CAMBRIDGE
Mathematics

Methodology

Formative evaluation

Others in this series

- 📄 Building the research base
- 📄 Glossary App

Author

Ellen Jameson

Representing the work of

Lynn Fortin, Tabitha Gould, Rachael Horsman, Ellen Jameson, Vinay Kathotia, Ray Knight, Lynne McClure, Darren Macey, Dominika Majewska, Nicky Rushton, Lucy Rycroft-Smith and Ben Stevens.

Methodology: Formative evaluation

Related design principles

- Collaboration and consultation

- Transparency

- Research-informed

Sections in this document

Introduction

Formative evaluation of the Cambridge Mathematics Framework

A design can be evaluated according to:

1. The relation of design goals and principles to the background and context of the design, including the problem the goals are intended to address and the theoretical and empirical justification for the design principles,

2. Whether the design is a faithful expression of those design principles, and

3. Whether the design goals can be met when the design is put to use.

This paper describes our formative evaluation efforts with respect to all three points.

We want the Cambridge Mathematics Framework to be meaningful, useful and used. Consequently, we are gathering evidence that can inform our design efforts while the core of the Framework is under active development, during which time this evidence can have the greatest impact on the overall design. To make *formative* evaluation possible, we produce expressions and extensions of the Framework design which can be shared with appropriate audiences while work is in progress. These include reporting of ongoing evaluation efforts so that potential users of the Framework will have the opportunity to understand what informs and influences it.

The overall rationale for our formative evaluation approach depends on the nature of the design itself, which is described in the Framework Update and will be detailed more fully in other papers. In summary, the Framework is intended to serve as an aid to teaching and designing in support of coherent, connected mathematics learning experiences for students. The experiences that students have are influenced by the coordination (or lack thereof) of the different perspectives and knowledge of people in a chain of related but distinct professions in mathematics education. Our design is intended to support coordination of perspectives and knowledge among and between members of these professional communities. This aim shapes the focus of our evaluation methods around inclusion of the viewpoints of representatives of multiple professional communities, seeking feedback from certain communities at certain times. The rationale for specific methods is presented as they are discussed.

Value and trustworthiness: Our goals for formative evaluation

Trustworthiness is what would enable the design of the Framework to be meaningful. The Framework will have *value*, relative to our goals for the design, if it is useful and used. Our perspectives on value and trustworthiness, explained in detail in Methodology: Building the Research Base, are the reason that formative evaluation is so central to our design efforts. We gather data that helps us to evaluate whether we are developing our content in a way that is likely to contribute to meaningful and productive engagement with mathematical ideas among users of the Framework.

Guiding questions

Our guiding questions for formative evaluation are specific to particular elements of the design effort, shown in Table 1.

Table 1: Guiding questions aligned to elements of the Cambridge Mathematics Framework design effort

Framework design element	Guiding questions
Aims for impact (described in the Framework Update)	Are there needs this work might fill? What should we prioritise?
Theoretical and practical influences, design goals, design principles	Are we building on solid foundations? Are we faithful to the meaning of the mathematics?
Understanding of curriculum design, educational design, teaching	Are we anticipating what is needed for the design to be useful?
Creation of tools and interfaces for end users	Are we anticipating what is needed for users to achieve their goals?

Each of these guiding questions cuts across multiple evaluation efforts and is focused appropriately according to the specific participants or audiences in each form of evaluation described below.

Signals and indicators of the value of the Framework

The Cambridge Mathematics Framework is within its development phase, and we are continuing to write its core content. In this phase we are seemingly faced with a paradox: it is essential for us to receive and act on feedback in all stages of development, but gathering data on outcomes or impact which might depend on the Framework as a connected whole will not be possible until after it is released and implemented in context. Moreover, while such data might provide a general picture of impact, it could be too far removed from specific aspects of structure or content to have implications which inform the refinement of the Framework.

Preliminary signals and indicators of the value of the Framework are therefore the main formative evidence we can obtain during initial development. In our context, a *signal* is something that members of the mathematics education community can observe about our project – whether in the content, the design itself, our process, goals, influences, etc. – which they might then interpret as a sign of quality or value. Likewise, an *indicator* is some response from them that we can interpret as feedback about their judgment of quality or value. This response might be direct, as in an interview or a survey, or indirect in the form of community engagement and critique of project outputs.

Wenger, Trayner, & de Laat (2011) suggest an approach to design project evaluation that looks for value to potential users in the appropriate place depending on how far along the project is – from the point where the design is hypothetical, through when it starts to produce artefacts, to when those artefacts can be implemented and their value directly experienced. In each stage, peer debriefing (review by external researchers), feedback from community members, and formal indicators of value are expected to provide complementary perspectives.

Signals are often used as proxy cues for judging quality in a wide variety of situations when quality cannot, or cannot yet, be directly observed or tested. There is evidence to suggest that the trust that potential users place in a database of knowledge, on the basis of these signals, affects the likelihood that they will choose to use it in their work or to spend their time making a contribution to it (Durcikova & Gray, 2009). Of course, whether the trust inspired in potential users is warranted – that is, whether the database of knowledge *would be* of value to them – depends on whether it sends reliable signals of its quality.

Table 2: Alignment of signals, audiences and indicators.

Audiences		Signals they encounter	Indicators we observe
Targeted audience	Internal reviewers	Research summaries Framework content Framework ontology report and other internal reports Research white papers Peer-reviewed research papers Published professional articles	Level of agreement Direct dialogue and negotiation of meaning
	External reviewers	Research Summaries Framework content	Data from semi-structured survey
	Delphi panel members	Framework ontology report Research Summaries Framework content (optional) Delphi round reports	Data from a series of semi-structured surveys with negotiation of meaning
	Collaborators	Research summaries Framework content	Direct dialogue
	Pilot test case partners	Research summaries Framework content	Semi-structured survey Direct dialogue
	Glossary survey participants	Sets of published definitions of key mathematical terms	Structured survey
	Pilot UX interview participants	None, except possible prior general audience knowledge	Semi-structured interview
General audience	Conference participants	Conference presentations Demonstrations of Framework design tools and content Espressos Conference tweets	Direct dialogue Agreeing to make a more active commitment in a category above Sharing our conference tweets Blogging about our presentations
	Website and newsletter readers Social media followers	Espressos Blogs Announcements of published professional articles Drafts of peer-reviewed papers Research white papers About Us page Events page Social media posts	Viewing our documents (for how long per view; how far do they read) Sharing our documents Blogging or tweeting about materials on the website

Not all signals or indicators are equally strong or good predictors of meaningful quality, nor of the same aspects of quality. The ethicist Onora O'Neill (2018) suggests a deceptively simple list of steps to take in order to earn meaningful trust: (1) "be trustworthy both in word and in deed" (p. 295), and (2) provide good evidence of trustworthiness. If we do everything we can to make sure that the signals we broadcast are an authentic representation of our content, processes, perspectives and intentions, for informed audiences whose judgments of quality are relevant and reasonable, then the feedback we receive from them will be able to inform and improve the meaningfulness and usefulness of the Framework.

What is good evidence of trustworthiness in our context, and for whom? Different audiences have different, partially overlapping areas of professional expertise, and we have developed our formative evaluation strategy to align signals, audience and indicators appropriately. Table 2 shows the signals that we consider to be good evidence of trustworthiness. It is readily apparent in the table that multiple audiences are engaging with some of the same signals in different contexts, and that we are getting different kinds of feedback as a result, each kind making a particular contribution to design.

Audiences

Our audiences are people in several different roles. For the Framework to serve as a shared representation of knowledge we need to treat feedback from each role as an important piece of the picture that the whole Framework might need to provide.

Some forms of feedback require an audience to engage more extensively with details of the project. In these cases, we use methods and materials for working with targeted audiences – those with whom we can directly arrange the necessary commitment of time and who can be chosen from among a variety of backgrounds and contexts to provide multiple perspectives. The following are our targeted audiences:

- **Internal reviewers** are members of the Cambridge Mathematics design team, primarily the Framework writing team.

  ◦ *Role*: All have more than one role/background, which helps them to act as knowledge brokers. Collectively, our past roles include primary and secondary specialist mathematics teaching, mathematics curriculum development, textbook authoring, task design, assessment and educational research.

  ◦ *Analysis*: Internal reviewers have invested time in developing a shared sense of appropriate use of the literature and structure of the Framework. Their expert heuristics apply but they will likely have blind spots.

- **External reviewers** have appropriate backgrounds for evaluating the selection and use of research in Research Summaries. They are usually academic researchers who also typically have experience as teacher educators, and sometimes also as classroom teachers.

- **Delphi Panel members** are senior academic researchers whose work has focused on relevant aspects of mathematics curricula and who have played national- or international-level roles in curriculum design.

- **Collaborators** are academic researchers in mathematics education, national-level curriculum developers and mathematicians. They also have experience with educational design projects, teacher education/professional development, and/or classroom teaching.

- **Pilot test case partners** are other teams, within and beyond Cambridge, that are doing work which allows them to trial specific uses of the Framework in an authentic context. Members of these teams are often curriculum, resource or assessment designers.

- **Glossary survey participants** will be mathematics teachers, teacher educators and/or educational designers who agree to provide some basic information about their areas of professional experience.

- **Pilot UX interview participants** are teachers, teacher educators, educational designers, and academic researchers in mathematics education, many of whom have had more than one of these roles.

General audiences might encounter us, our website materials or our social media posts at any time. They could be anyone, but they are most likely to be those who are either actively searching for the kind of information that we publish on our website, or who have actively connected with us through subscribing to our newsletter, direct following on social media, attending one of our conference sessions or encountering our published content through social networks. The following are our general audiences:

- **Conference participants** have professional roles in mathematics education, and may be academic researchers, teachers or teacher educators depending on the conference. We report internally on the discussions we've had and with whom, and follow up with those who have expressed interest in further collaboration.

- **Website and newsletter readers** are the audience we have the least direct contact with, though there may be brief interactions through comments. For the most part what we can see are viewing statistics indicating engagement: the number of people who have visited specific pages on the website (usually Espressos, blogs and research documents), how long they've spent on each page, whether they scrolled to the end of the page, and whether they've downloaded documents.

- **Social media followers** have subscribed to be notified when we post on a particular platform. We often post to publicise new material on our website, or conference presentations by our team or others. Our largest following is currently on Twitter, which provides a range of statistics for audience engagement.

Signals

Our documents and activities which serve as signals for a general audience are meant to have the lowest barriers to engagement possible so that we can (a) reach a larger number of people, and (b) obtain feedback from teachers and teacher educators engaged with our presentation of research. This feedback might be indirect; for example, hearing that our Espressos are being independently shared and used in professional development. In contrast, the documents and activities which serve as signals for a targeted audience involve engaging in more detail with the aims, goals, approaches, and outputs of the Framework project itself. This requires a greater time commitment and often a background in working with the research sources which form the basis for Framework writing, and/or the design methods which form the basis for our methodology.

- **Research Summaries** are short papers which explain how research has informed the structure and content of a section of the Framework. They tell the story of the mathematical ideas and relationships represented in the content. Research Summaries are not yet available to a general audience; we give them to invited reviewers, collaborators and pilot case partners for direct evaluation and feedback.

- **Framework content** mainly refers to the names, descriptions and examples given in waypoints, the names of themes, and how waypoints are connected by themes. An overview of these features of the Framework is available in An Update on the Cambridge Mathematics Framework. Framework content is also not yet available to a general audience, but is made available to the same audience described for Research Summaries.

- **Internal reports** are working summary documents that help us to keep track of various aspects of Framework design, including documentation of how we are systematising Framework content, how we interpret feedback for design, and our overall narrative of design influence and changes made over time. Internal reports often form the basis for reports published on the website.

- **Research white papers**, like this one, detail our perspectives, approaches, justifications, methods, and specific issues that we have engaged with in our work. They are published on our website and publicised through our regular newsletter and on social media. While they are available to anyone, they represent part of our ongoing conversations with researchers and speak most directly to that audience.

- **Peer-reviewed research papers** are similar to white papers but are more focused on specific aspects of our work which we feel could be of interest to particular academic communities and which could benefit from peer review feedback. So far these have mainly been papers for conference proceedings, but we now have data from formative evaluation that could be productive to report in journals.

- **Published professional articles** appear in widely-read publications with an audience of teachers, school administrators, teacher educators, and educational researchers. They focus on current issues in mathematics education which intersect with our team's professional development activities, Framework writing or current collaborations.

- **Delphi round reports** were read by Delphi panel members as part of a recently concluded Delphi study, summarised below, with more detail given in a later final report. These reports provide a summary and analysis of opinions and points raised by the panel after each survey round.

- **Published definitions of key mathematical terms** are definitions that may later be offered in the Framework glossary. The definitions we are evaluating in this case have already been published in leading dictionaries and glossaries for mathematics education.

- **Conference presentations** might be for academic or teacher audiences, or a mix of the two. Conversations at conferences help us to connect with potential collaborators, become better known in relevant communities, and develop ways of communicating more effectively about the Framework.

- **Demonstrations of Framework design tools and content** provide a first impression of the content, structure and design tools for audiences at conferences and meetings who have not encountered it by other means.

- **Espressos** are published monthly for an audience of teachers and teacher educators and are freely available on our website. They are two-page discussions of research and its implications for mathematics teaching and learning. Topics include those that audiences have asked about or are currently debating, and intersections with our work on the Framework. Writing Espressos is sometimes a way for the team to work more closely with collaborators from specific roles or contexts.

- **Blogs** are wide-ranging, informal and brief discussions of topics in mathematics education we are currently engaged with in some way. These may or may not be directly related to the Framework. Blogs are published weekly, and are meant to reach the widest casual audience, who may then decide to read more or engage with the project in other ways.

- **About Us** is a page on our website which, among other things, provides brief biographies of our team members to help our audiences build their own picture of our perspectives, backgrounds and qualifications.

- **Events** is a page on our website which shows events we've hosted, presented at or attended, as well as other events we wish to highlight, giving an indication of the mathematics education communities we're involved in, where, and how actively.

Indicators

Indicators from our targeted audiences often include detailed feedback in direct response to questions we have developed based on our internal discussions, research, and feedback. In the case of survey data and pilot test case observations, this can require detailed analysis on our part to determine the implications of feedback for design and writing. Indicators from our general audiences may include some detail pertaining to structure and content, but mainly involves a scale of engagement, where increasing engagement indicates to us that we are doing something of relevance to mathematics education communities. In turn, those who find our work most relevant to their own may choose to become more directly involved.

Formative evaluation methods

In this section we describe the types of formative evaluation we do, the feedback we receive as data, and how each type of evaluation is integrated into our design process (Fig. 1). Each design method, shown in red, is currently happening as the writing team continues to refine the structure, produce content and prepare for implementation. Design features and design projects, shown in dark blue, continue to be planned and produced. When we receive feedback that informs any of these methods, features or projects, we can make adjustments which help us revise our current and planned work.

Figure 1 on next page

Figure 1: Evaluation in the Cambridge Mathematics Framework design process
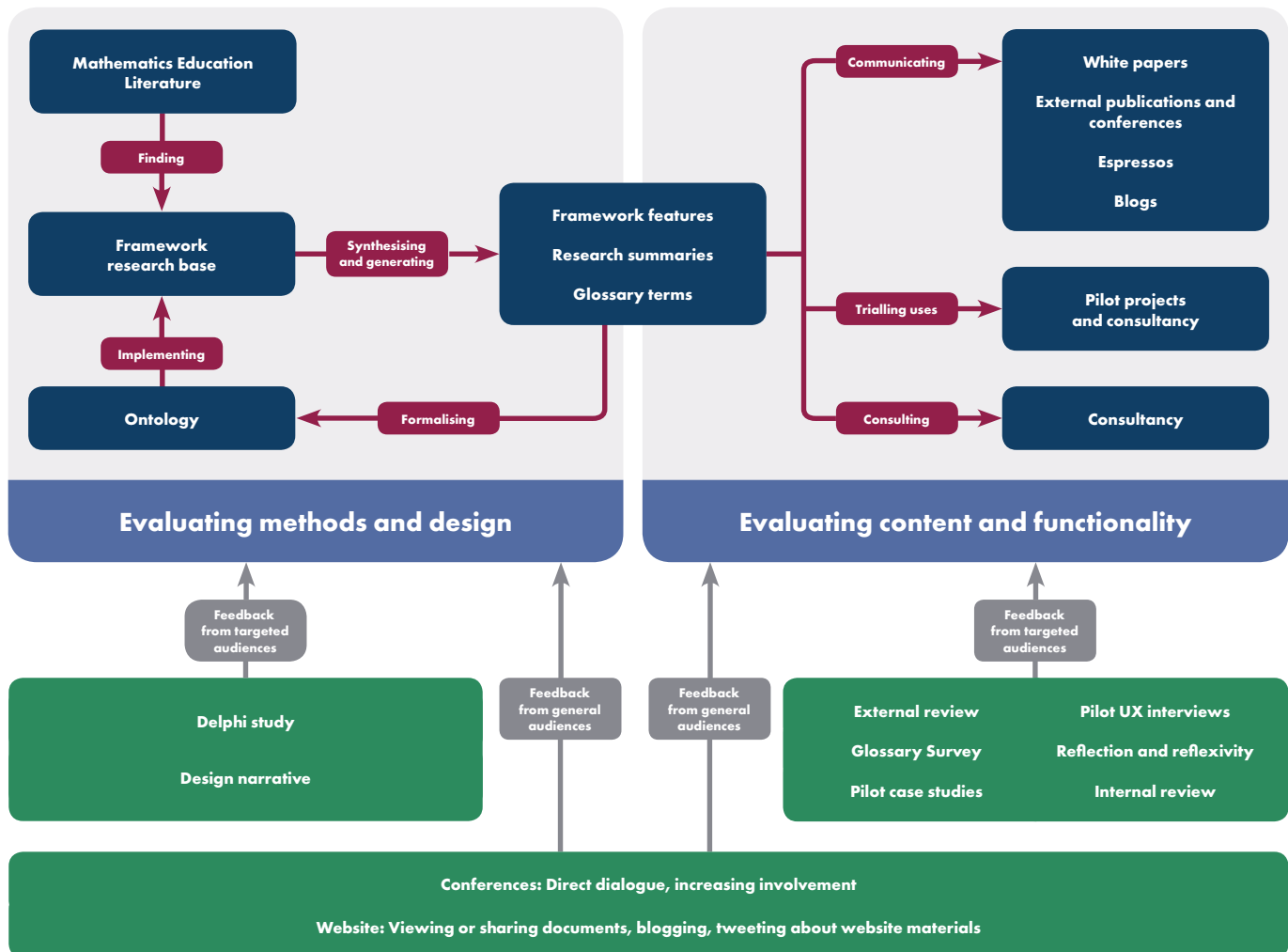
Evaluating methods and design

Decisions about our research and design methods and the *ontology*, or guiding structure of the design of the Framework, fundamentally shape what is explicitly represented in its content and structure, and how trustworthy and useful that form of representation can be. We began by informing these decisions with research and the experience of the team, but there are areas of meaningful disagreement between researchers, and areas in the literature and in our own experiences where questions relevant to our design have not been specifically addressed. To get external feedback on these elements of Framework design, we have conducted a *Delphi study* with participants whose professional judgment can help to navigate areas of uncertainty.

External: Delphi Study

We were interested in finding external evaluators who had both the experience and the willingness to engage with a set of the more complex and central aspects of Framework design, and who would take each other's opinions seriously as being worthy of response. We also looked for a method that was appropriate for eliciting their feedback and discussion on matters that are more a matter of professional judgment than known solutions. After reviewing the literature on focus group methods, the Delphi method stood out for affording the level of engagement, expertise and professional standing among participants which was required, with the added benefit that we could mediate a group discussion without the impossible task of convening participants from multiple countries and continents in the same place at the same time.

Delphi is a group survey method designed to identify areas of consensus and disagreement among experts around issues for which existing evidence doesn't point to a clear solution (Clayton, 1997). Experts on a Delphi panel fill out a questionnaire, receive an anonymised report of the group's opinions and statements, and then have a chance to react to the group in a new questionnaire. These rounds provide the mechanism for the panel to converge on specific points of agreement and disagreement through mediated discussion. This method and our rationale for using it will be reported in more detail in a forthcoming paper, along with analysis and discussion of results, but a summary is provided here.

We recruited 16 participants for the Delphi panel in Round 1, of whom 13 also participated in Round 2 and 11 in Round 3. These participants were senior researchers and designers of mathematics curricula from six countries whose work has had an impact on the field nationally or internationally. Known drawbacks of the Delphi method applied in our case: the analysis and reporting were time-consuming. However, we completed three rounds of semi-structured questionnaires and reporting, with the last two rounds repeating key questions so that we could look for opinion change and movement towards consensus as a result of the group discussion. The first questionnaire was exploratory and focused on outstanding issues developed into questions by the team. This helped us to identify areas of disagreement that could be particularly fruitful for further discussion. The Round 2 questionnaire posed questions that were developed out of Round 1, and the Round 3 questionnaire gave the panel the chance to express whether their opinion had changed on key questions from Round 2. A report was produced for the panel after each round, and a final report has been produced for team discussion of the implications for design. The results of that discussion will be added to the final report, and a version of this report will be sent to the panel and published on our website.

Internal: Design narrative

The Cambridge Mathematics team is the group of people closest to the Framework design process. We maintain internal notes from literature reviews and meetings, recording major design decisions about the structure and the tools we use to work with it. In addition, research sources are annotated and decisions regarding related sets of waypoints in the framework are explained in Research Summaries describing those waypoints. This serves some of the function of an audit trail, which Lincoln & Guba (1985) describe as a set of records adequate for making the research process transparent to aid the interpretation of outcomes. In design research, this includes being able to describe how and why the thing being designed came to be the way it is, and perhaps how it might best be developed further (McKenney & Reeves, 2012). The design narrative informs our internal work and our communication to wider audiences about the Framework and our methodology.

Evaluating content and use

Internally, we examine and develop our own ways of working and thinking, and the writing team reviews Framework content among themselves as it is written. This provides a good starting point for the process of refining content because the members of the writing team collectively have backgrounds like those of knowledge brokers – that is, they have each played multiple roles in mathematics education, including teaching, teacher professional development, instructional design, curriculum design, assessment design, and educational research. From this starting point, we can bring Framework content to specific audiences recruited to participate in our external feedback processes, including external review of content, pilot test cases which let us implement parts of the Framework, and interviews with members of professional communities in education. For a wider audience, we present at conferences and publish white papers, blogs and research digests for teachers called *Espressos*.

Reflection and reflexivity

Reflexivity is a way of examining our own thinking about the content and structure of the Framework. It is one method of confirmability, the ability to trace whether links between data and researchers' interpretations are reasonable, as discussed in Lincoln & Guba (1985) and employed in other research on curriculum frameworks (Cunningham, 2017; Ferrini-Mundy & Martin, 2003) and conceptual representation in mathematics learning (Confrey & Lachance, 2000). In our work, this involves both a shared research process and the reflective practice of each writer when interpreting the literature and creating Framework features. We have developed a shared process for finding and reviewing literature when

writing Framework content and research summaries, described in Methodology: Building the Research Base. Our individual interpretations of research implications for the Framework emerge according to our own perspectives, beliefs, experiences and interpretation of feedback from others. This results in individual contributions to particular approaches and decisions about structure and content. These contributions are a frequent part of our internal discussions.

Internal review of Research Summaries and Framework content

The members of the writing team are most familiar with the emerging norms and practices for Research Summaries and Framework content. We have developed an internal review process which serves as a form of reliability exercise. In this process, a Research Summary must be reviewed by a second member of the writing team according to criteria for the assumptions, connections and supporting arguments for implications in the narrative (relative to the Framework content described in the Research Summary), tone, length, style, formatting of text and references, and links. The original writer(s) of the Research Summary and Framework content will discuss recommendations with the internal reviewer and make adjustments as needed.

External review of Research Summaries and Framework content

Once internal reviews of Research Summaries and accompanying collections of Framework content are complete, they are sent out for external review. External reviewers are academic researchers, often with additional experience in teacher education, who have knowledge of the mathematics education literature in appropriate areas of the domain for the Research Summary they are asked to review. They are recruited individually from among the team's existing contacts, from recommendations, and from the wider set of authors of the research sources the team have been using. External reviewers are volunteers, with the understanding that they will be acknowledged if they wish and that their work will contribute to refining the core of the Framework, which is to be publicly available upon release.

Up to the point of publication (May 2019) this external review has consisted of informal discussions with reviewers. We are now launching our formal external review process. External reviewers access their Research Summaries online. The Framework content to which each Research Summary refers is embedded within it and the details can be accessed on demand, shown in Fig. 2(a). We have developed a semi-structured questionnaire (Fig. 2(b)) which includes the same items for all Research Summaries. The questionnaire asks reviewers to tell us:

- whether any important research sources are missing that we should include,

- whether our interpretation and use of the literature we have included seems reasonable and is appropriately represented in the content,

- how accessible the Research Summary might be for others,

- whether the set of content seems coherent and consistent with the story of the Research Summary, and

- whether the examples of actions students may take in order to have some experience of the content are appropriate.

Reviewers may also add any other comments. Leading up to the development of the formal process, Framework writers had informal review conversations with several collaborators, and the questionnaire was developed on the basis of these together with discussions during the internal review. We developed an additional questionnaire for a short pilot of the formal external review process so that we would be able to improve it if indicated before launching it in full. Framework writers take reviewer responses into account when adjusting the content of the Framework and the narratives of the research summaries.

Figure 2: (a) Online Research Summary with dynamic Framework content visualisation window; (b) questionnaire

(a)

(b)

Pilot case studies

Specific uses of the Framework are piloted so that we can identify features that need to be refined or added before release. Feedback from these pilot test cases can help us to identify the key goals and strategies of Framework users, the ways in which they feel the Framework adds to their process, and specific feedback and suggestions related to problems that they encountered and processes they developed when working with the Framework. Currently, we have observed cases in which the Framework was used in curriculum design, textbook authoring, assessment item design and mapping to a segment of a curriculum. Cases of mapping to tasks and assessment items are underway, along with cases involving curriculum refinement and sequencing, and professional development. We will be reporting the results of selected pilot test cases on our website when they have concluded.

Pilot UX interviews

The roles listed in Table 2 have been further characterised by data from interviews conducted by the Cambridge Assessment BTS Innovate UX (user experience) team. These roles had emerged as a primary focus from preliminary data gathered from the Cambridge Mathematics team through card sorting, an activity which helps people to identify and build a shared model of the relationships between important considerations, and group discussion. The pilot interviews helped us build a preliminary set of potential use cases and a sense of what kinds of differences in the use of the Framework might be expected, either between individuals in the same role or different roles. This work has primarily informed our pilot cases so far, and may be the basis for additional interviews based on outcomes from the pilot case studies.

Glossary term definition survey

Shared language is important for effective communication (Akkerman & Bakker, 2011; Stahl, 2006). The Framework glossary is being populated with definitions of key mathematical terms to provide the most essential support for shared language, and we would like to be able to offer a set of definitions which will be meaningful and useful for people in different roles. We want to develop a sense of which existing definitions might be useful to consider for our glossary, why, and for whom.

To gather data which helps us to do this, we are recruiting mathematics teachers and educational designers to participate in a survey exploring what they find useful and/or correct in a set of existing published definitions for some of the terms in the Framework glossary. In order to obtain enough feedback to work with, it is important that we match our method of data collection to the availability of participants. We have developed a survey app called *CM Define It* which prompts participants to provide data on one word per week. A diagram of the app's three main screens is shown in Fig. 3. We collect informed consent and data on participants' language, country, main professional role, additional

professional roles, and the range of students they teach (Fig. 3(a)). Each week, the app gives participants a set of definitions for a word and asks them to give a rating from one to five stars for each definition, followed by further ratings of each according to their accuracy, usefulness in their chosen context, and other aspects we are currently trialling. We are aiming to recruit over 100 participants for a limited trial of the app before starting the main survey. More detail on this project's methodology will be available on our website in a forthcoming paper.

Figure 3: A diagram of the main screens in the CM Define It glossary term definition survey app: (a) data on participants' professional context; (b) definition survey; (c) feedback/reward

(a) Registration	(b) Survey: Word of the day	☆☆☆☆☆	(c) Thank you
User data: • First Language • Country • Main role • Additional roles • Range of students	Word +5 alternate definitions; rate for: • Technical accuracy • Emphasis of key points • Accessibility for intended audience • Sufficiency for intended audience • Whether it added to or clarified their own understanding	Free response: Any other feedback	Feedback / reward: • User stats, history

Reporting and integrating feedback

The evaluation feedback from selected participants helps us to gauge and improve the trustworthiness and usefulness of the Framework's structure and content and of our design methods. Feedback from general audience engagement with our presentations and published writing provides additional indication that an audience of potential users recognises signs of usefulness and relevance in our work. Feedback and observations from our pilot case studies give us a sense not only of the usefulness of the content but the usability of the tools that allow users to work with content, and helps shape the direction of our further efforts to identify and support the most essential uses of the Framework.

We regularly meet to discuss the implications of incoming feedback and plan how and when we will be able to incorporate it. A great deal of the earlier feedback we received from conferences, meetings and collaborations has already contributed to the structure, content, tools and processes that make up the Framework project, and contributions from our evaluation studies are ongoing.

Reporting the results and design impacts of formative evaluation back to our audiences is an important part of maintaining transparency and providing the basis for judging the trustworthiness of various aspects of the project. We plan to publish reports of selected formative evaluation studies on our website as the relevant data become available.

References

Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*(2), 132–169.

Clayton, M. J. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology, 17*(4), 373–386. https://doi.org/10.1080/0144341970170401

Confrey, J., & Lachance, A. (2000). Transformative Teaching Experiments through Conjecture-Driven Research Design. In *Handbook of Research Design in Mathematics and Science Education* (pp. 231–265). Mahwah, NJ: Lawrence Erlbaum Associates.

Cunningham, M. (2017). *Method for Developing an International Curriculum and Assessment Framework for Mathematics* (pp. 1–15) [Internal Report]. Retrieved from UNESCO, Institute for Statistics (UIS) website: http://inprogressreflections.ibe-unesco.org/wp-content/uploads/2018/01/Math-Content-Reference-List-and-Coding-Scheme_Methods-Paper.pdf

Durcikova, A., & Gray, P. (2009). How Knowledge Validation Processes Affect Knowledge Contribution. *Journal of Management Information Systems, 25*(4), 81–108. https://doi.org/10.2753/MIS0742-1222250403

Ferrini-Mundy, J., & Martin, W. G. (2003). Using Research in Policy Development: The Case of the National Council of Teachers of Mathematics' Principles and Standards for School Mathematics. In *A Research Companion to Principles and Standards for School Mathematics* (pp. 395–413). National Council of Teachers of Mathematics.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Sage.

McKenney, S., & Reeves, T. C. (2012). *Conducting Educational Design Research*. Routledge.

O'Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies, 26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637

Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge, MA: MIT Press.

Wenger, E., Trayner, B., & de Laat, M. (2011). *Promoting and assessing value creation in communities and networks: a conceptual framework* (Rapport No. 18; p. 56). Retrieved from Ruud de Moor Centrum, Open Universiteit Nederland website: http://wenger-trayner.com/wp-content/uploads/2011/12/11-04-Wenger_Trayner_DeLaat_Value_creation.pdf