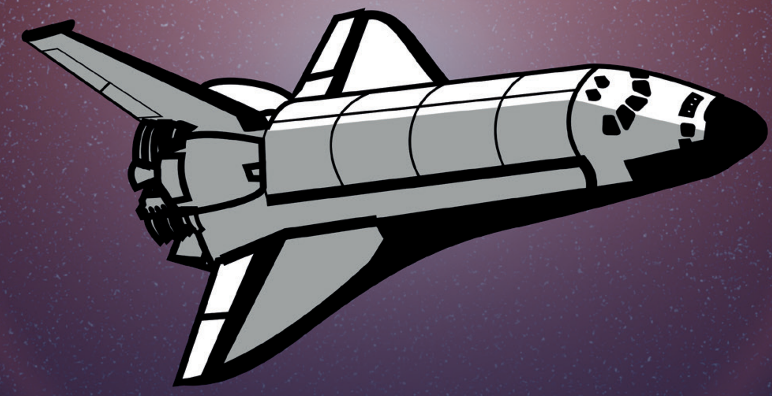
CAMBRIDGE Mathematics

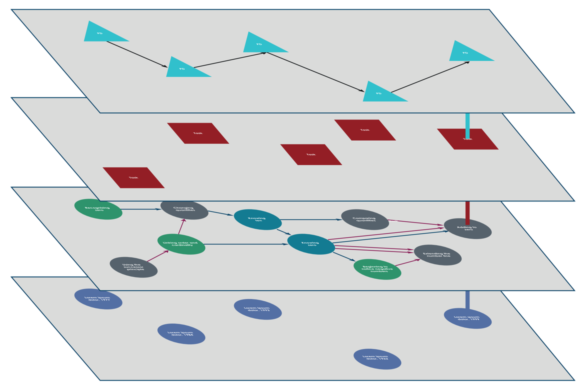
Ground tests and test flights: Refinements from formative evaluation of the Cambridge Mathematics Framework

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



1. Introduction

The Cambridge Mathematics Framework is a research-informed, digital, connected framework of mathematics learning for 3-19 year olds, built using a graph database.



Since one of the aims of our work is to use good quality evidence from research and expertise, we evaluate our content, structure and methods throughout the writing of the Framework. This poster focuses on the latter.

How might Cambridge Mathematics:

- obtain feedback from expert opinion to inform design, and evaluate structure and methods;
- use this evaluation to begin to refine the structure, ontology and products of the Framework?

2. Methods

To make sure our approach is as transparent, effective and useful as possible, we have incorporated traits which McKenney and Reeves (2012) classify as being associated with design research:

- using analysis and formative feedback according to frameworks derived from literature review;
- formally documenting our aims, goals, methods and results for a wide audience; and
- *seeking to learn from other design efforts with similar goals and/or contexts.*

As part of this wider design intention, we have used many kinds of formative evaluation: consideration of our work against existing frameworks; feedback on a glossary app pilot; informal expert advice; piloting a formal External Review process; a Delphi panel exercise; a series of user experience (UX) interviews and several case studies. We consider one of these – the Delphi panel – in greater detail, in order to illustrate some key features of our approach.

The Delphi method is an exploratory and advisory structured group survey method designed to identify areas of consensus and disagreement, particularly for ambiguous or hard to pin down issues (Clayton, 1997). We completed a Delphi panel evaluation in three rounds during 2018-2019. We recruited 16 international experts in mathematics education for the study. We used six-point Likert-type rating items in our questionnaire and after each round participants were sent summaries of anonymous responses for the previous round. Where consensus had not been achieved, or for areas of particular focus for the team, the questions were refined.

Although there was some attrition over three rounds, we remained within the panel size commonly recommended for the Delphi method (eight people): Round 1 n=16; Round 2 n=13; Round 3 n=11.

Based on the goals and context of this study, our criteria for examining the range of opinion were:

- consensus was indicated if 70% of ratings fell into one category and/or 80% fell into two adjacent categories;
- when consensus was indicated, the amount of support that consensus showed for the statement being rated was expressed by the value of the measure of central tendency, which we report as the median. Support was categorised as strong support, moderate support, weak support, weak opposition, moderate opposition, strong opposition (de Loe, 1995, p. 62).

3. Results

The Delphi study collected descriptive statistics for exploratory analysis so that the Cambridge Maths team could prioritise the implications of this feedback for our writing and design decisions. The small sample sizes in each round mean there is no statistical power: we are using statistics as a way of giving ourselves an overview of distributions of opinions in items relative to one another within Round 3 and between Rounds 2 and 3 so that we can decide on the clearest feedback and how to act on it.

Our choice of the Delphi method for this expert study was made because of the way the method's unique affordances aligned with our goals for the study, but the lack of power is an inherent trade-off. The literature on the Delphi method consistently recommends panel sizes which are below the threshold for achieving statistical power for typical within-subject correlations of opinions between rounds. The method is usually chosen for its potential to raise unique points in expert decision-making, to evaluate their importance relative to each other and the specific scenario under consideration, and to identify areas of consensus rather than to draw absolute or generalisable conclusions about the population to which panel participants belong (e.g. mathematics education researchers).

THE FOCUS OF EACH ROUND IN THE DELPHI STUDY

ROUND	FOCUS
1	Evaluating theoretical influences, starting assumptions, and general design choices for structuring and representing mathematics content in the Cambridge Mathematics Framework
2	Managing complexity in the Framework and research methods for reviewing the literature and evaluating Framework content
3	The nature of big-picture mathematical thinking and doing, what progression in this big picture there might be, and how it might be usefully represented

There was a great deal of consensus and much of the response analysis was confirmatory, supporting our priorities and aims whilst noting that implementation might uncover unanticipated issues.

One widely shared opinion from Round 1 was that the Cambridge Mathematics Framework should facilitate understanding and working with a big-picture view of mathematical thinking and doing, and hence the design should support this aim. Users' attention is finite and we will need to try to focus on the most important things to convey about mathematical thinking and doing and on supporting common understanding of what they mean. We should do this in a way that remains approachable, describing what is necessary when necessary, and using examples.

Some panel members indicated that they saw elements like abstraction as important, but requiring a different sort of treatment to other elements. Their suggestions are being included in our design discussions. There was a wider range of opinion than expected in the importance panel members assigned to a few elements of mathematical thinking. They disagreed particularly about motivation, with little change of opinion between rounds 2 and 3. The free response comments revealed that some interpreted it as tied directly to conceptual understanding whilst others considered it purely affective. This disagreement has prompted us to review additional sources and has contributed to our design discussions, particularly about support for teacher development.

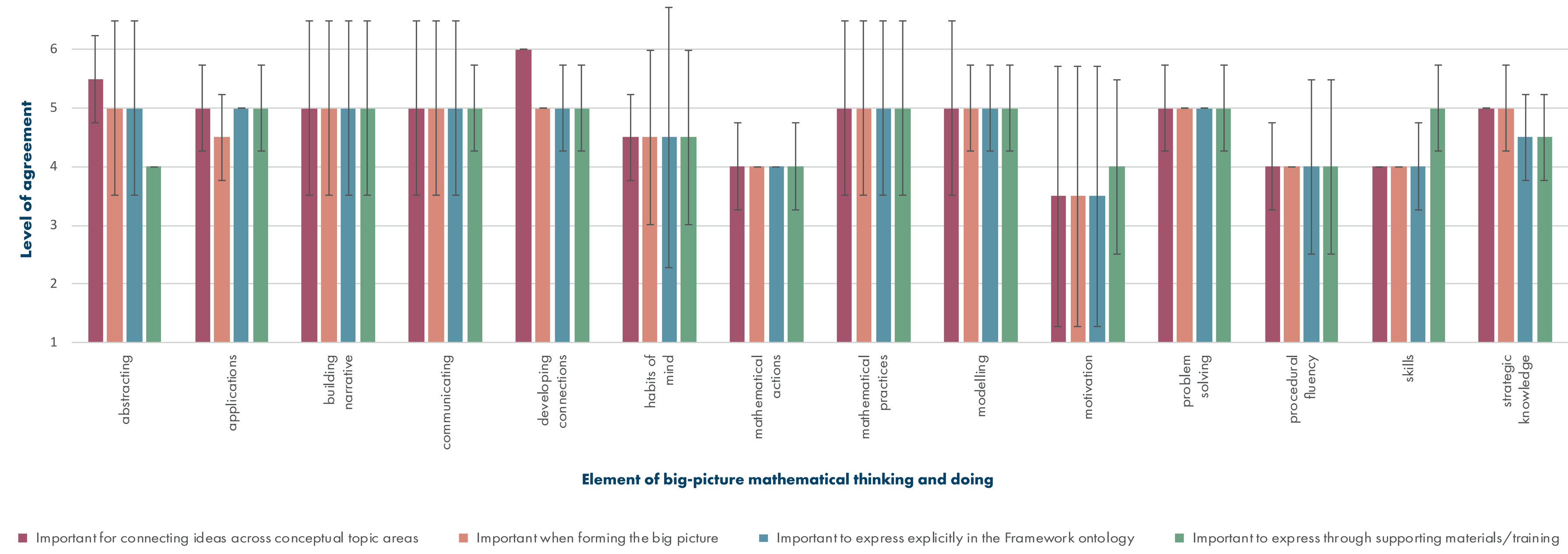
Consensus ratings of elements of mathematical thinking and doing, ordered by importance for connecting across topic areas: strongly agree (6) to strongly disagree (1)

Element	Median, MAD	Difference R2 to 3	Spearman's rank correlation	Round consensus first reached
Develop, exercise, strengthen connections	6 ± 0	1	$\rho(9) = .76$, $p < .02$ two-tailed	**3**
Conceptual understanding	6 ± 0	n/a	n/a	**2**
Reasoning	5 ± 0	n/a	n/a	**2**
Pervasive themes/big ideas	5 ± 0	n/a	n/a	**2**
Strategic knowledge	5 ± 0	1	$\rho(9) = .75$, $p < .02$ two-tailed	**3**
Communicating	5 ± 1.48	0	$\rho(9) = .54$, n.s.	**2**
Modelling	5 ± 1.48	0	$\rho(9) = .47$, n.s.	**3**
Applications	5 ± 0.74	0	$\rho(9) = .49$, n.s.	**2**
Skills	4 ± 0	0	$\rho(9) = .54$, n.s.	**2**
Procedural fluency	4 ± 0.74	0	$\rho(9) = .25$, n.s.	**2**

Design implications from the Delphi study apply largely to the waypoints layer, with secondary implications for the planned layers of tasks and teacher support. These include:

- ensuring that Research Summaries develop and highlight a meaningful perspective on the connections between waypoints; i.e. what is developed or leveraged from one to the next, and how this occurs and can be supported;
- treating the top five elements in the list (see table, above) as the most uncontroversially important (according to the panel); this implies addressing these elements explicitly and consistently within Research Summaries as appropriate, but also designing and trialling other ways to make them explicit, tailored to different types of end user; and
- using free-response feedback and sources suggested by panel members to develop our ideas further around the elements in the lower half of this list so that more specific design implications can come into focus.

Round 3 median rating of the importance of elements of mathematical thinking and doing according to the four categories of importance listed in the key*



*error bars represent 1 MAD, expected 50% responses to fall within the bars due to non-normal data

4. Conclusions

The Delphi study complements our other formative evaluation methods because it involved experts who were very senior, international and often interdisciplinary whilst our other methods have been mostly localised and specialised. The Delphi study was a particularly useful addition because it was timely (enough of the Framework had been written to begin the study in 2018) and particularly well-suited to address more generalised and difficult questions about the structure and methods we have used in the Framework, rather than just the content.

Analysis of the results of the Delphi study corroborates some of our initial thinking: that we need to pay attention to 'big picture' ideas and how they are encoded in the Framework, with greater detail being provided about which and how; these may be 'pervasive ideas' within mathematics itself, or elements of mathematical thinking and doing. It has highlighted the need for a clearer vocabulary around them and to define what we mean by these higher-level terms in some detail for those interpreting the Framework. However, some panel members noted that it is not always possible or desirable to clarify everything behind these terms. Leaving room for people to relate to them from their own context, even if this results in some differences in what is understood, might be valuable.

In reconsidering how and where we represent ideas that seem to fit at a higher level than the waypoints layer, we will need to design some overarching perspective. Some types of big ideas such as *Doing mathematics; Rich, challenging tasks; Problems;* and *Modelling* seem to fit into our plans for a Task layer. We are currently discussing the possibility and functionality of 'aggregation layers' based on saved searches/Research Summaries, themes, landmark waypoints, or even higher-level collections.

Acknowledgements

The Cambridge Mathematics team share, consult and work in partnership with knowledgeable colleagues from all over the world, many of them ISDDE fellows. We are extremely grateful for their time and expertise which has informed the development of the Framework thus far. Their contributions have been freely given but all actions arising from them are, of course, the responsibility of the team and the team alone.

Literature cited

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

de Loe, R. C. (1995). Exploring complex policy questions using the policy Delphi: A multi-round, interactive survey method. *Applied Geography, 15*(1), 53–68.

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

Further information

Please see www.cambridgemaths.org for more information.

You can follow us on Twitter @CambridgeMaths

You can also email us at info@cambridgemaths.org if you have a question or comment.

Our partners

Cambridge Assessment

Cambridge University Press

University of Cambridge Faculty of Education

University of Cambridge Faculty of Mathematics