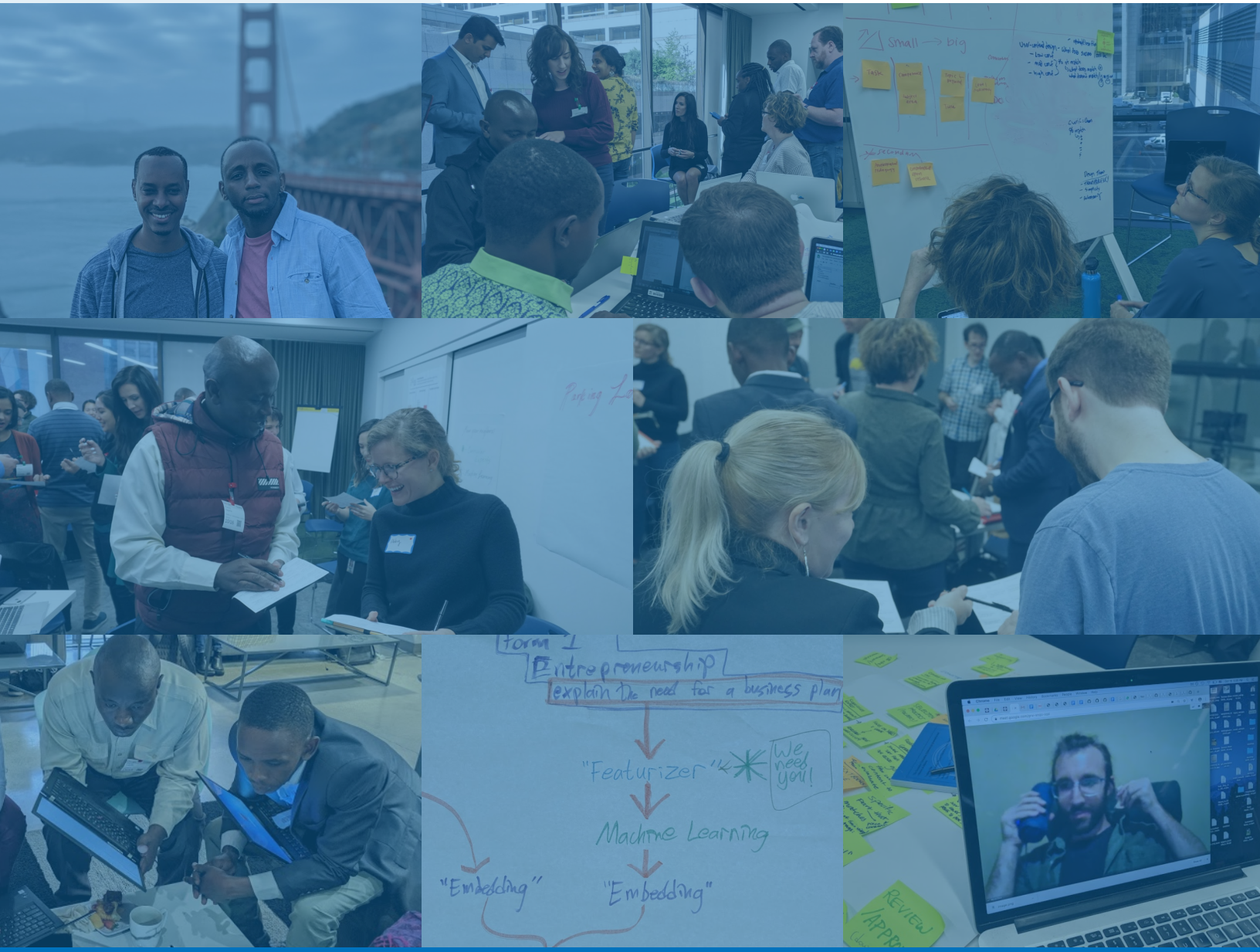
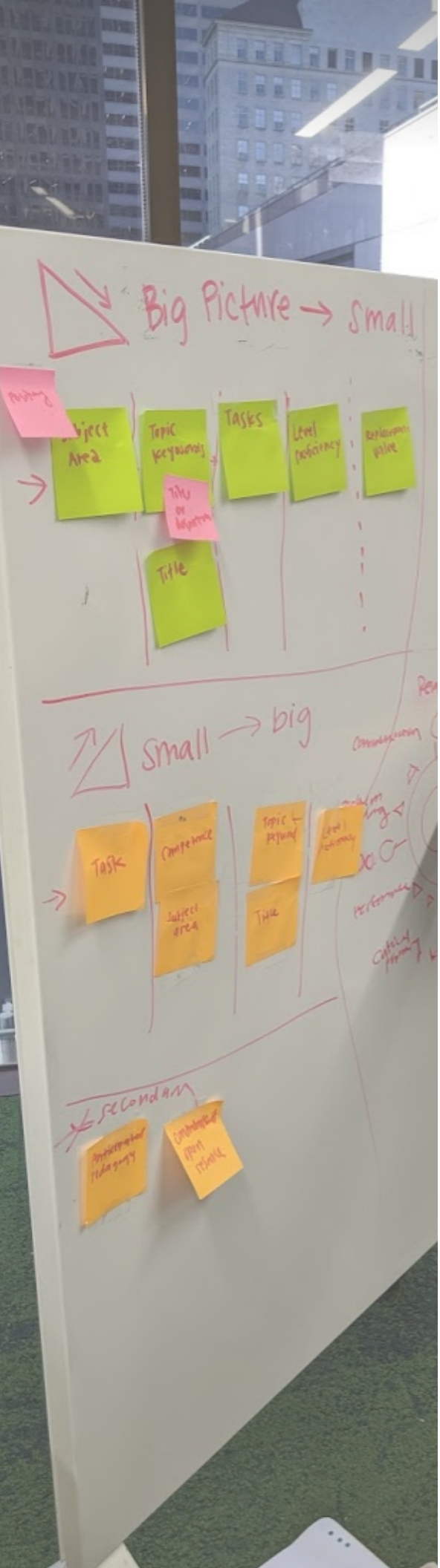
HACKATHON ON CURRICULUM ALIGNMENT

SYNTHESIS AND NEXT STEPS

16 - 18 OCTOBER 2019
SAN FRANCISCO, CALIFORNIA, USA



Google.org | Vodafone Foundation | UNHCR The UN Refugee Agency | LEARNING EQUALITY | UNESCO Education Sector
United Nations Educational, Scientific and Cultural Organization



REPORT CONTENTS

SUMMARY

On October 16–18, 2019, UNHCR, Learning Equality, Google.org, Vodafone Foundation, and UNESCO convened a hackathon in San Francisco, California at Google's offices. Ths hackathon aimed to prototype a tool or set of tools to automate aspects of the curriculum alignment process. The intention is that this will become a public good that can support curriculum alignment in emergency and crisis contexts, show the potential to be integrated with policy work, and strengthen cohesion around existing projects in the space to benefit both refugee and Open Educational Resource (OER) communities.To take an initial step in prototyping this work, we started by focusing on the curriculum itself (and not the content within curriculum) as an entry point. The hackathon focused on five project areas and achieved the following:

1. CURRICULUM DIGITIZATION

Goal: Get curriculum documents into a machine-readable format.
Achievement: Imported parts of curriculum documents for four countries, and created a workflow that partially automates the digitization of curriculum documents and also incorporated human review to correct software mistakes that could not be automatically detected or addressed.

2. HUMAN JUDGMENTS

Goal: Make relevance judgments to inform and improve the machine learning models.
Achievement: Developed a UI (iterated on as a result of user testing) to efficiently collect evaluations of relevance between parts of two curricula. This allowed us to collect enough initial data to help improve the models. Created an exploration UI to help understand the types of recommendations the models were making, as one moved through a curriculum.

3. RUBRIC DEVELOPMENT

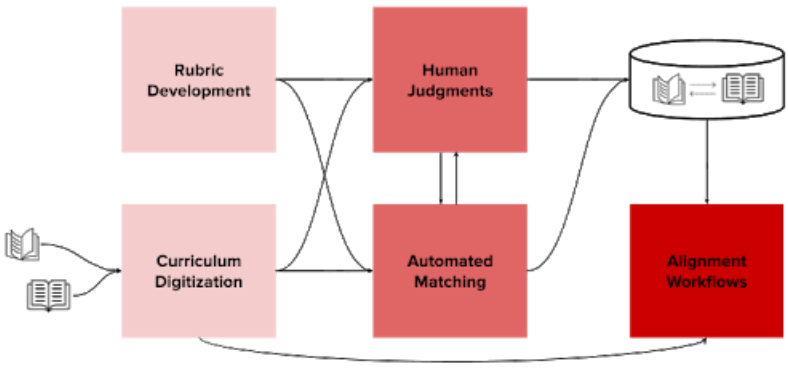
Goal: Articulate and understand the criteria for relevance judgments between two or more curricular standards.
Achievement: Co-created an initial rubric to serve as a basis for making comparisons between objectives across curricula, along with a set of concrete features that can be extracted from curricular documents to support automation of this process.

4. AUTOMATED MATCHING

Goal: Build and train algorithms to identify relevance correspondences across curricula.
Achievement: Explored a range of tools, including semantic text encoders (Universal Sentence Encoder, BERT, TF-IDF) and methods for factoring in hierarchical context as an input feature. Identified several human judgment criteria, and integrated methods to extract these as input features. Saw demonstrated ranking improvement using the initials available small dataset.

5. ALIGNMENT WORKFLOWS

Goal: Explore ways to present the outputs of the machine learning recommendation models and integrate then into a curriculum building workflow.
Achievement: Built designs to demonstrate what it could look like for users to engage with these types of recommendation systems within the context of a tool designed specifically for the rapid alignment of OERs to local curricular standards.



The diagram here illustrates how the projects fit together as part of the overall goal of using machine learning techniques to support the task of curriculum alignment. By starting with a clear definition of the task (rubric) and curriculum documents in machine-readable format (curriculum digitization), we were able to collect lots of training examples (human judgments), and develop machine learning models for identifying similarities across curricula (automated matching). We also designed the user interface for incorporating the machine learning recommendations into an overall user journey (alignment workflows). The results of this multi year engagement, and the direct results of this hackathon, have made important strides towards helping to automate parts of this process in order to significantly improve efficiency when performing curriculum alignment. In terms of what is next, we intend to continue to gather and generate data to improve the machine's ability to judge the similarities between learning objectives in different curriculums in order to provide increasingly accurate recommendations for those performing curriculum alignment.

PROBLEM SPACE

This hackathon built on two years of previous consultation and experience in the curriculum alignment process aimed at addressing the challenges faced in crisis contexts when there is a need to quickly categorize and align digital materials to national systems of hosting countries. While this process is also a need at the national level, there have been limited achievements in aligning digital curriculum to date. This is because this process is often time-intensive, conducted manually and is never ending since new content is being created that requires constant alignment, and there are ongoing changes to the curriculum.

HACKATHON APPROACH & AGENDA

The format of the hackathon was designed to be iterative. There were design sprint elements to one of the projects, but on the whole, the work that took place followed directly from the Paris design sprint which took place in March 2019 (read the report here). In taking a "healthy hackathon" approach, the hackathon itself took place over the course of three days and was also somewhat defined and limited based on the timing of the event, availability, and skillset of the ultimate participants. We continued taking a user-driven design and cross-sectoral collaboration approach to achieving these milestones in order to help move together the development of these public goods in a meaningful way. To achieve these milestones, organizers convened 40 hackathon participants, both in-person and remote, who helped to inform, build and test aspects of the set of tools we are collectively aiming to develop (see complete list of hackathon participants here). This includes individuals with expertise in machine learning, UX design, and curriculum design, as well as refugee leaders who spoke to the specific needs that we are designing for. It began with an in-depth discussion with five individuals supporting learners in emergency and crisis contexts in Kenya. They provided background on the problem space and particular contextual issues of note. From there, we presented an overview of the intention behind the tool we are aiming to build (as well as some of the work done to date), and divided participants into project groups for the remainder of the hackathon, focusing on different aspects of the curriculum mapping process. Individuals within each project group were assigned a role for them to take on for the remainder of the hackathon, with rotations and cross-pollination between groups happening as needed. The crux of the hands-on work took place over two days, with opportunities for cross-project collaboration as well as iteration. Through this process, the initial missions and challenge statements presented at the outset were refined based on feasibility and discussion. The hackathon concluded with presentations on each project, as well as a debrief to inform next steps and to enhance the conveners understanding of hosting such an event.

OPEN APPROACH

In the spirit of the open source community, all ideas and concepts produced as a result of this hackathon are intended to be taken forward with the appropriate MIT/Creative Commons licensing. All efforts will be made to make them as interoperable, widely sharable, and generally applicable as possible.

> ***I think it's important to have a variety of people... because each one will complement what the other person is saying...This is something that I feel would not be able to be done on a non-interdisciplinary level."***
>
> ***Flora Michti, TES***

PROJECT 1: CURRICULUM DIGITIZATION

Follow this project here: https://tinyurl.com/hack-project1

Why does this project exist?

Transforming scanned curricular documents into a meaningful machine-readable format ready for further processing.

What specific challenges did we tackle?

Get as many curriculum objectives and their structure into the database as possible in a relevant and usable format, while exploring approaches for addressing optical character recognition (OCR) limitations and making the process as streamlined as possible.

What was our approach?

We first approached the project from a primarily technical perspective, trying to see if we could create a fully automated workflow. We quickly realized, however, that even state-of-the-art OCR tools make a lot of mistakes in ways that varies between documents and is hard for code to consistently account for. We then adjusted the process to add a human review component to clean up and fix structural issues with the text, so that they could be machine-readable for import.

What did we achieve?

A workflow of using OCR plus code to digitize text and layout from curriculum documents, then having humans review the output to correct OCR mistakes and produce a spreadsheet that can be directly input into the database.

What is needed to take this forward?

- Explore tools for improved cleanup of OCR outputs.
- Gain access to more curriculum documents to increase the curriculum data in the database.
- Iterate further on what data is and is not useful for the machine learning process.
- Simplify, standardize and document the human review process.
- Find ways to incentivize and distribute human review, to allow for a quick and efficient review process.

PROJECT 2: HUMAN JUDGMENTS

Follow this project here: https://tinyurl.com/hack-project2

Why does this project exist?

This interface will produce data for the machine learning models in Project 4 - Automated Matching and give insights into how curriculum experts are making similarity judgments between curriculum standards.

What specific challenges did we tackle?

Designing and developing interfaces that prioritize ranges of quantity or quality of data collection.

What was our approach?

We used a prototype interface for curriculum standard comparison to elicit judgments from curriculum experts. We used observation and unstructured interviews about how the interface supported their judgment making, what information they were using, and what additional information might further support it. Through iterative design and development, we made gradual improvements to the interface.

What did we achieve?

We created a more engaging and streamlined user interface to collect user data. This reduced the cognitive load of the task, by doing sequential comparisons with a single reference standard.To support collecting more nuanced data about the reasons for users' judgments, we developed an additional design in order to understand in more depth not just the primary criterion for the judgments but individual judgments against each criterion. Leveraging the work of Project 3 - Rubric Development, we created a design to balance the two competing priorities of Quality and Quantity by integrating the rubric into the judgment interface.

What is needed to take this forward?

Implementation of the additional designs, including integration of the rubric from Project 3 - Rubric Development, combined with testing of each design to understand which produces the best tradeoff of quantity of data collected vs quality of judgments. It may also be useful to prompt for additional clarification and to provide collaborative tools for controversial judgments.

PROJECT 3: RUBRIC DEVELOPMENT

Follow this project here: https://tinyurl.com/hack-project3

Why does this project exist?

In curricular equivalence, "similarity" between learning objectives is not about equivalent meaning, but whether assessments and activities planned for one objective can be used for another. Some objective measure must guide effective judgment-making, provide a common language to discuss possible biases, and serve as a basis for judgment calibration.

What specific challenges did we tackle?

How might we present the criteria for judgments of similarity between curricular standards?

What was our approach?

We broke down the judgment-making process into 9 key dimensions of similarity and voted on their importance while making group judgments, with 5 emerging as paramount. We then discussed how to infer these from limited curricular text, creating approaches for algorithmic featurization. Identifying upwards of 40 such approaches, we selected 10-15 of the most feasible.

What did we achieve?

Our rubric identifies 5 key dimensions of similarity used in human judgment (summarized below) as well as a set of several recommended featurization approaches to extract or infer them, and indicators of high, medium, and low confidence of success in each.

Dimension	Definition	Possible Featurization Approaches
Keywords	Specific terms, concepts, and phrases.	• Linguistic/semantic matching
Level	Indication of learner proficiency required, inferred via age, grade level, prerequisites, etc	• Extract level indicator keywords • Average level through sibling nodes • Phrase/vocabulary lexile level • Semantic sentence complexity
Tasks	Skills, competencies, or desired outcomes.	• Blooms' Taxonomy verbs • Hierarchical classification as activity
Knowledge area	Domain of subject or topic.	• Identify cross-disciplinary keywords • Classify title case and font size
Specificity level	Degree of granularity within subject/topic.	• Allow/reject levels of depth

What is needed to take this forward?

Machine learning experts will need to assess the feasibility of the possible featurization approaches, and they will need to be reevaluated in light of the growing judgment dataset.

PROJECT 4: AUTOMATED MATCHING

Follow this project here: https://tinyurl.com/hack-project4

Why does this project exist?
The process of manually identifying correspondences between entries across two curricular structures is time-consuming and intensive, and would benefit from automated support tools.

What specific challenges did we tackle?

The process of manually identifying correspondences between entries across two curricular structures is time-consuming and intensive, and would benefit from automated support tools.

What was our approach?

We first set up an overall architecture upon which to build out and test various models:

- Featurizer: Take a curricular entry as input, extract useful features from its text (e.g. title, description) and its positioning in the structural hierarchy, and output a “feature vector”.
- Embedding Model: Feed the feature vector as the input into a multi-layer neural network that outputs a learned “embedding”, as a normed vector.
- Relevance Model: Take two curricular entries, pass them through the embedding model, and take the dot product of the embeddings to calculate a value representing their “relevance” (1 meaning perfectly relevant, and $\leqslant 0$ meaning not at all relevant).

We then evaluated performance based on how closely these relevance scores match the judgments made by humans in Project 2 - Human Judgments. We train the model by presenting these human judgments and adjusting the weights of the embedding model to better separate non-relevant pairs of curricular entries, and bring relevant pairs closer together.

What did we achieve?

We explored a range of tools for the Featurizer, including semantic text encoders (Universal Sentence Encoder, BERT, TF-IDF) and methods for factoring in hierarchical context as an input feature. Working with the Rubric Development project, we identified several human judgment criteria, and integrated methods to extract these as input features (for example, using Bloom's verb taxonomy to determine the level of complexity of a learning objective). Using the initial small dataset available to us, we already started to see ranking improvements using these methods.

What is needed to take this forward?

We need support to increase the size of the dataset by adding curricular data, through further digitization/extraction of source curricula; and adding human judgment data, through sharing the judgment user interface with a broader set of curricular experts. We also need to iterate on feature extraction and data pre-cleaning (with ongoing contributor support plus a potential Kaggle competition).

PROJECT 5: ALIGNMENT WORKFLOWS

Follow this project here: https://tinyurl.com/hack-project5

Why does this project exist?

The process of building a curriculum aligned to local educational standards is a time consuming task that requires curriculum experts to manually find and vet thousands of content items.

What specific challenges did we tackle?

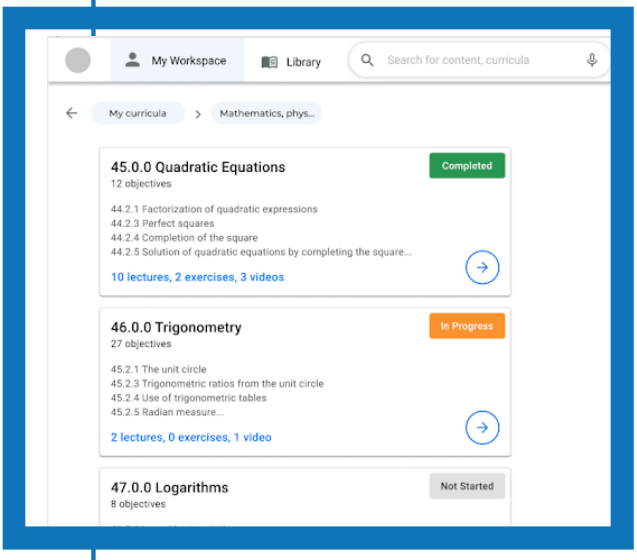
We designed the user-facing components of a platform for curriculum experts to build curricula efficiently based on machine learning recommendations for matching curriculum topics.

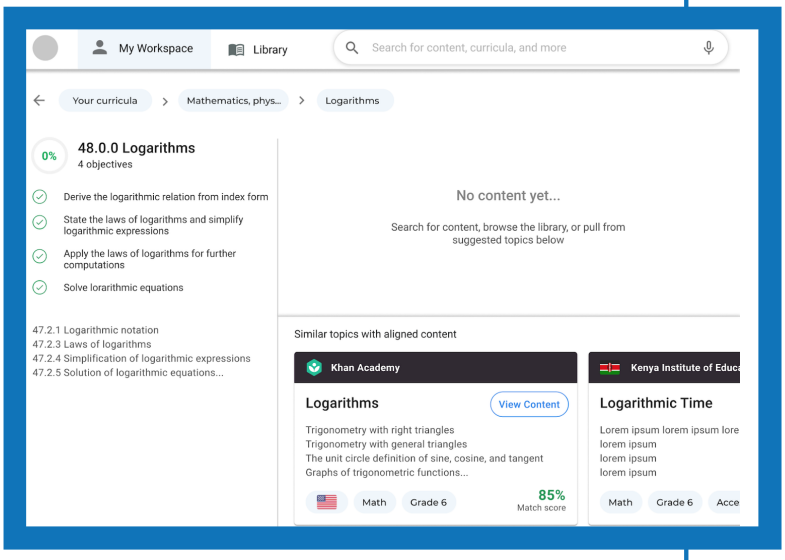
What was our approach?

We studied the user journey of a curriculum expert building a locally-aligned curriculum and identified the key moments. We then designed an end-to-end workflow for the task that guides users through the multi-step process by integrating information from different systems.

What did we achieve?

We designed a prototype for the user-facing component of the platform that guides curriculum experts through the process: A) selecting the target curriculum document (output of extraction phase), B) managing progress throughout the alignment process (shown below left), C) reviewing of "bulk recommendations" for entire folders of matching curriculum items from other countries (shown below right), and D) fine-grained selection and vetting of individual content items. The designs set the foundation for multi-user collaboration including splitting up the work into manageable chunks, progress tracking, peer review, and external approval.





What is needed to take this forward?

Build a prototype based on the user interface designs that connects to the live recommendations for matching curriculum standards and content items available in various OER repositories.

LESSONS LEARNED

As with any multi-stakeholder collaborative process in developing something completely new, there are lessons learned along the way. On a positive note, one exciting output of the hackathon is that we developed a strong foundation for how to effectively convene this type of diverse stakeholder group for future hackathons or other co-creation sessions, as there is more work to be done. We also validated what we already knew from our consultative process, which is that fully automated digitization is not going to work given that the structure of each curriculum document is very different. As anticipated, while we started the hackathon with some data thanks to pre-work done to establish an outline for several proposed projects with some existing curriculum, that only got us so far. Additionally, the challenges of digitizing curricular documents to OCR were further reinforced through this process. Lastly, while we had a sense of some of the constraints going into the event related to the availability, and skillset of the ultimate participants, we were fortunate as to the uniqueness of a multi-stakeholder hackathon and still arrived at concrete outputs that put us on a path towards our goal.

> ***This is something that's really happening. It's not all talk. It's action. We're all doing this, and that's so deeply empowering to be at an event that is, 'Yes, and...' and is inclusive and is thoughtful. We can make a difference in this space, and we're going to do it.***
>
> ***Abby Daniels, Google (volunteer)***

CALL TO ACTION

It is our intention building these tools will help to support:

- Governments to quickly evaluate whether digital resources created for another system/country could be used within their schools to support quality learning opportunities
- Educators to access diverse digital content, providing great autonomy and easier access to categorized resources that map to the level and learning objectives of their classes
- Content designers to distribute their digital resources to global audience
- Schools to refugees or new learners into their system, evaluating learning equivalence and identify materials to assist in bridging gaps
- Students to receive differentiated digital content and learning pathways to support their personal development
- Parents to access a range of free open-source digital resources to support their children's revision at home

Therefore, to help its continued development, we call upon the following stakeholders to support us in taking this work forward in the short-term.

CURRICULUM DESIGNERS/DEVELOPERS, EDUCATORS AND TEACHER TRAINING BODIES:

- Identify curriculum documents and participate in the crowdsourcing of digitization (likely manual data entry) of these documents to have more curricular data for judgment making.
- Use curricular expertise by support in the making of judgments to improve the machine's learning.
- Organize local events (in the next few months, especially) to use the UI developed through these projects to support the making of judgements/evaluations.

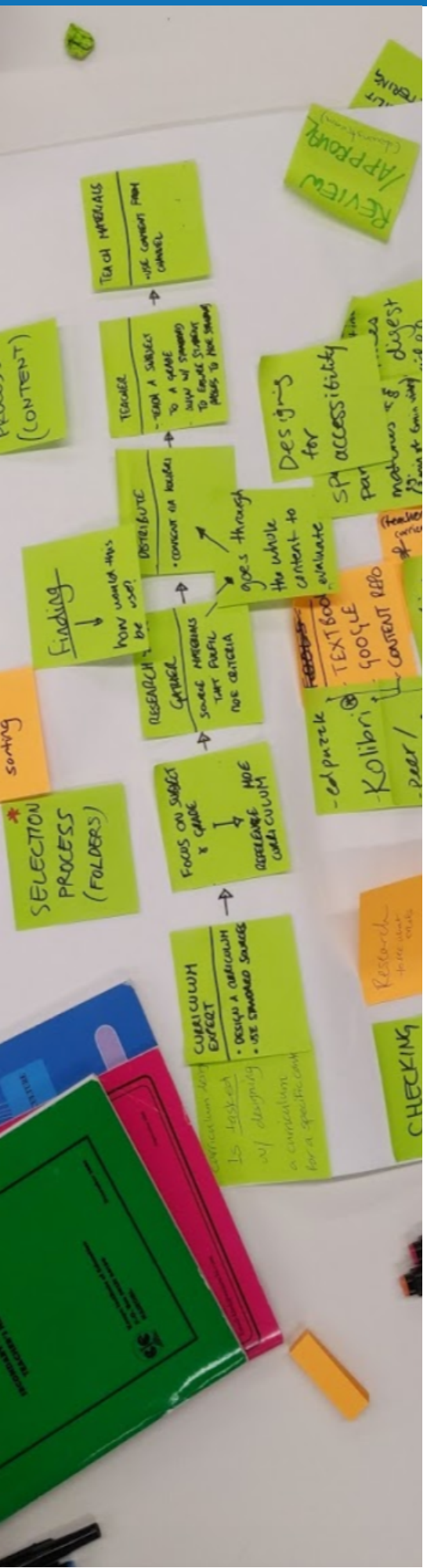
DEVELOPERS:

- Participate in upcoming Kaggle competition (stay tuned!) once enough data has been collected, and leverage your experience in machine learning, artificial intelligence, or natural language processing.
- Follow and contribute to the projects related to this effort using Google Colab notebooks and relevant GitHub repositories.
- Explore more efficient and effective ways of automating OCR and OCR post-processing.
- Build web-based applications and tools for fostering collaboration among the various types of project stakeholders, from curriculum experts to technologists.

GOVERNMENTS:

- Provide curricular documents and work with the organizers to support the digitization of curriculum that can also be made available for your own use.
- Consider digitization needs and best practices (i.e. use of consistent numbering terminology instead of font differences) for OCR when writing curriculum.

Want to get involved?
Email design2align@learningequality.org.



APPENDIX 1: FUNCTIONAL GLOSSARY

To address the needs of technologists and curricular experts having a common language during the hackathon, Project 3 - Rubric Development developed a functional glossary of terms. The complete list is linked above, but here is a subset of terms that may be helpful.

Curriculum [1] - The curriculum is a whole document (implicit or explicit) of all the standards (mandates which all educators must fulfill within that context, no matter what individual activities they do), and possibly the activities that match those standards, the proficiency at which they should be assessed, and so on.

Learning objective - This is what the student should know, understand, or be able to do at the end of the activity, class, or semester. Learning objectives should always be framed as tasks with verb statements, not noun phrases. For example: “Seek and give factual information” Synonyms in this context: learning content

Content - Knowledge components necessary for meeting the learning objective. Usually the information/knowledge that lesson plans should be able to deliver. (e.g., parts of a computer, functions of a computer, power safety, etc.)

Topic - Under a given subject area, a subfield within which a given learning objective falls; what the learning objective is “about.” For example, algebra.

Match - A similarity between two curricular nodes, or a piece of content and a learning objective. Criteria for the precision of the match--the number of attributes on which there is accordance - may vary depending on what it is used for.

Align/Alignment - considered equivalent from the perspective of an educator. When learning objectives are aligned, this means that they lead to similar skills or knowledge. For example, “Identifying prepositions in English” and “Understand prepositions in the English language”. The term can also be used to talk about a learning resource that can be used to teach a specific learning objective.

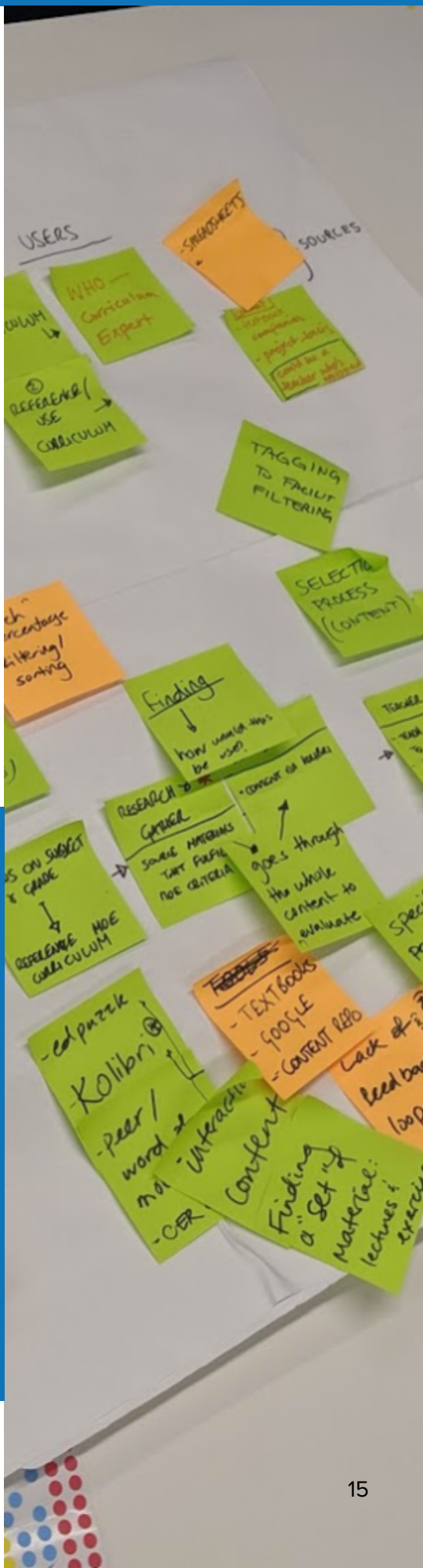
[1] For the purposes of our hackathon, we’ve standardized most curriculum documents to the fields in this example of a template.

APPENDIX 1: FUNCTIONAL GLOSSARY

Relevance - In this context, a level of similarity sufficient that an educator can use the standard or piece of material in question to fulfill a given curricular goal, without modification. For example, “The content of the video is relevant to the learning standard.”

Curriculum mapping - Visually (or otherwise) representing the connections between two parts of a curriculum that match, or between a part of a curriculum and a piece of content. The output of “curricular mapping” often takes the form of a data structure, map, diagram, index, spreadsheet, or re-organization in an interface.

Learning resource - A piece of material (digital or print) that an educator can use to deliver a lesson, whether designed for that purpose or not. For example, a lesson plan, audiovisual material (such as an educational video), a complete lesson in the form of a presentation, an activity, a worksheet, an assessment, a test, a quiz.



It’s important to bring everybody together because we have expertise in different areas...we’re trying to find where the common language lies.

Dr. Wanjira Kinuthia

APPENDIX 2: HACKATHON PARTICIPANTS

The following individuals contributed to the hackathon projects, either in-person (traveling into Google's San Francisco office from Canada, the United Kingdom, Denmark, Switzerland, India, Kenya, and around the United States) or remotely.

- Abby Daniels, Google (volunteer)
- Alan Mwicka, UNHCR
- Armina Foroughi, Google (volunteer)
- Aron Asor, Learning Equality
- Blaine Jester, Learning Equality
- Byaruhanga Nestori, Windle International Kenya
- Caren Watkins, Inclusive Design Research Center at OCAD University
- Duna Tatour, Harvard Graduate School of Education
- Flora Michti, TES
- Francis Kagutha, Windle International Kenya
- Helen Crompton, Old Dominion University
- Ivan Savov, Learning Equality
- Jackie Strecker, UNHCR
- Jamie Alexandre, Learning Equality
- Jason Griffey, Consultant
- Jayson Akilimali Mushagalusa, Windle International Kenya
- Jordan Yoshihara, Learning Equality
- Josephat Ewaat, Windle International Kenya
- Kevin Ollivier, Learning Equality
- Kiza Mauridi, Instant Network Schools
- Lauren Lichtman, Learning Equality
- Lisa Lee, Case by Case Consulting
- Meghan Kellner, UNHCR
- Micah Fitch, Learning Equality
- Mohamud Hure, UNHCR
- Nicole Smith, The Hive, USA for UNHCR
- Nishant Baghel, Pratham Education
- Olly Farshi, Outside
- Pablo Duboue, Consultant
- Regina Nguyen, Google (volunteer)
- Richard Tibbles, Learning Equality
- Ruben Santa, Google (volunteer)
- Sarah Loos, Google (volunteer)
- Shivi Chandra, Learning Equality
- Solange Lalonde, Consultant
- Tabitha Yong, Google
- Tyler Hou, Google
- Vikas Ramachandra, Onward Assist
- Wanjira Kinuthia, Consultant
- Wendy Yang, Google (volunteer)