sounds for families of languages. This may have implications for policy around language of instruction.

In addition, the development of AI in assessment presents an opportunity to rethink the modes of assessment that are used to assess FLN. However, it is difficult to get FLN on the political agenda, and doubly difficult to get EdTech in FLN discussed at a policy level, participants agreed.

A barrier is that the costs of data analysis are significant but hidden. There are initiatives that are currently running which are being held back because the costs of file transmission, connectivity and servers are prohibitive. Finding a way to collectively address high infrastructure costs would support innovation in this area.

Seven

Next steps

The long-term goal for this project is to ensure that the supply of AI tools for education in low and middle-income countries (LMICs) are high quality, and available to meet the needs of their students.

We want to create a vibrant community of AI in education in LMICs. This challenge is too great for any single organisation. However, by sharing our ideas and building collaboratively, we aim to help shape the development of AI use in education in LMICs.

We will shortly be launching an Open Knowledge Platform. This will host materials used in the convening, and will grow to host ideas, guidance, shared knowledge and open-source code and tools to help the education community learn how to sensibly integrate AI into their work helping children learn.

It will be a space that helps shape the development of new tools and products, and the integration of AI into existing ones. It will help tech-innovators and educators by lowering the cost of their tools and shaping them to best practice for FLN. We welcome any contributions here - this is a space to share your ideas and learnings.

More details will be shared soon! In the meantime please contact Sirin Tangpornpaiboon if you have any queries.

Sirin.Tangpornpaiboon@fabinc.co.uk

AI for Education

FAB INC.

FAB DATA.IO