

The Calculus Baseline Assessment:

**A diagnostic tool that aims to capture student voice at scale,
using text analytics and data visualization**

Caroline Junkins (McMaster)

Lindsey Daniels (UBC)

Connor Gregor (McMaster)

James Colliander (Crowdmark and UBC)



Our Context: McMaster University

- McMaster is a medical doctoral public university located in Hamilton, Ontario, Canada.
- We recognize and acknowledge that students of McMaster University meet and learn on the traditional territories of the Mississauga and Haudenosaunee nations, and within the lands protected by the "Dish With One Spoon" wampum, an agreement to peaceably share and care for the resources around the Great Lakes.



Our Context: Undergraduate Calculus Students



UNDERGRADUATE:
86% | 32,174

GRADUATE:
14% | 5,363

○ Full-Time: **96%** | **30,792**
○ Part-Time: **4%** | **1,382**

○ Full-Time: **83%** | **4,435**
○ Part-Time: **17%** | **928**

○ Domestic: **85%** | **27,197**
○ International: **15%** | **4,977**

○ Domestic: **71%** | **3,787**
○ International: **29%** | **1,576**

Each year, 4000-5000 of these students will enrol in “Calculus I”:

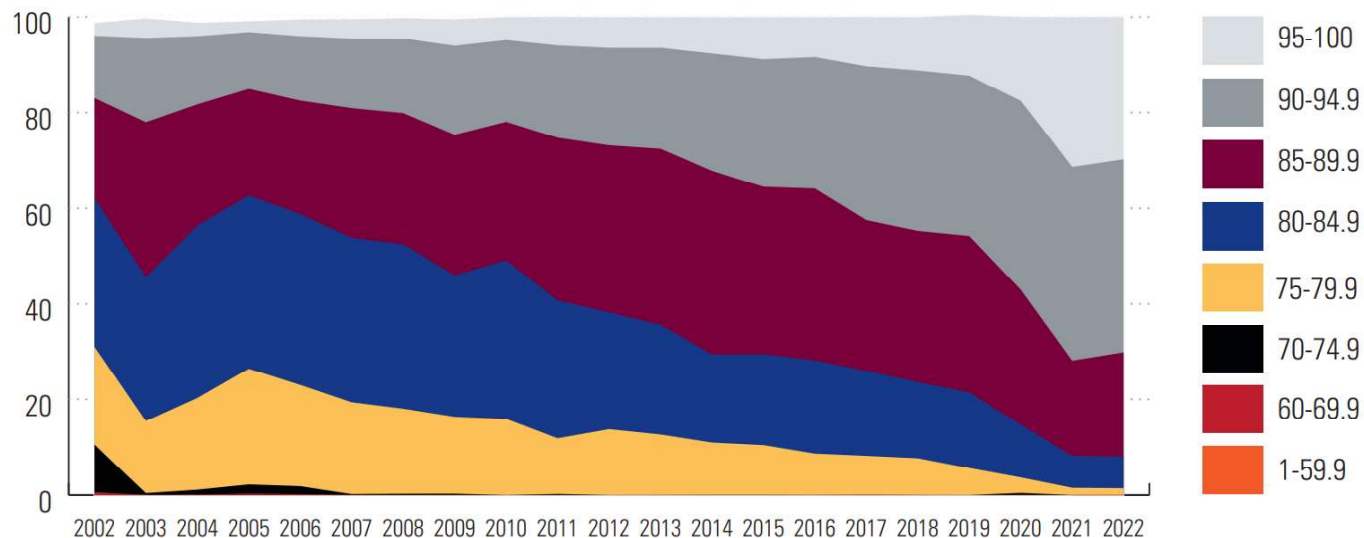
- **Math 1LS3: Calculus for the Life Sciences I (~2000 students/year)**
- Math 1ZA3: Engineering Mathematics I (~1400 students/year)
- **Math 1MM3: Applied Calculus (~900 students/year)**
- Math 1A03: Calculus for Sciences I (~500 students/year)
- Math 1X03: Calculus for Math & Stats I (~300 students/year)

Our Context: Prerequisite Math Courses

For programs including **Science, Business, Economics, Engineering***, admission is determined by completion of a High School Diploma plus grades in six Grade 12 courses including:

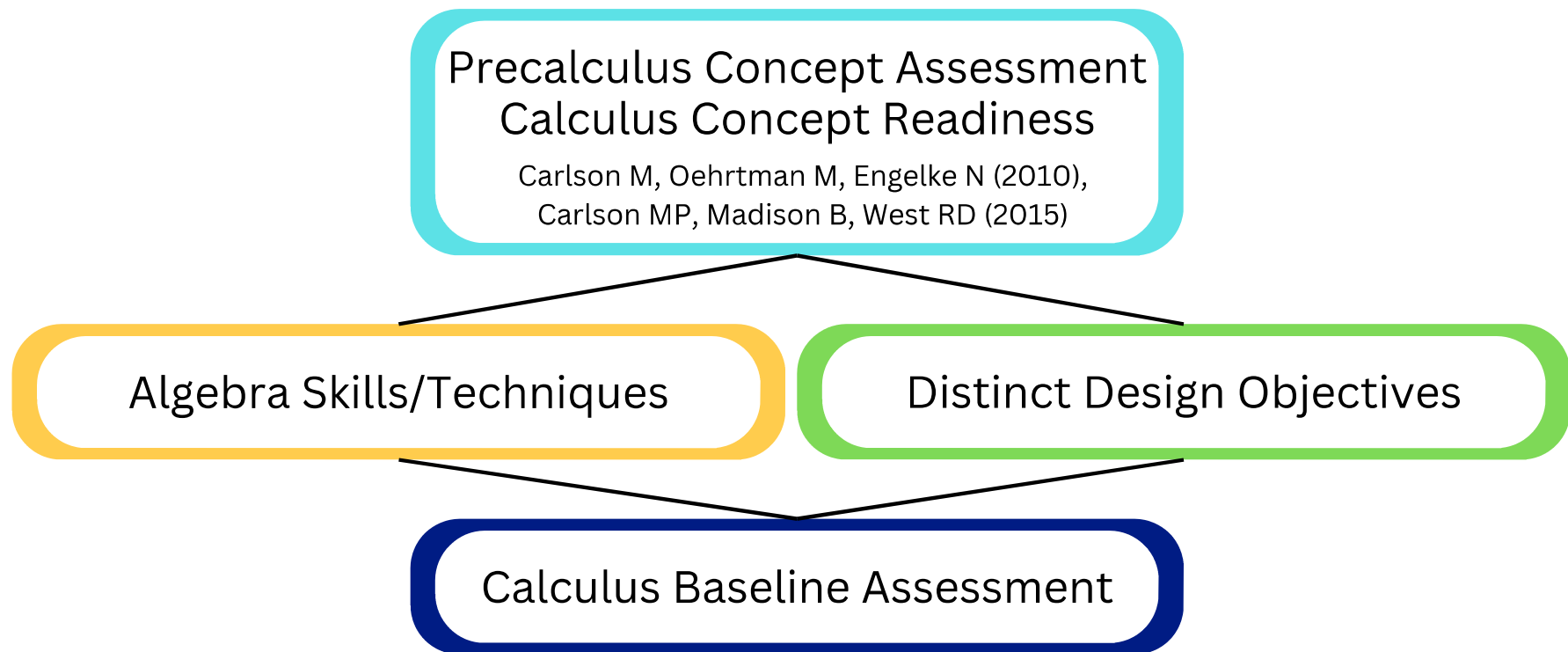
- ENG4U: English
- **MHF4U: Advanced Functions (Precalculus)**
- **MCV4U: Calculus and Vectors (similar to AP Calculus AB)**

Grades of Entering Full-Time Undergraduate Students from Ontario High Schools—University Total, 2002-2022

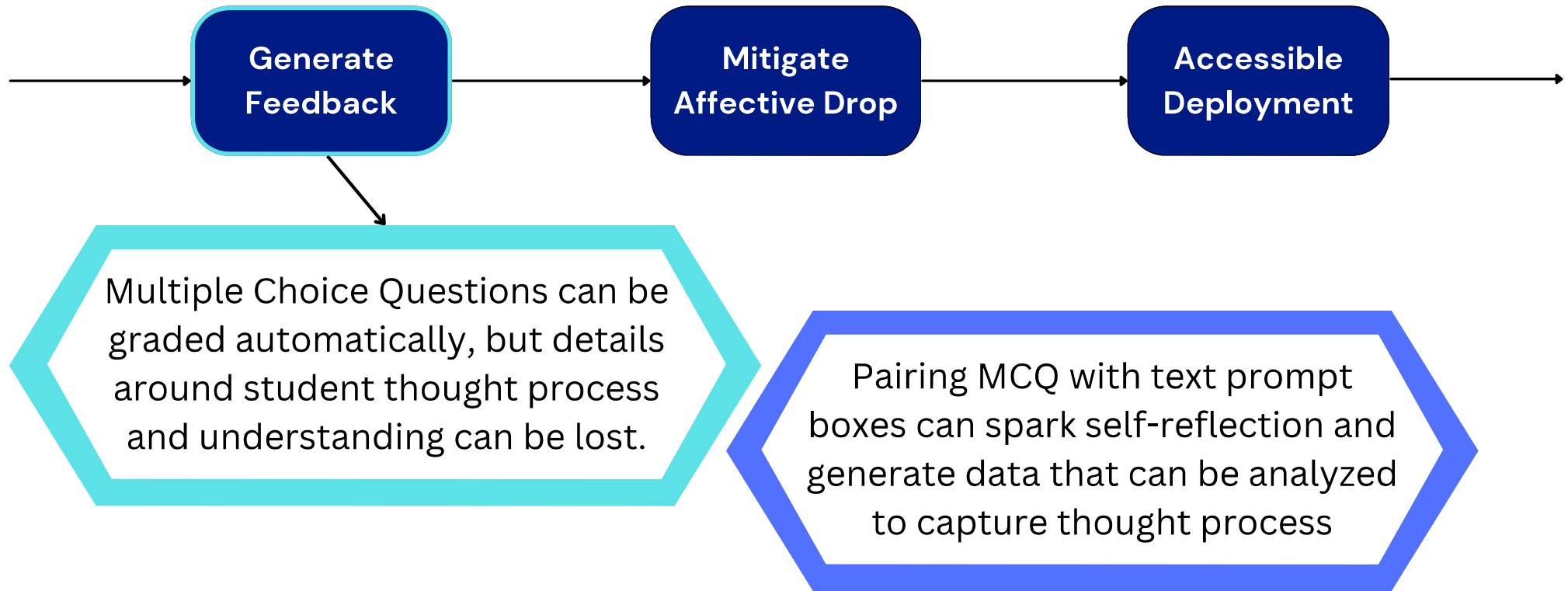


Our Motivation

Develop a tool to measure preparedness for university-level calculus, with an emphasis on providing nuanced student-level and cohort-level information in a scalable way.



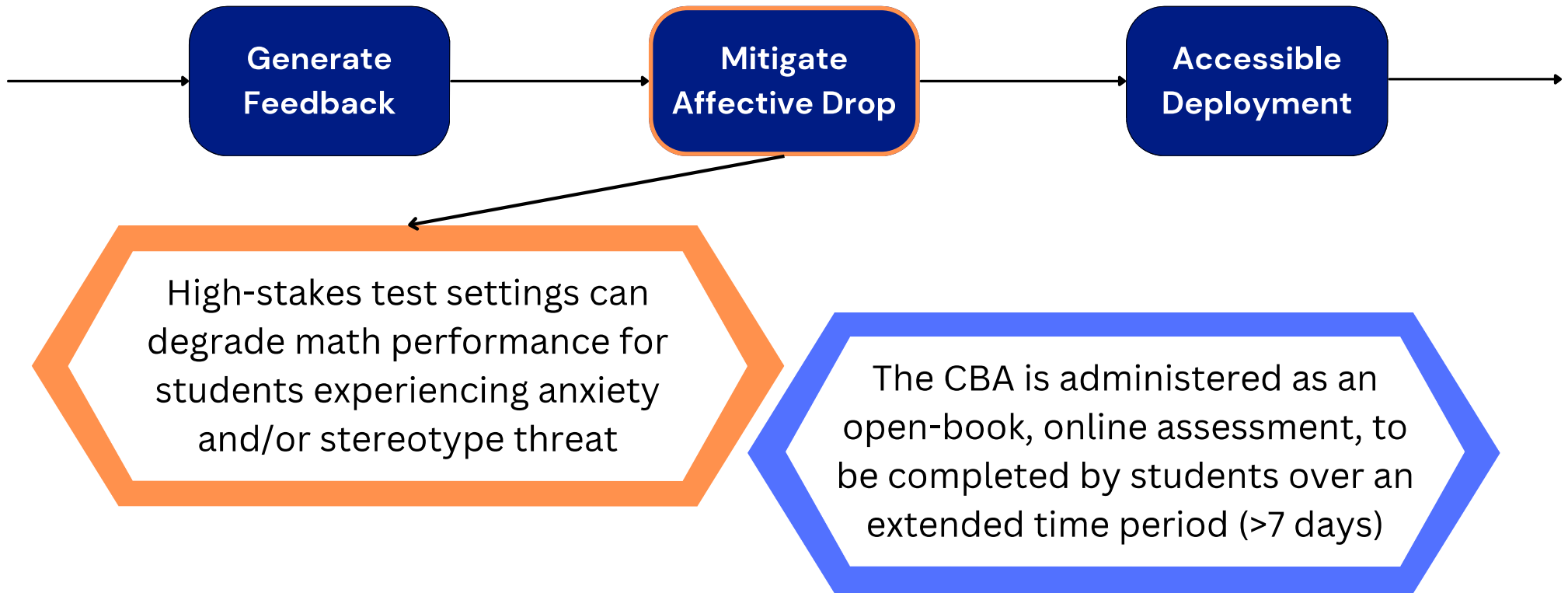
Our Design Objectives



Lombrozo T (2006) The structure and function of explanations. Trends in Cognitive Sciences 10(10):464–470.

Asano Y, Solyst J, Williams JJ (2020) Characterizing and influencing students' tendency to write self-explanations in online homework. In: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. Association for Computing Machinery, New York, NY, USA, LAK '20, pp 448–453.

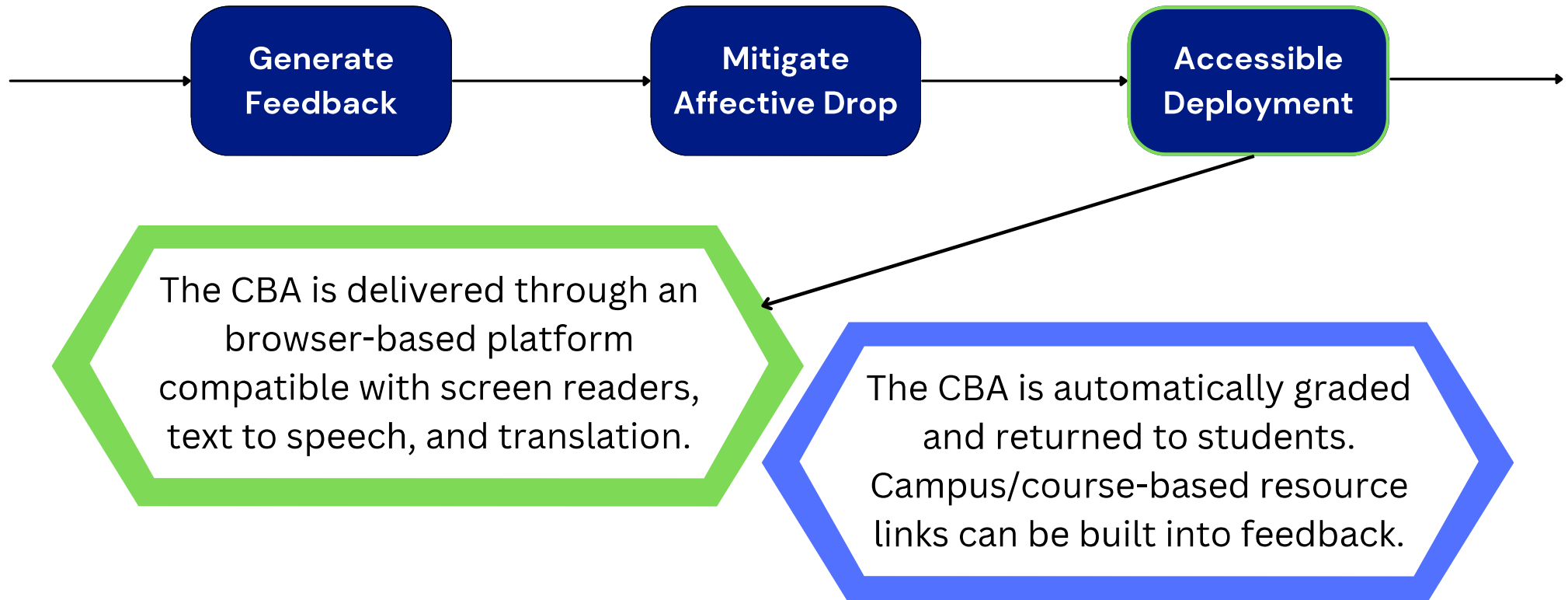
Our Design Objectives



Ashcraft MH, Moore AM (2009) Mathematics Anxiety and the Affective Drop in Performance. *Journal of Psychoeducational Assessment* 27(3):197–205.

Beilock SL, Rydell RJ, McConnell AR (2007) Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General* 136(2):256–276.

Our Design Objectives

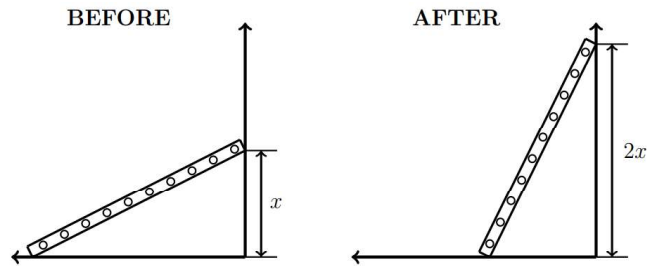


The platform used for this study aims to meet compliance with the Web Content Accessibility Guidelines (WCAG) 2.1 Level AA standards, Caldwell B, Cooper M, Reid LG, et al (2008) Web content accessibility guidelines (wcag) 2.0. WWW Consortium (W3C) 290:1–34

CBA Sample Question

Q13a (1 point)

A ladder - of fixed length - is leaning against a wall. The ladder is adjusted so that the distance of the top of the ladder from the floor is twice as high as it was before it was adjusted.



The slope of the ladder is:

- ☐ Less than twice what it was before
- ☐ Exactly twice what it was before
- ☐ More than twice what it was before
- ☐ The same as what it was before
- ☐ There is not enough information to determine if any of a through d is correct.

Q13b (0 points)

Explain the reasoning for your answer to Q13a in the box below.

[Edit](#) [Preview](#)

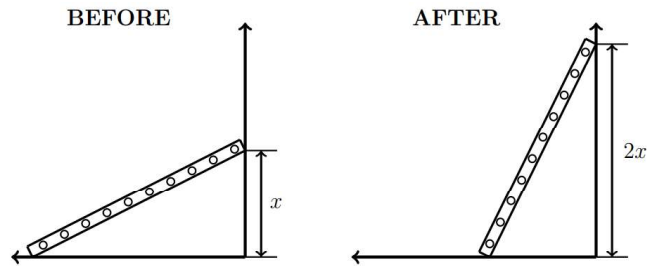
Please enter your response to Q13b

[Attach files](#) [Formatting tips](#)

CBA Sample Question

Q13a (1 point)

A ladder - of fixed length - is leaning against a wall. The ladder is adjusted so that the distance of the top of the ladder from the floor is twice as high as it was before it was adjusted.



The slope of the ladder is:

- ☐ Less than twice what it was before
- ☐ Exactly twice what it was before
- ☒ More than twice what it was before
- ☐ The same as what it was before
- ☐ There is not enough information to determine if any of a through d is correct.

Q13b (0 points)

Explain the reasoning for your answer to Q13a in the box below.

[Edit](#) [Preview](#)

Please enter your response to Q13b

[Attach files](#) [Formatting tips](#)

The slope is equal to rise over run. The rise is being doubled, but since the ladder is fixed length this decreases the run. Because the numerator is doubling and the denominator is decreasing, the fraction will be more than doubled.

*
* Fabricated example, not student data.

Qualitative Coding Procedure

- We construct a qualitative codebook using emergent codes.

The slope is equal to rise over run. The rise is being doubled, but since the ladder is fixed length this decreases the run. Because the numerator is doubling and the denominator is decreasing, the fraction will be more than doubled.

Qualitative Coding Procedure

- We construct a qualitative codebook using emergent codes.
 - Implements mathematical definition(s)

The **slope is equal to rise over run**. The rise is being doubled, but since the ladder is fixed length this decreases the run. Because the numerator is doubling and the denominator is decreasing, the fraction will be more than doubled.

Qualitative Coding Procedure

- We construct a qualitative codebook using emergent codes.
 - Implements mathematical definition(s)
 - Uses details/context provided in question

The slope is equal to rise over run. The rise is being doubled, but since the **ladder is fixed length** this decreases the run. Because the numerator is doubling and the denominator is decreasing, the fraction will be more than doubled.

Qualitative Coding Procedure

- We construct a qualitative codebook using emergent codes.
 - Implements mathematical definition(s)
 - Uses details/context provided in question
 - Compares/describes fractions, percentages, proportions

The slope is equal to rise over run. The rise is being doubled, but since the ladder is fixed length this decreases the run. Because the **numerator is doubling and the denominator is decreasing**, the fraction will be more than doubled.

Qualitative Coding Procedure

- We construct a qualitative codebook using emergent codes.
 - Implements mathematical definition(s)
 - Uses details/context provided in question
 - Compares/describes fractions, percentages, proportions
 - Comprehensive written process

The slope is equal to rise over run. The rise is being doubled, but since the ladder is fixed length this decreases the run. Because the numerator is doubling and the denominator is decreasing, the fraction will be more than doubled.

Qualitative Coding Procedure

- Initial codes are bundled into broader themes that apply across questions.

Algebra Skills

Algebra Traps

Math Relationship Skills

Misconceptions/Interpretations

Solution Framework

Content/Knowledge Gaps

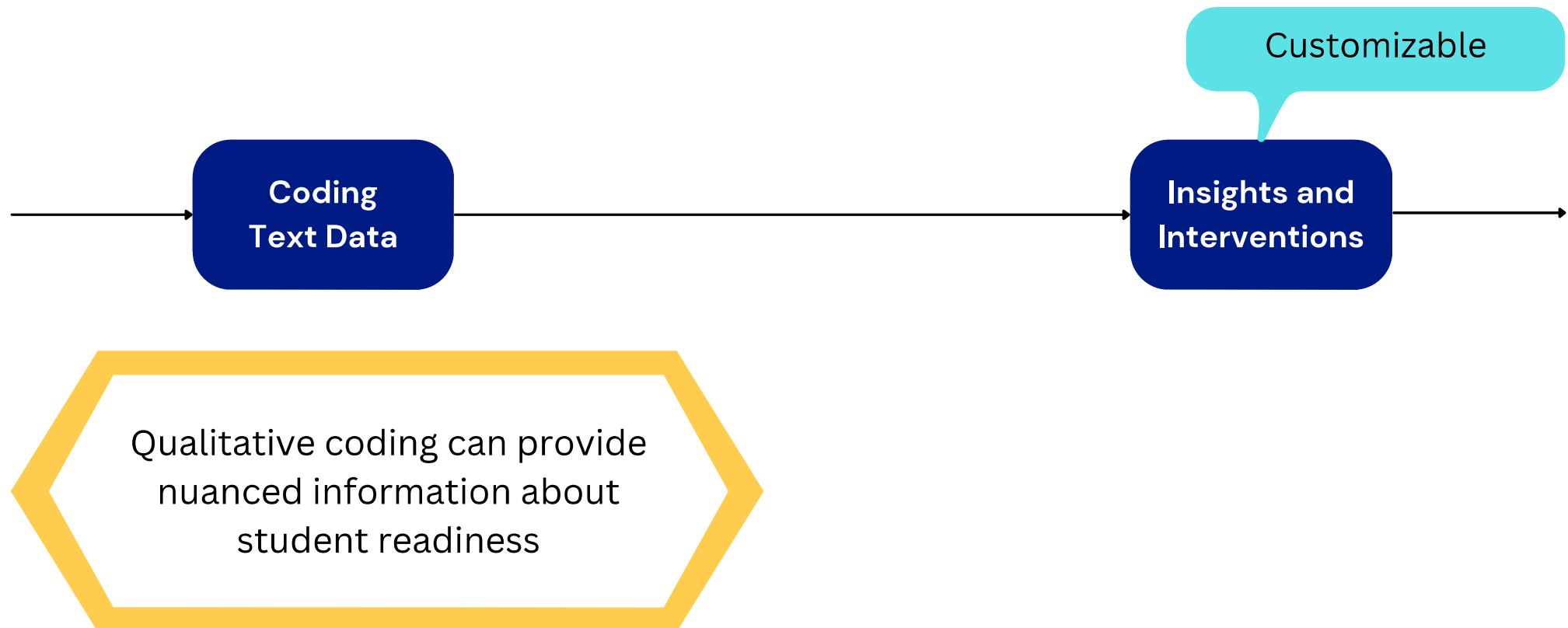
Visualization Use

Mathematical Language

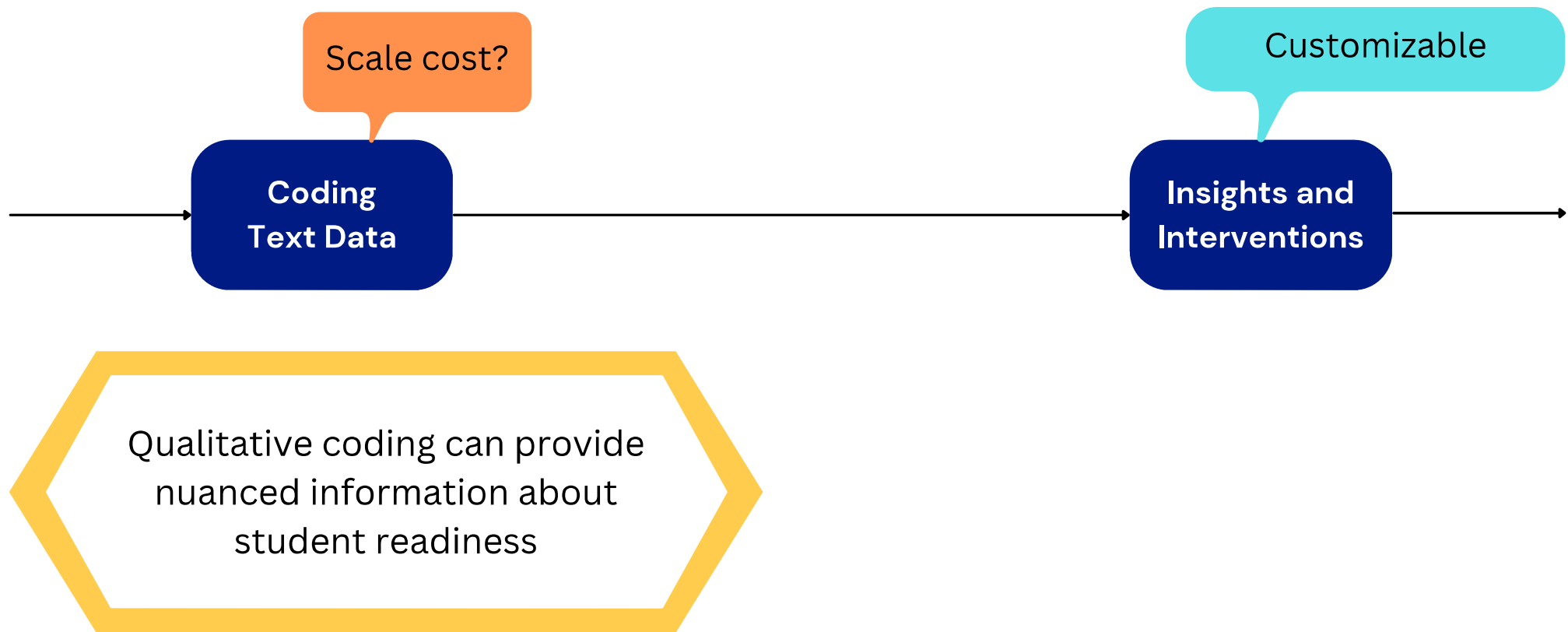
Contextual Reasoning

Heuristic View

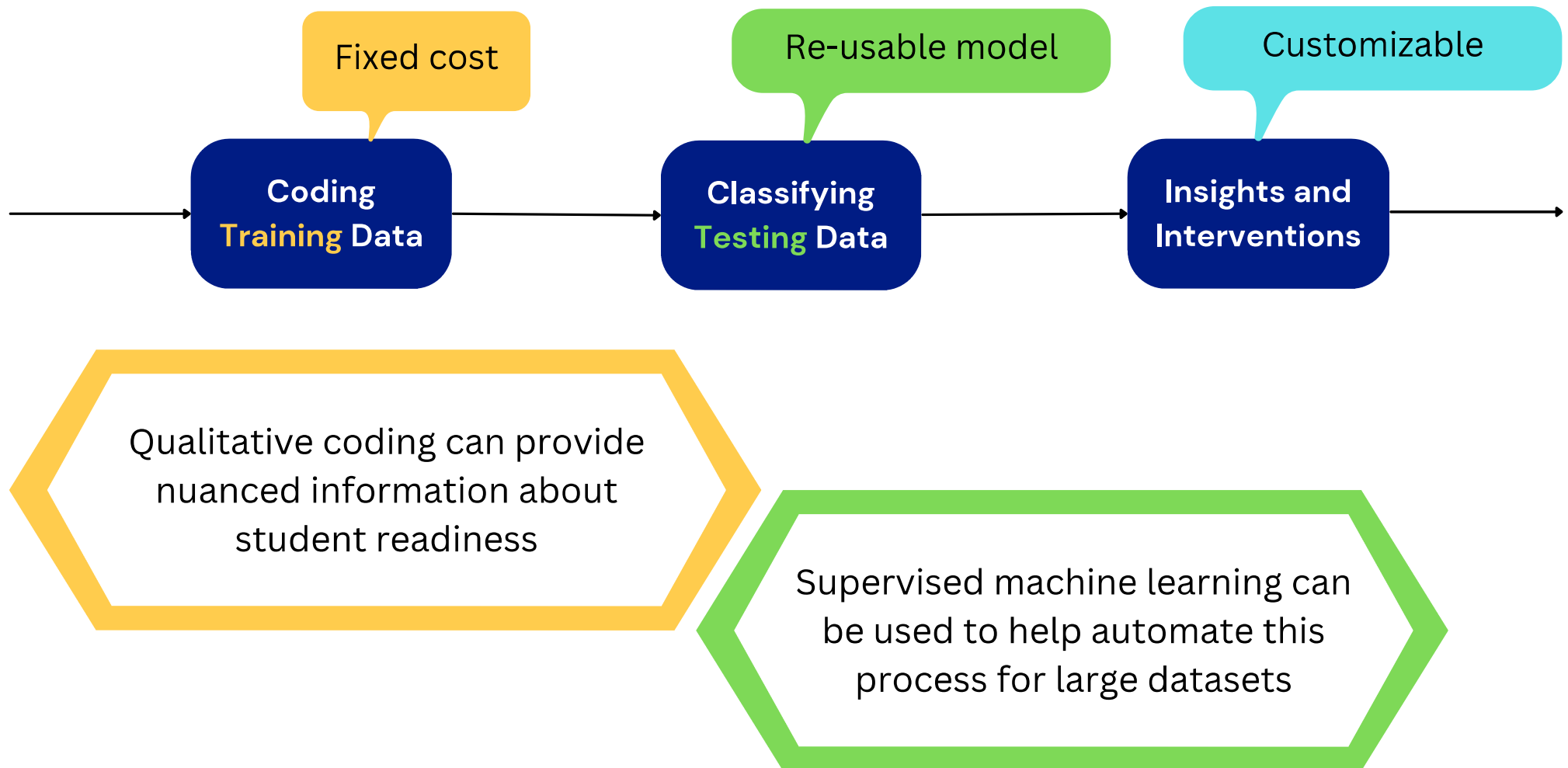
Our Methodology



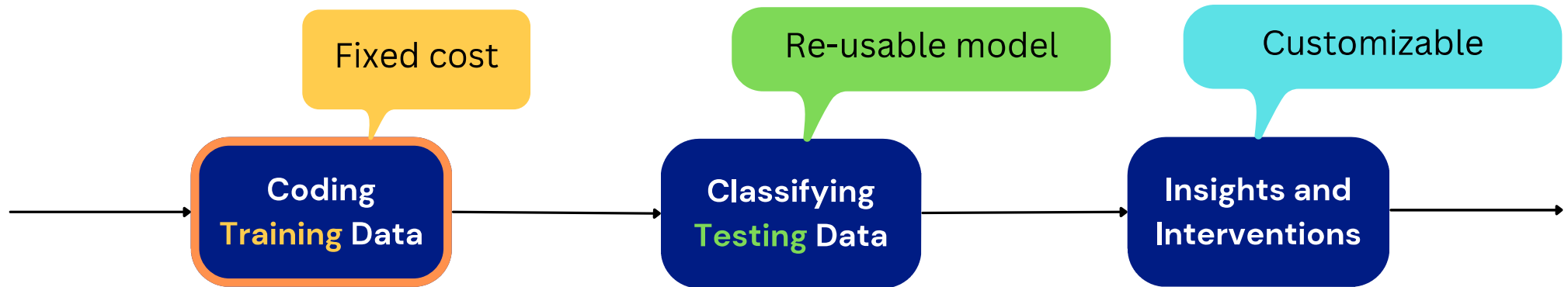
Our Methodology



Our Methodology

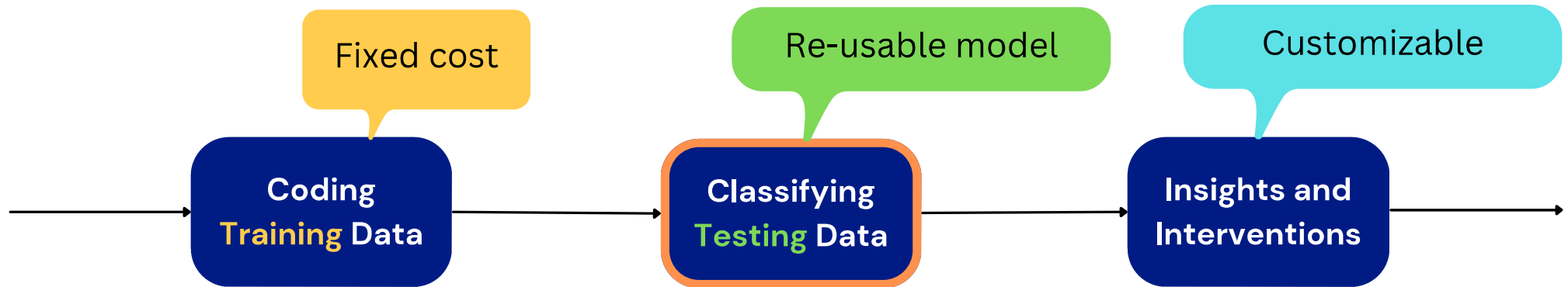


Our Methodology



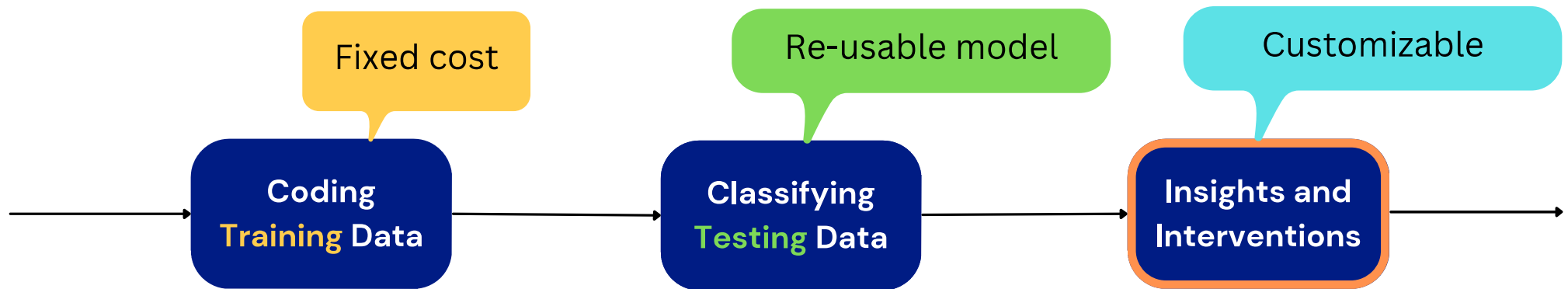
- Each response is first transformed using **Natural Language Processing**:
 - a. Raw text is converted to a list of lemmatized tokens, with 'stop' words removed by a customizable filter.
 - b. Recurring n-tuples of tokens (called n-grams) across responses are used to define the dimensions of a vector space.
 - c. Each token list is transformed into a vector in this space by counting the number of times a particular n-gram appears in the list.

Our Methodology

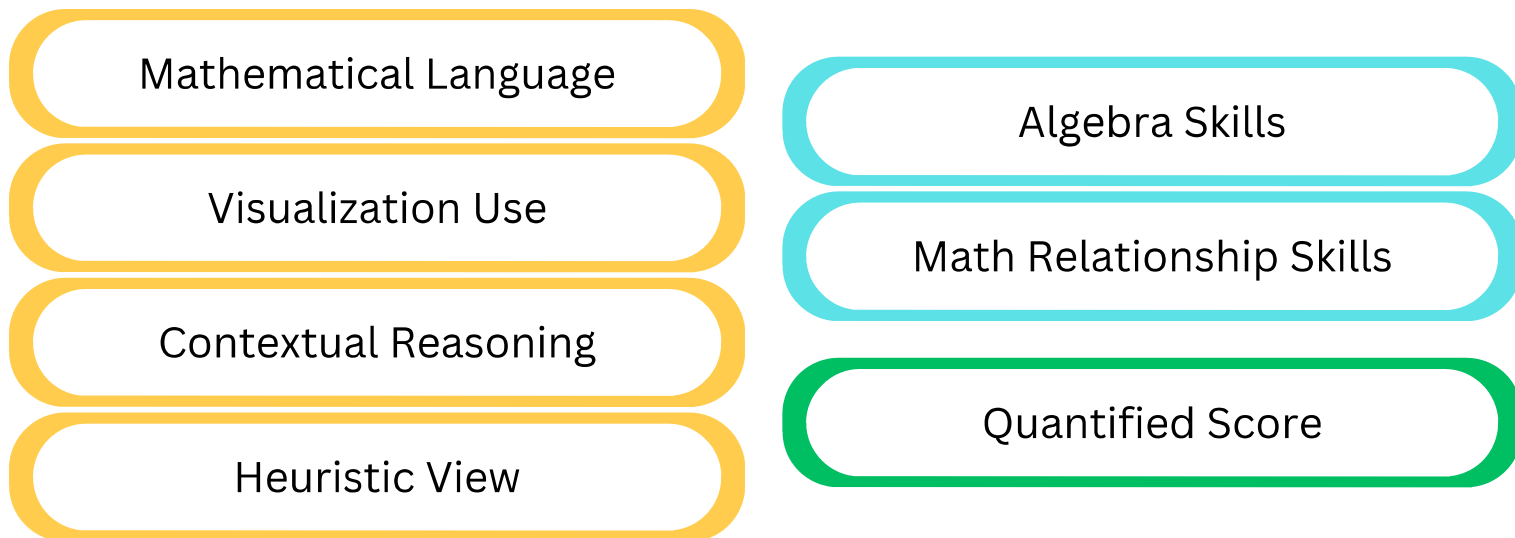


- These transformed vectors serve as **inputs** for Machine Learning (ML) models
- The **output** is a multi-label classification provided by the qualitative themes.
- Using these input-output pairs, ML models can automate qualitative coding:
 - We use gradient boosting machines (GBMs)
 - A GBM iteratively designs a sequence of decision trees to classify vectors
 - A separate GBM is trained for each question where we have a sufficient number of student responses in the training set showing a given theme

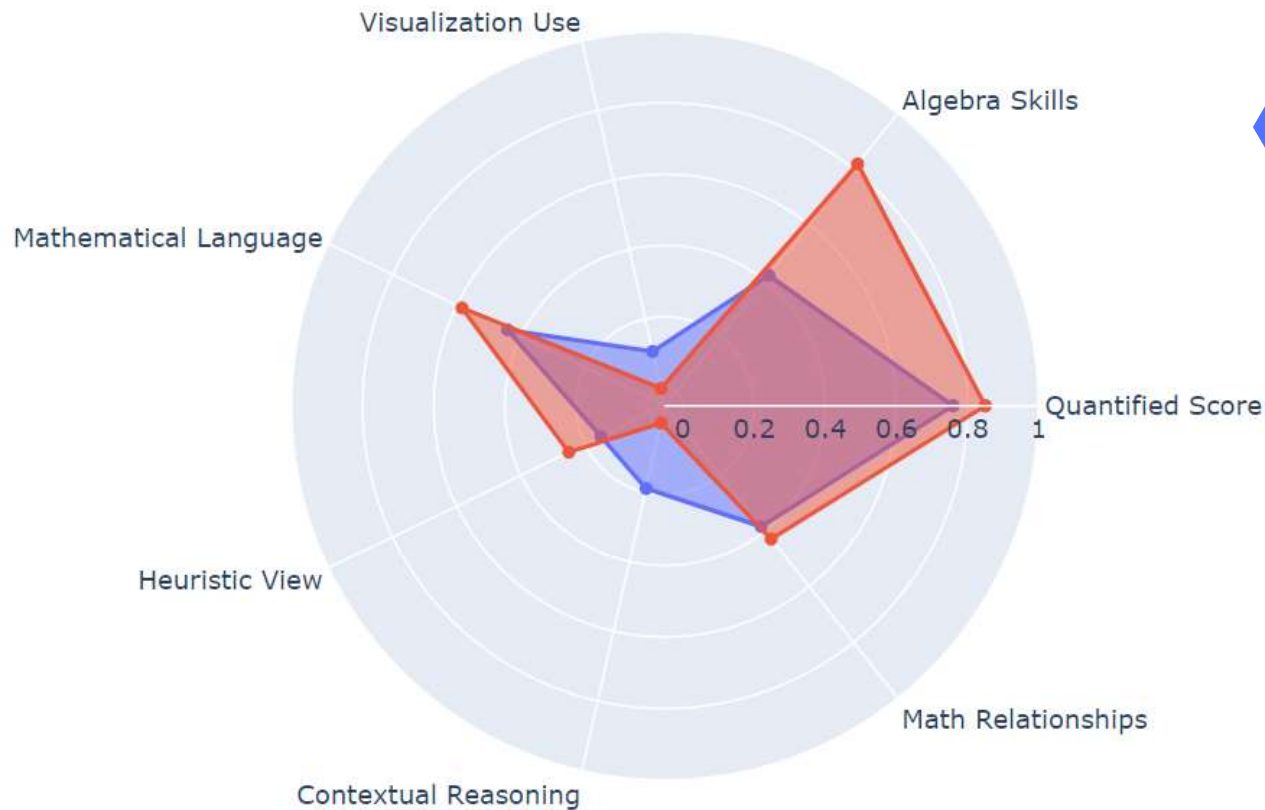
Our Methodology



- Taking a growth-mindset approach, we opt to focus on “neutral/positive” themes:



Data Visualization with Spider Plots

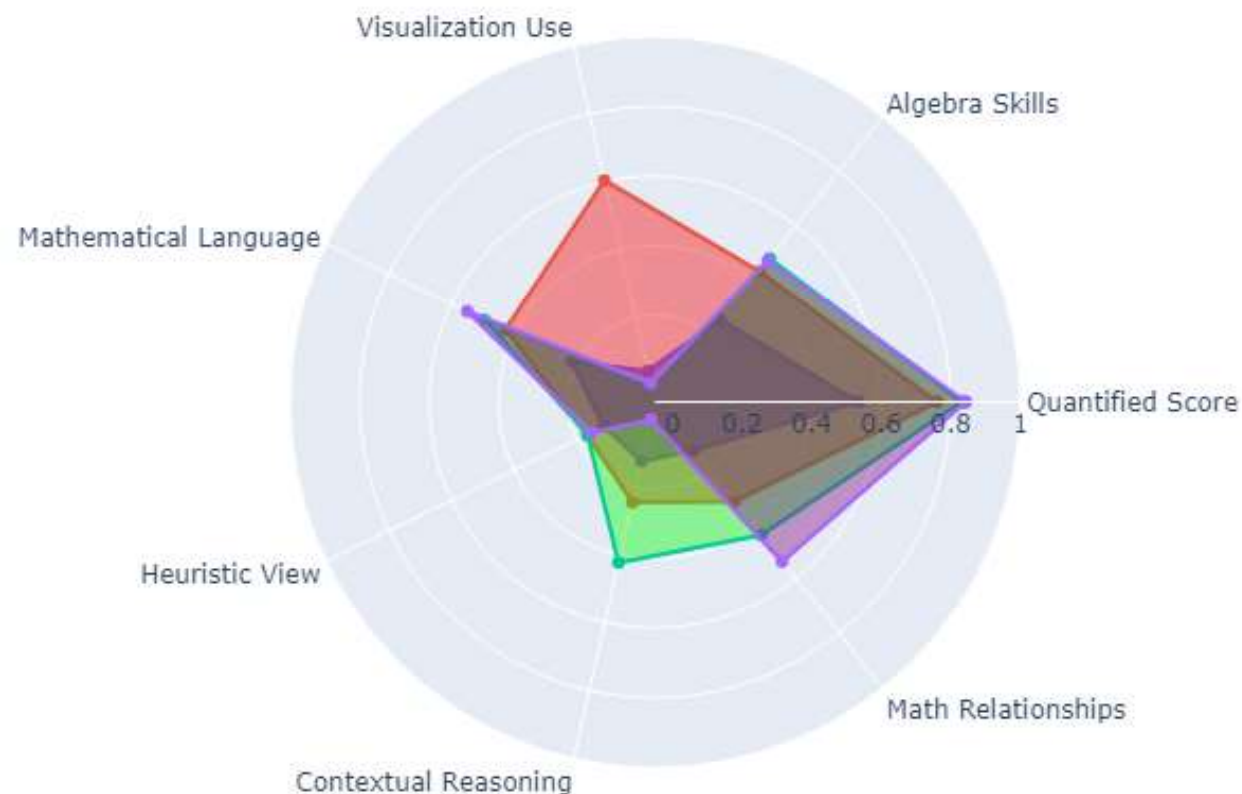


Defining Thematic Dimensions:

Each ray in the plot shows the proportion of times a student's text exhibited the given theme

The red area shows the profile of a random student while the blue area shows the averaged profile of every student in the cohort.

Clustering Student Cohort



Defining Thematic Dimensions:

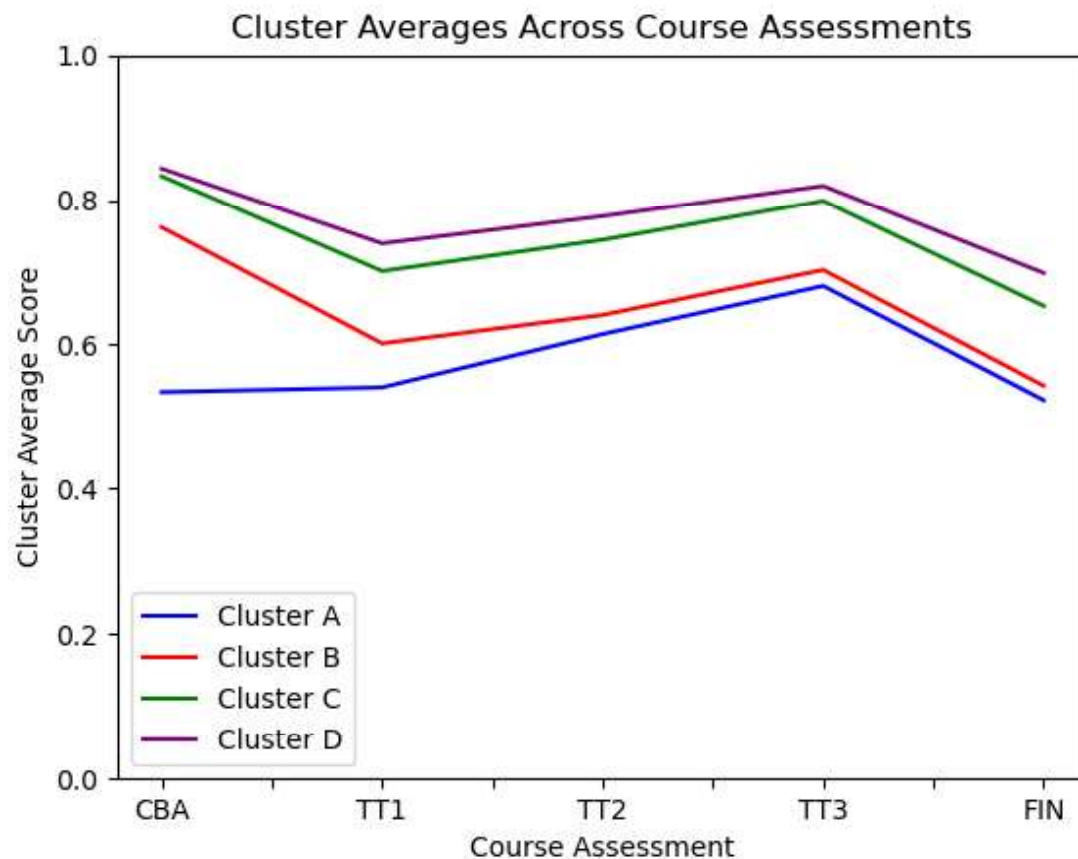
Each ray in the plot shows the proportion of times a student's text exhibited the given theme

Using k-Clustering:

Student results are partitioned based on dimensional distance, using the 6 chosen themes.

Each plot shown here is the centroid of a student cluster. cluster sizes range from 106 students (red) to 222 students (purple).

Clustering Student Cohort



Defining Thematic Dimensions:

Each ray in the plot shows the proportion of times a student's text exhibited the given theme

Using k-Clustering:

Student results are partitioned based on dimensional distance, using the 6 chosen themes.

How do clusters perform?:

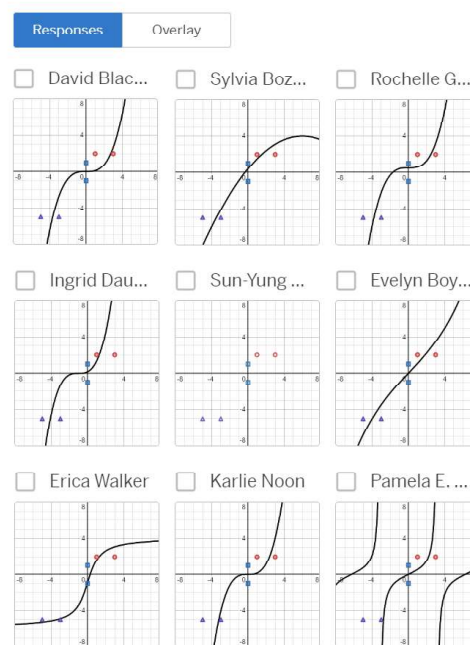
Each cluster's performance is tracked across course assessments for longitudinal observation.

Where do we go from here?

Using information these clusters, targeted student interventions can be designed and deployed by instructors.

Future CBA implementations can allow for these interventions and supports to be automatically delivered upon completion.

Desmos Function Challenge



Write an equation for a curve $y = f(x)$ that passes between each set of gates (red, blue, and purple) Then press "Try It."

EXPRESSION	STUDENTS
$y = .1x^3$	David Blackwell
$y = \frac{-1}{10}(x-6)^2 + 4$	Sylvia Bozeman
$y = (0.1)x^3 + 0.5$	Rochelle Gutierrez
$y = \frac{1}{6}(x+1)^3$	Ingrid Daubechies
$y = 1/80x^3 + 1.05x$	Evelyn Boyd Granville
$y = 3.2 \tan^{-1}(x) - 0.9$	Erica Walker
$y = 0.15x^3$	Karlle Noon
$y = \tan 0.5x$	Pamela E. Harris

Example of an intervention designed to support effective exploration and learning through desmos visualization

THANK YOU FOR YOUR TIME!

Please ask away with any questions
you may have!

*Visit our
Github repo
for a demo!*



Contact our team:

- Caroline Junkins, Assistant Professor, Department of Mathematics and Statistics, McMaster University (junkinc@mcmaster.ca)
- Lindsey Daniels, Assistant Professor of Teaching, Department of Mathematics, University of British Columbia (ldaniels@math.ubc.ca)
- Connor Gregor, Postdoctoral Fellow, Department of Mathematics and Statistics, McMaster University (gregoc9@mcmaster.ca)