Understanding course readiness through student text responses: a scalable framework

Caroline Junkins (McMaster)
Lindsey Daniels (UBC)
Connor Gregor (McMaster)
James Colliander (Crowdmark and UBC)

MAA Recording Policy

- The MAA is recording this event.
- We retain the right to show it again and to distribute it.
- By participating, you are agreeing that your contributions (i.e. text in the chat, your video if enabled, etc.) become part of the recording.
  - Note: Breakout rooms are NOT recorded.
- The recording is available by request to registered participants. Please email programs@maa.org to request.

- As reminder, we ask you be open to new ideas and kind in the comments.

Our Context: McMaster University

- McMaster is a medical doctoral public university located in Hamilton, Ontario, Canada.

- We recognize and acknowledge that researchers and 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.

- We recognize and acknowledge that the University of British Columbia, Point Grey (Vancouver) campus sits on the traditional, ancestral, unceded territory of the xʷməθkʷəy̓əm (Musqueam) First Nation.

Our Context: Undergrad Calculus Students

Undergraduate: 85% – 32,105	Full-Time: **96%** – **30,747**	Domestic: **85%** – **27,427**
	Part-Time: **4%** – **1,358**	International: **15%** – **4,678**
Graduate: 15% – 5,487	Full-Time: **84%** – **4,597**	Domestic: **69%** – **3,780**
	Part-Time: **16%** – **890**	International: **31%** – **1,707**

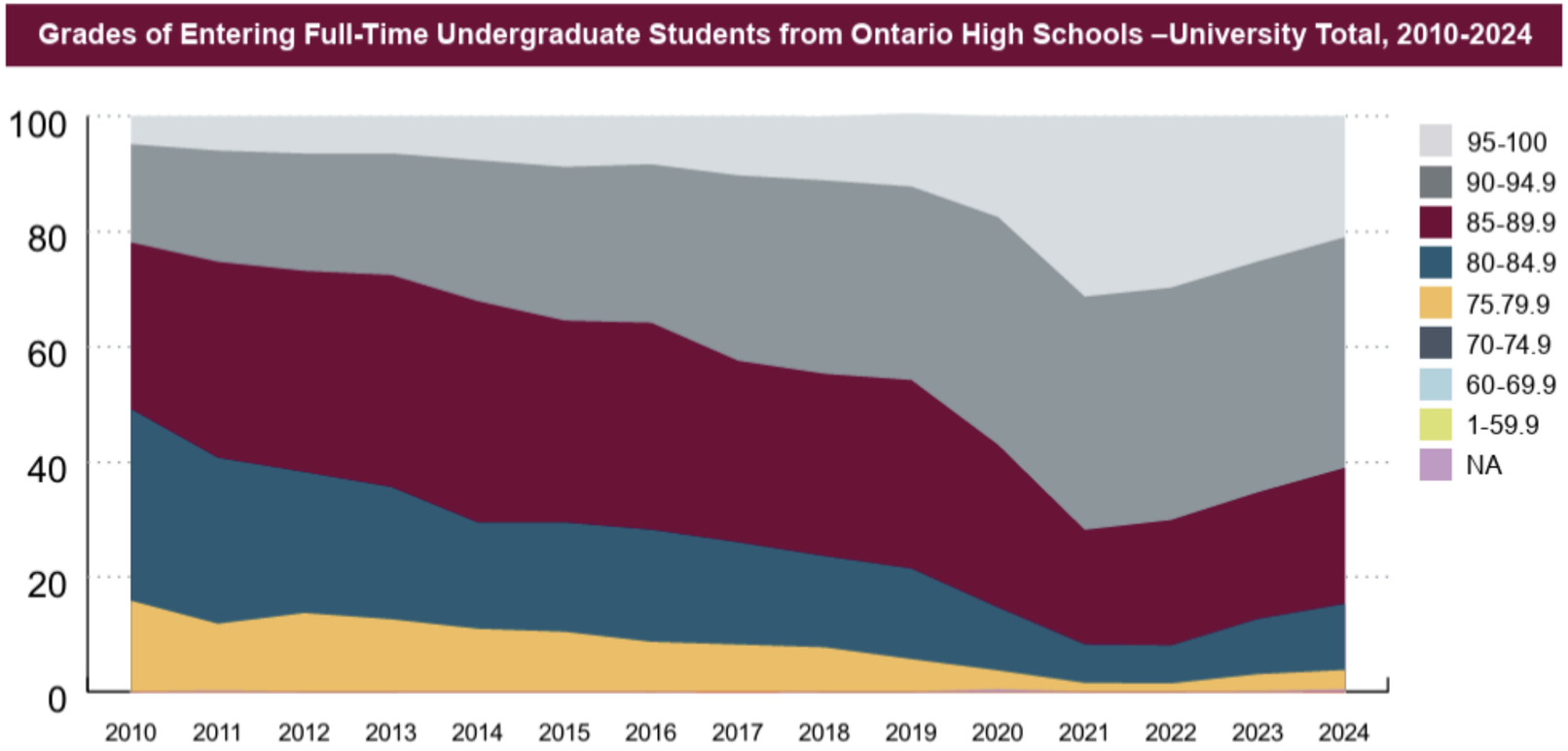
Each year, approximately 5000 of these students will enrol in “Calculus I”:

- **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:

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

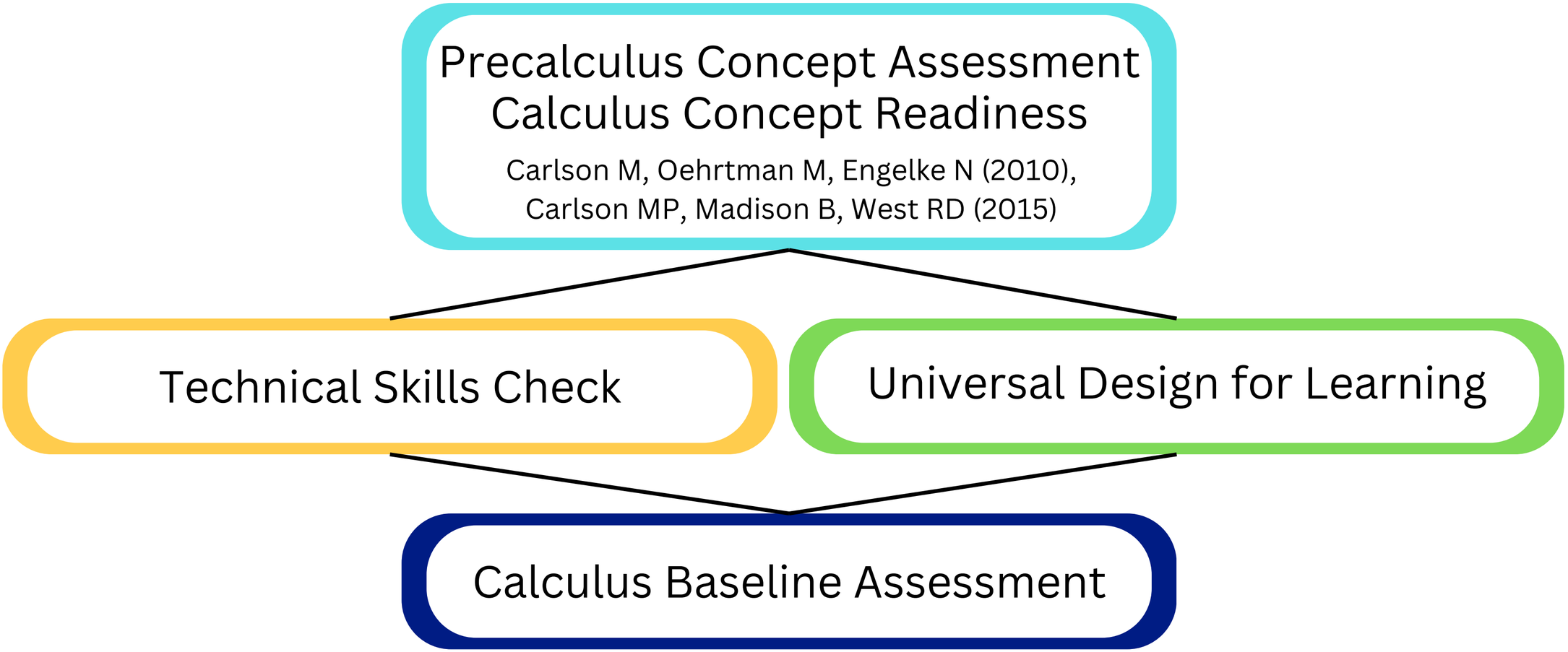


Readiness for University Calculus

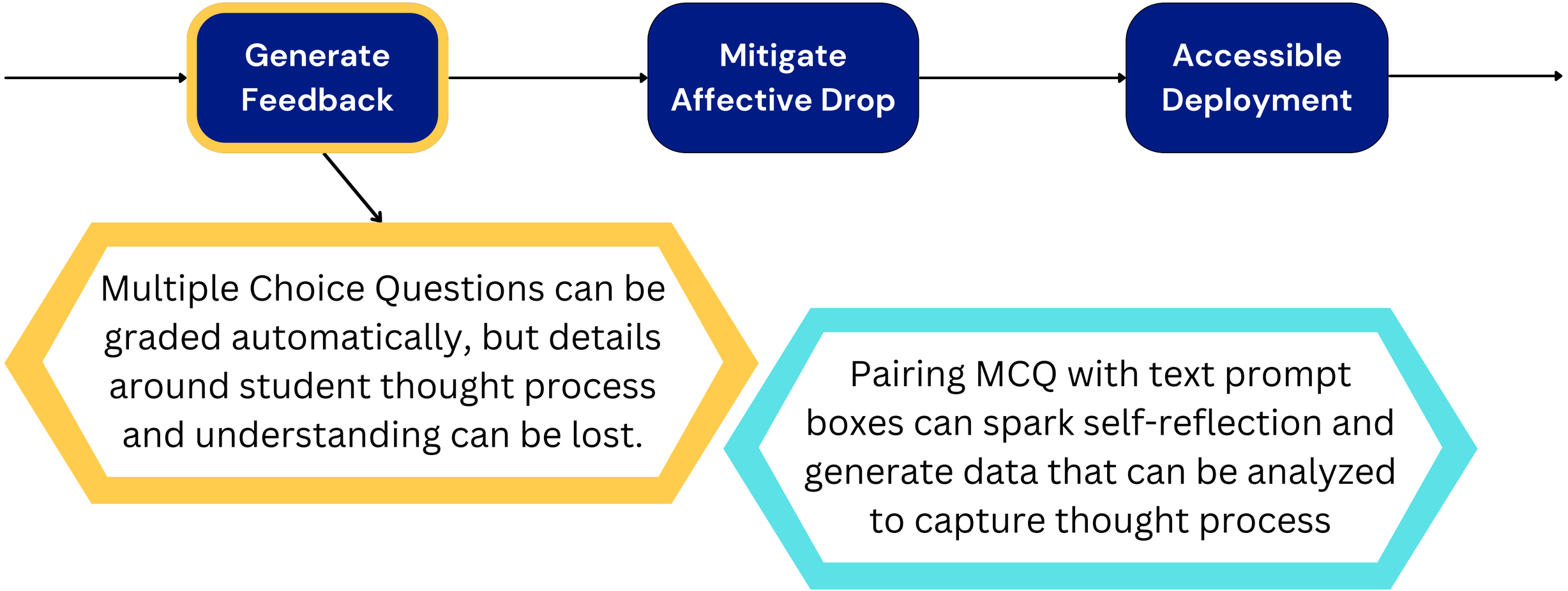
- Can we assess which students are well-prepared for their required calculus course, and which may need **additional support**?
- Can we identify **existing gaps** in pre-requisite skills and conceptual understanding for a given cohort of students?
- Can we encourage students to demonstrate **critical thinking** and **communication** skills in a calculus setting?
- Can we do this in a scalable way for **large-enrollment courses**?

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.



CBA 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.

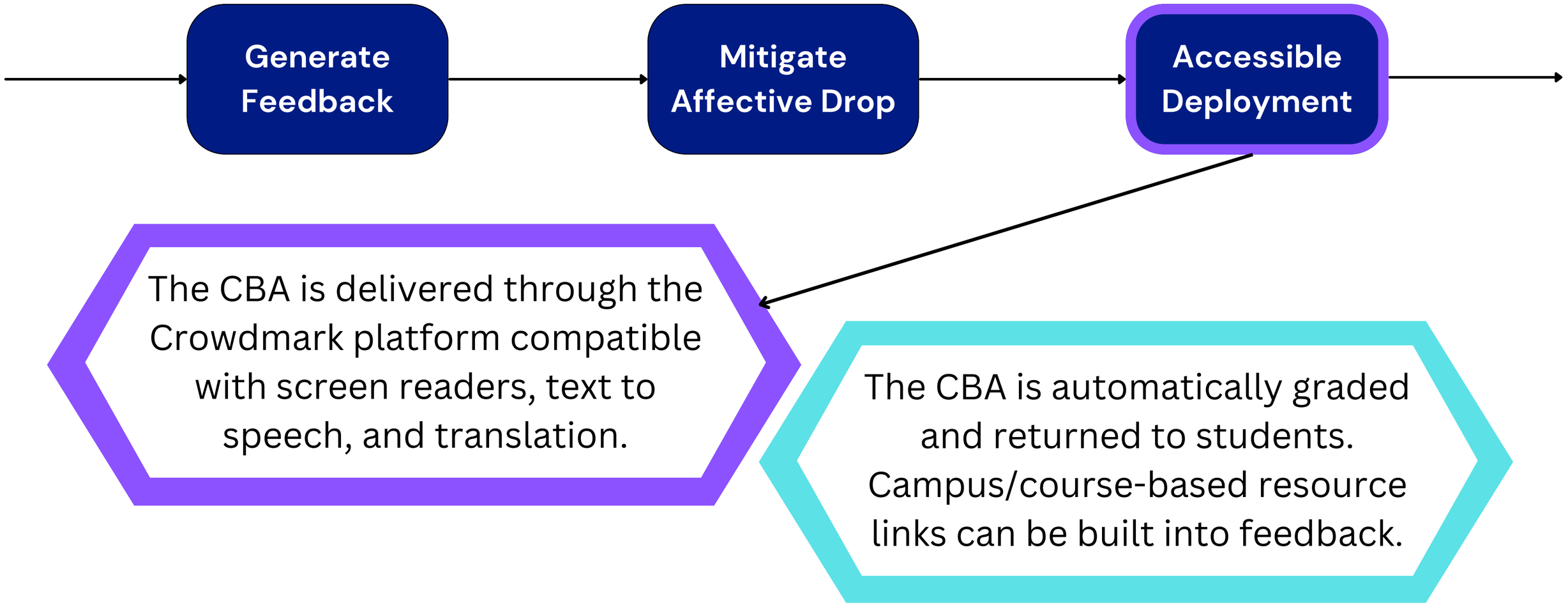
CBA Design Objectives

High-stakes test settings can degrade math performance for students experiencing anxiety and/or stereotype threat

The CBA is administered as an open-book, online assessment, to be completed by students over an extended time period (>7 days)

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.

CBA 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

Q15a (1 point)

How does the graph of $g(x) = \cos(x)$ compare to the graph of $h(x) = \cos(2x)$?

- The period of $h(x)$ is double the period of $g(x)$.
- The period of $h(x)$ is half the period of $g(x)$.
- The amplitude of $h(x)$ is double the amplitude of $g(x)$.
- The amplitude of $h(x)$ is half the amplitude of $g(x)$.
- There is no difference between the graphs.

Q15b (0 points)

Explain the reasoning for your answer to Q15a in the box below.

Edit Preview

Please enter your response to Q15b

Attach files Formatting tips

CBA Sample Question

Q15a (1 point)

How does the graph of $g(x) = \cos(x)$ compare to the graph of $h(x) = \cos(2x)$?

- The period of $h(x)$ is double the period of $g(x)$.
- ✓ The period of $h(x)$ is half the period of $g(x)$.
- The amplitude of $h(x)$ is double the amplitude of $g(x)$.
- The amplitude of $h(x)$ is half the amplitude of $g(x)$.
- There is no difference between the graphs.

Q15b (0 points)

Explain the reasoning for your answer to Q15a in the box below.

Edit Preview

Please enter your response to Q15b

Attach files Formatting tips

After looking at the graphs, I could tell that h(x)'s period is half of g(x)'s period. Another reason is because the 2 in cos(2x) compresses cos(x) by a factor of two, "cutting" the period in half.

CBA Sample Question

Q15a (1 point)

How does the graph of $g(x) = \cos(x)$ compare to the graph of $h(x) = \cos(2x)$?

- The period of $h(x)$ is double the period of $g(x)$.
- ✓ The period of $h(x)$ is half the period of $g(x)$.
- The amplitude of $h(x)$ is double the amplitude of $g(x)$.
- The amplitude of $h(x)$ is half the amplitude of $g(x)$.
- There is no difference between the graphs.

Q15b (0 points)

Explain the reasoning for your answer to Q15a in the box below.

Edit Preview

Please enter your response to Q15b

Attach files Formatting tips

After looking at the graphs, I could tell that h(x)'s period is half of g(x)'s period. Another reason is because the 2 in cos(2x) compresses cos(x) by a factor of two, "cutting" the period in half.

We use semi-automated text analysis to flag evidence of thought processes

- ✓ Visualization Use
- Mathematical Language
- ✓ Math Relationship Skills

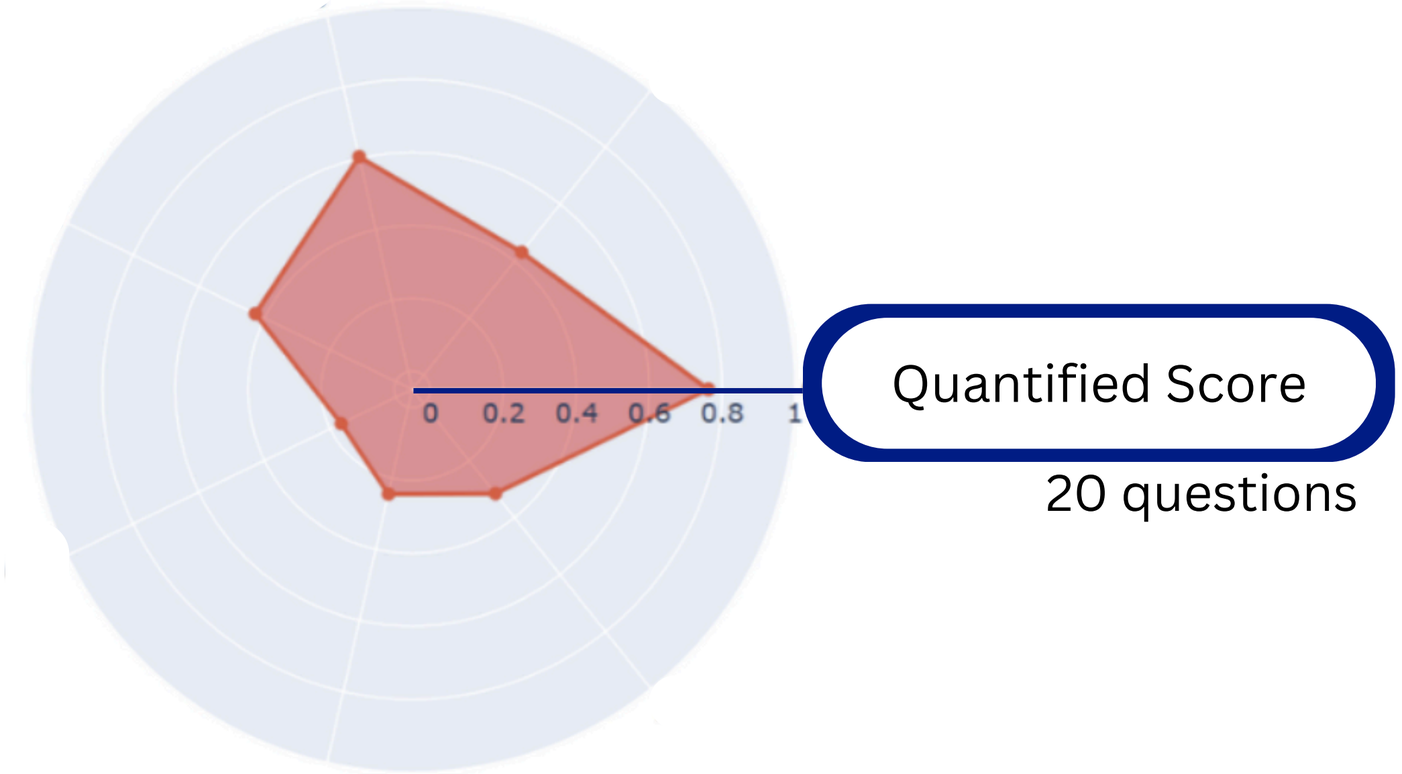
Qualitative Coding Procedure

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

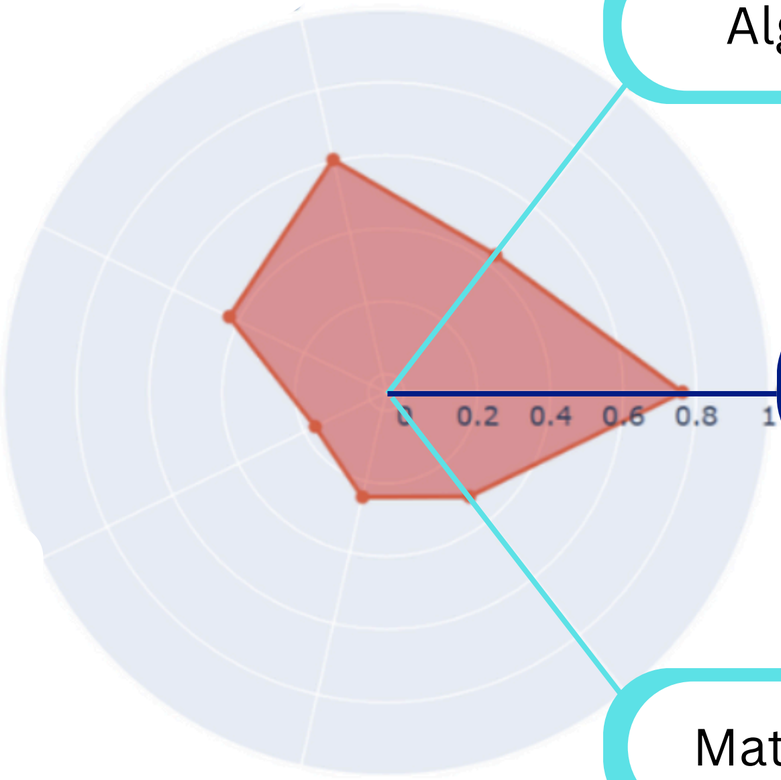
CBA Results Dashboard: Student Level

We can combine numercial score with qualitative info from select questions



CBA Results Dashboard: Student Level

We can combine numercial score with qualitative info from select questions



Algebra Skills

11 questions

Quantified Score

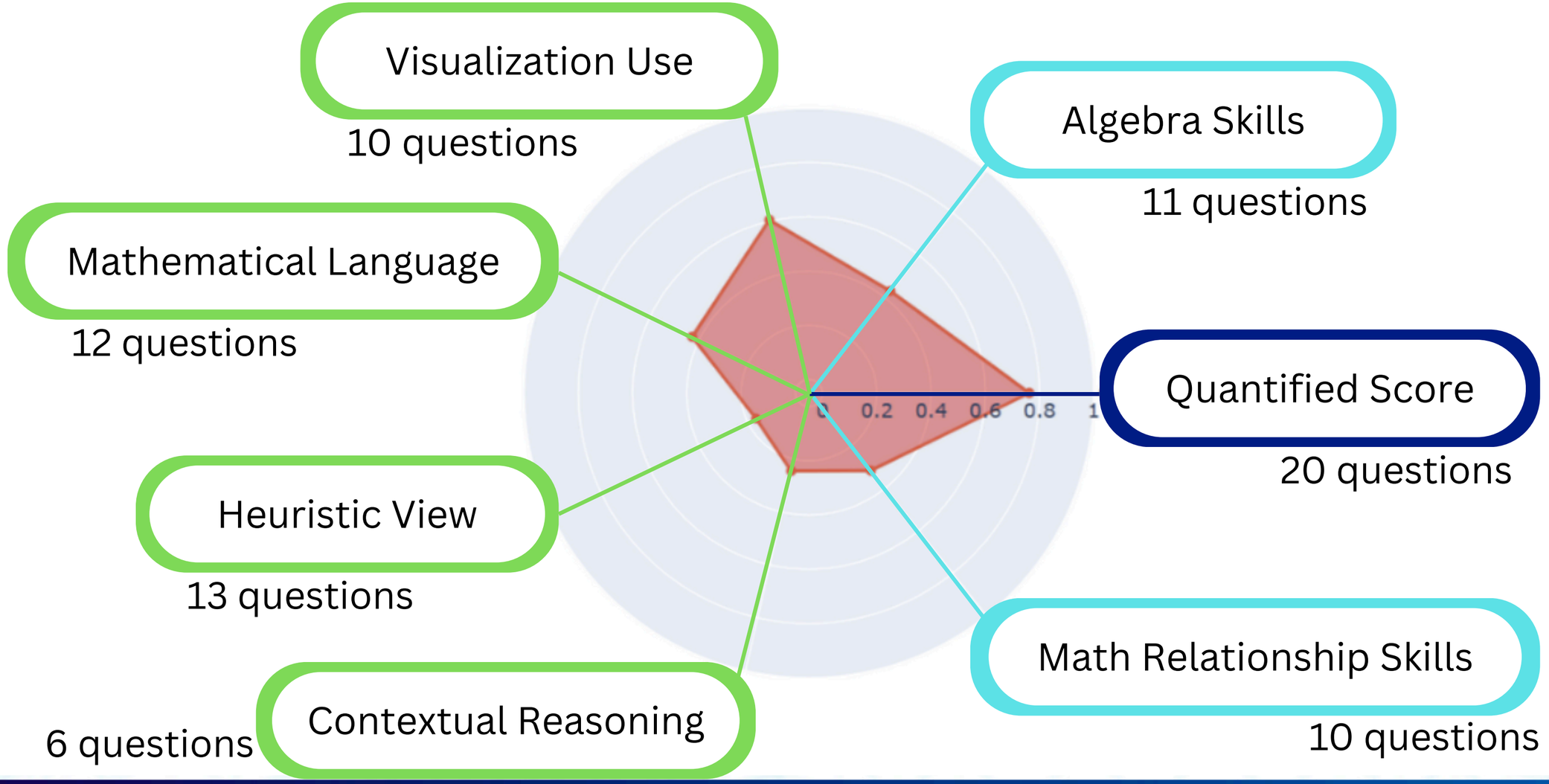
20 questions

Math Relationship Skills

10 questions

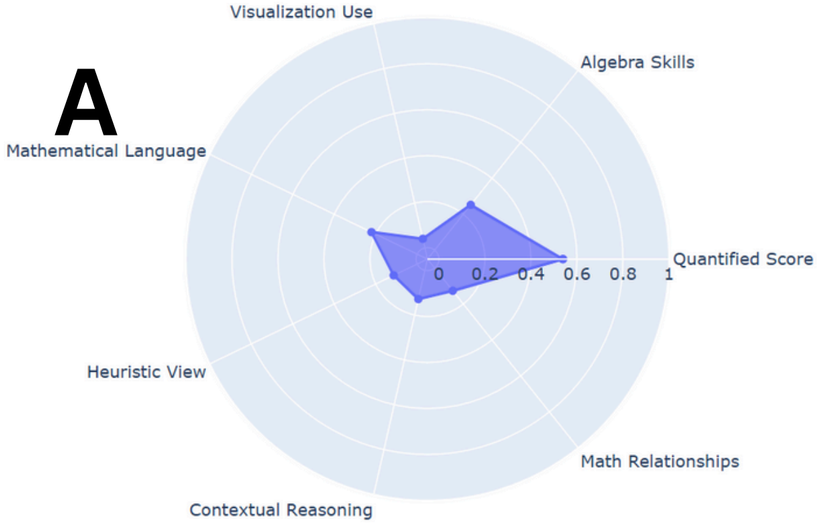
CBA Results Dashboard: Student Level

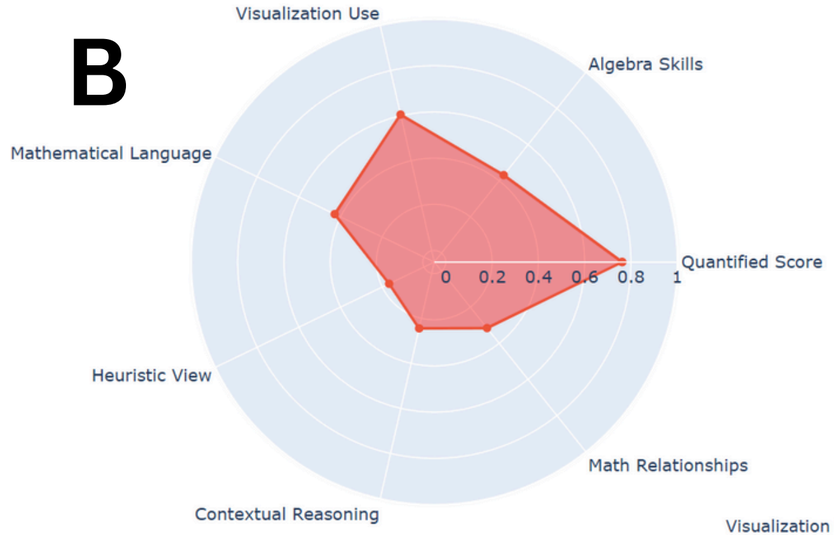
We can combine numercial score with qualitative info from select questions

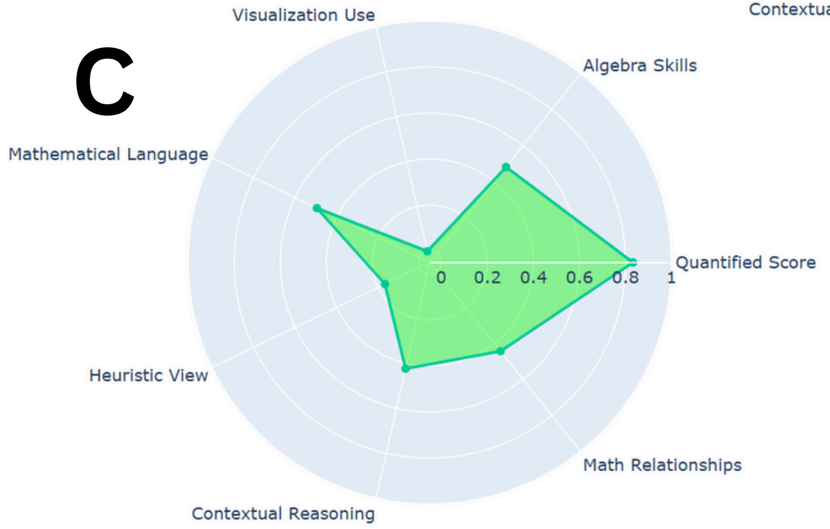


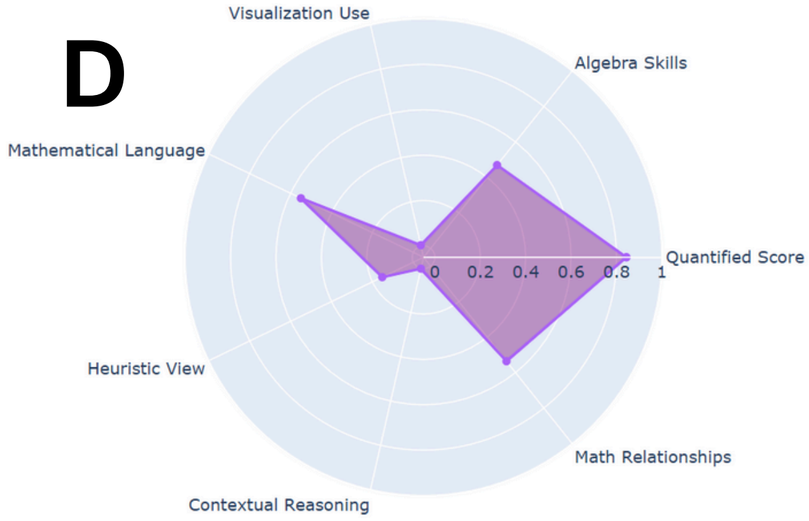
CBA Results Dashboard: Cohort Level

k-clustering based on dimensional distance of the 6 qualitative themes.









A: 147 students

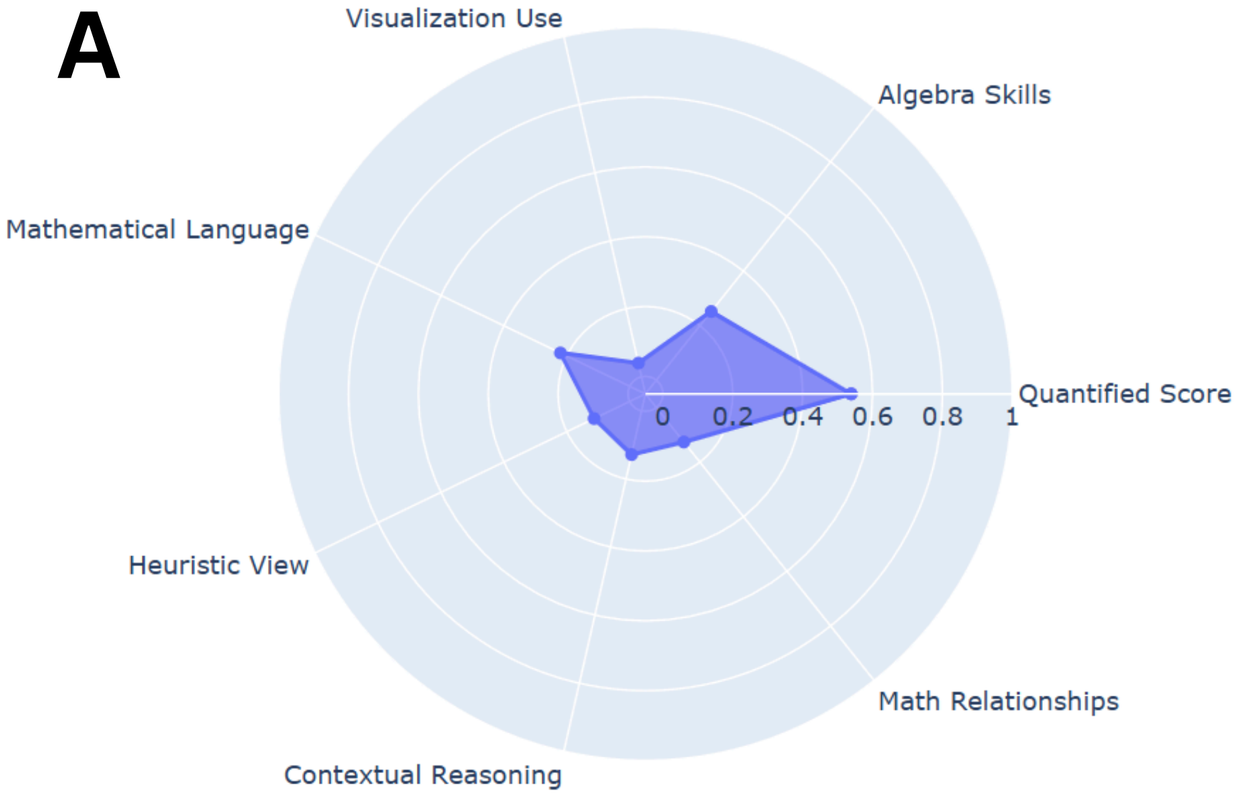
B: 106 students

C: 188 students

D: 222 students

CBA Results Dashboard: Cohort Level

A



- lowest numerical score
- fewer themes detected from student text

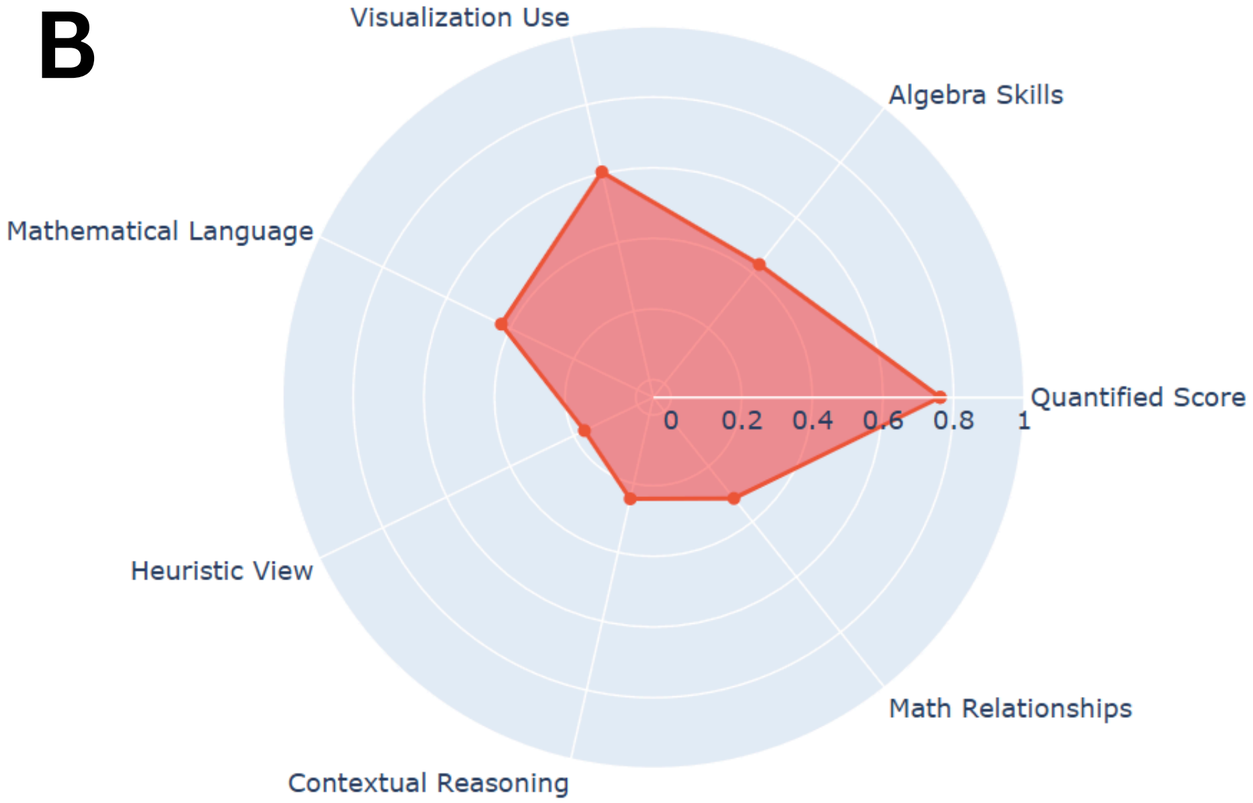
I guessed.

I just remember

anything in the bracket is opposite

CBA Results Dashboard: Cohort Level

B

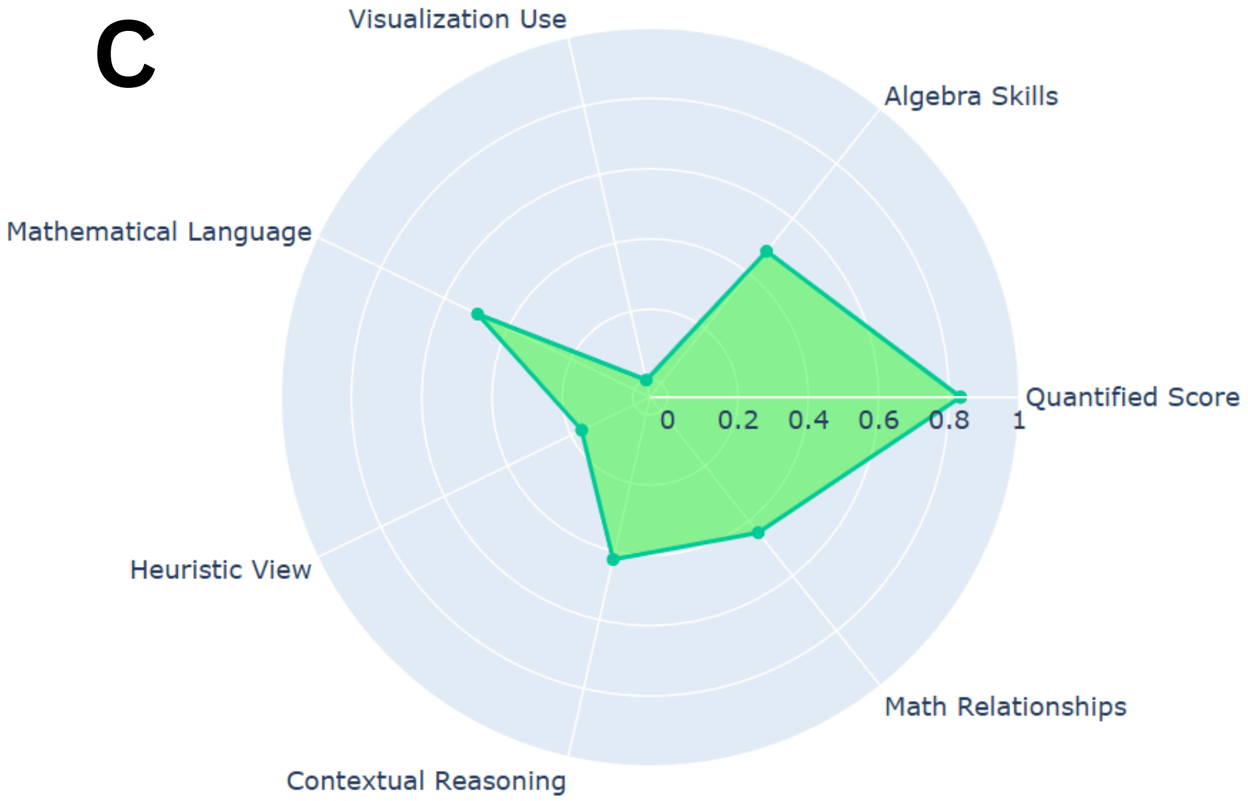


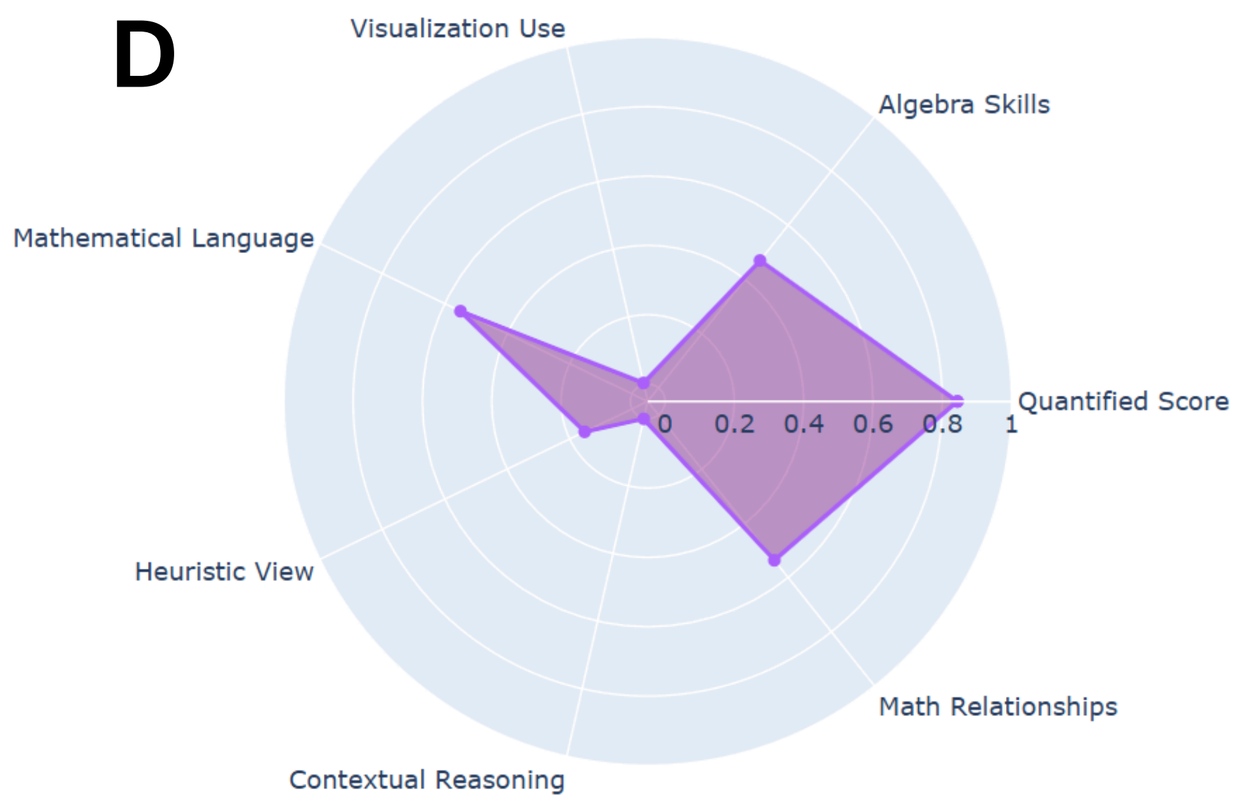
- more themes detected
- high visualization use

I graphed the functions and saw that the 2 was affecting the period for h(x) by approximately 1/2.

To obtain this answer, I used Desmos Graphing Calculator. By observing the graphs, I concluded that the period of h)x) is half of g(x).

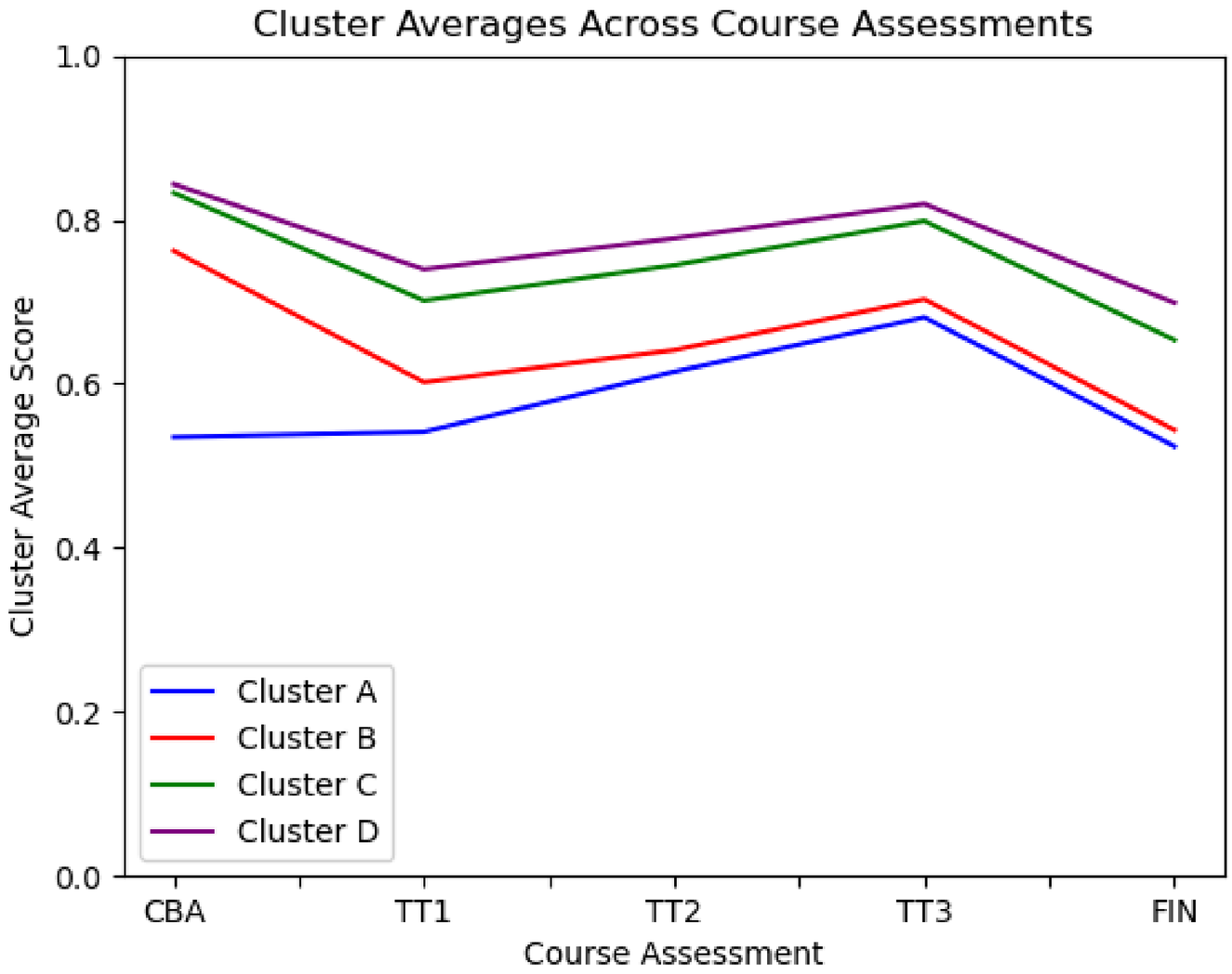
CBA Results Dashboard: Cohort Level



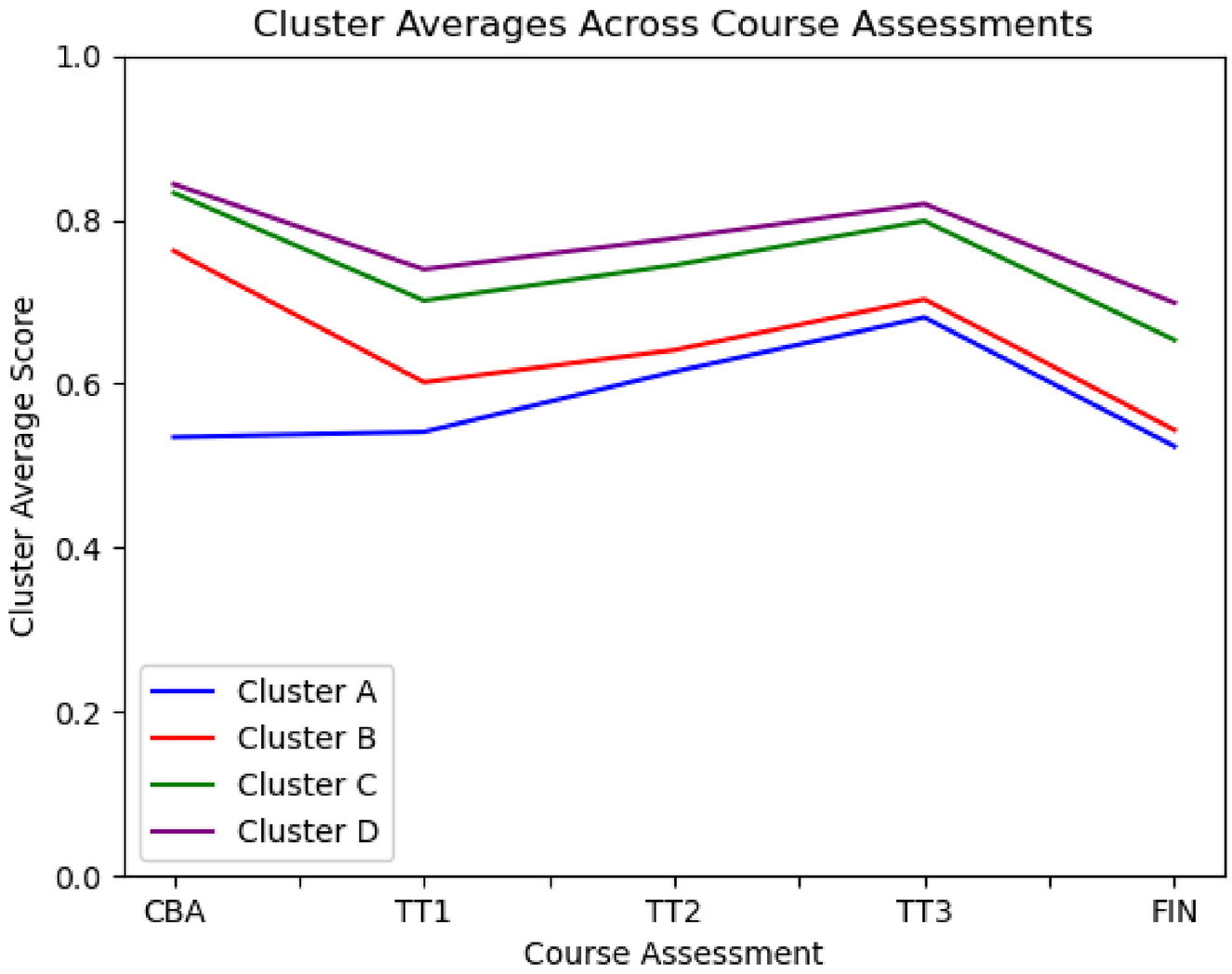


- similar in all dimensions except contextual reasoning

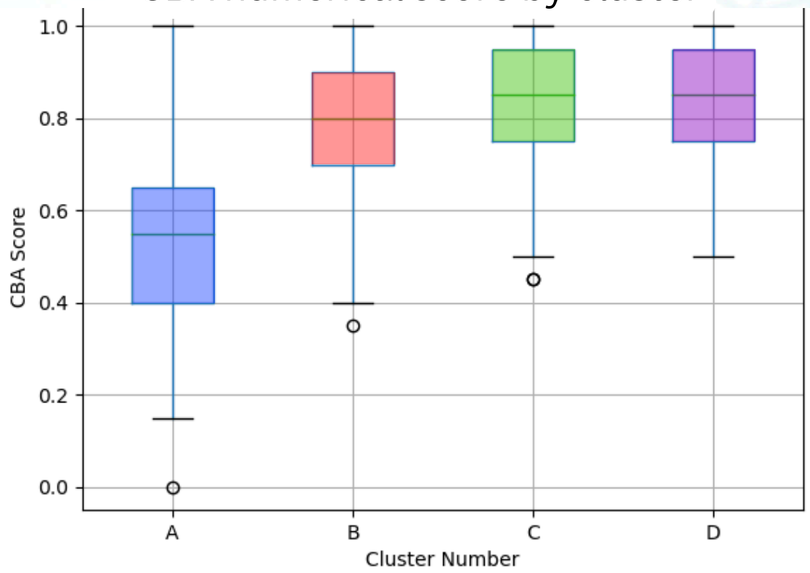
Course Progress by Cluster



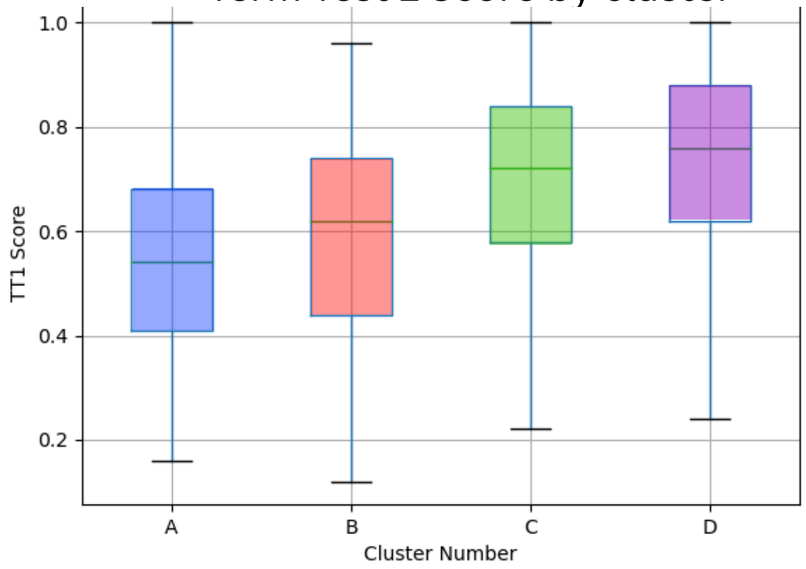
Course Progress by Cluster



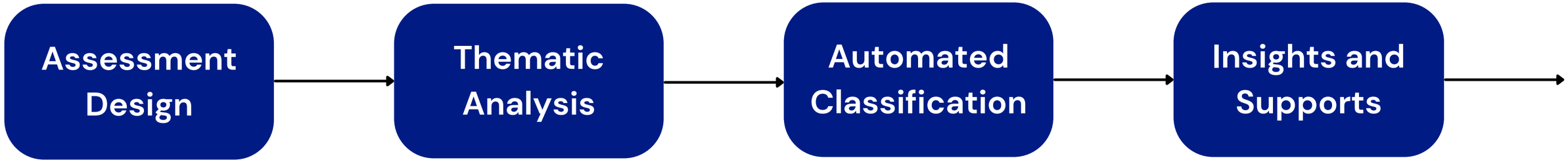
CBA numerical score by cluster



Term Test 1 Score by cluster



CBA Framework Overview



CBA Framework Overview

Assessment Design → **Thematic Analysis** → **Automated Classification** → **Insights and Supports** →

Adaptable for any course context

Precalculus Concept Assessment
Calculus Concept Readiness

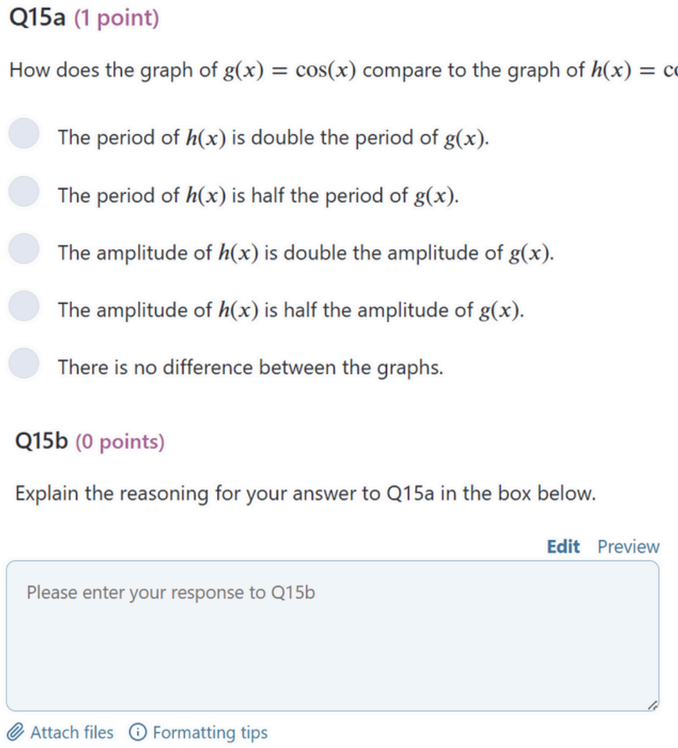
Carlson M, Oehrtman M, Engelke N (2010),
Carlson MP, Madison B, West RD (2015)

+

Technical Skills Check

=

Calculus Baseline Assessment



Q15a (1 point)

How does the graph of $g(x) = \cos(x)$ compare to the graph of $h(x) = \cos(2x)$?

- The period of $h(x)$ is double the period of $g(x)$.
- The period of $h(x)$ is half the period of $g(x)$.
- The amplitude of $h(x)$ is double the amplitude of $g(x)$.
- The amplitude of $h(x)$ is half the amplitude of $g(x)$.
- There is no difference between the graphs.

Q15b (0 points)

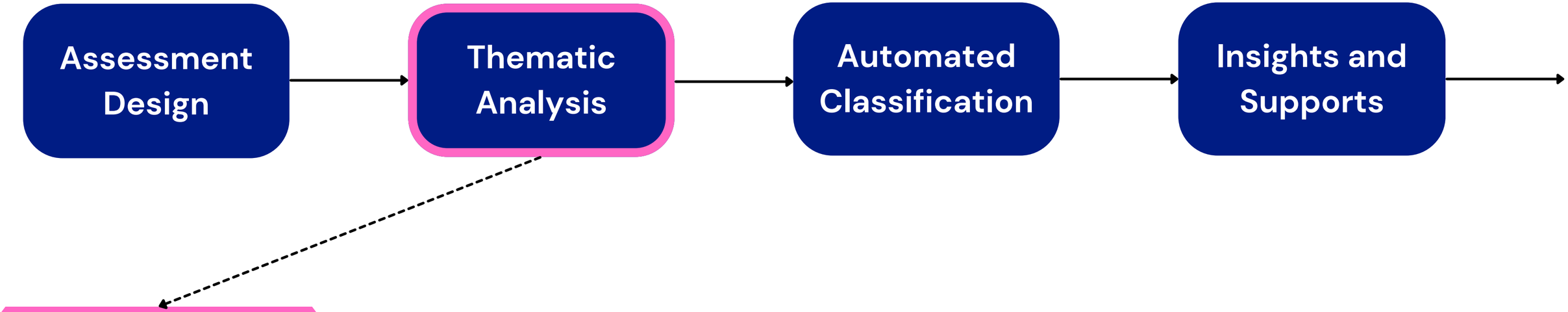
Explain the reasoning for your answer to Q15a in the box below.

Edit Preview

Please enter your response to Q15b

Attach files Formatting tips

CBA Framework Overview

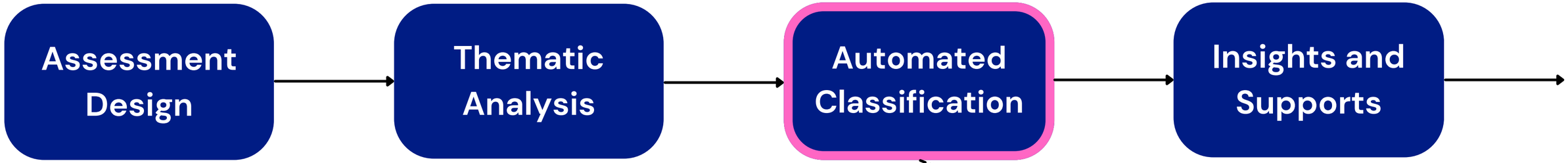


Fixed Cost on initial deployment

Themes aligned with learning objectives:

- Algebra Skills
- Math Relationship Skills
- Solution Framework
- Visualization Use
- Contextual Reasoning
- Mathematical Language

CBA Framework Overview




```
import numpy as np
import NLPmod
import re
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.corpus import words
from nltk.corpus import stopwords
from nltk import pos_tag


from sklearn.ensemble import GradientBoostingClassifier
from sklearn.feature_extraction.text import TfidfVectorizer
import plotly.express as px
import plotly.graph_objects as go

from random import randint
import plotly.express as px
import plotly.graph_objects as go
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
from collections import Counter
```

Natural Language Processing

+

Gradient Boosting Machines

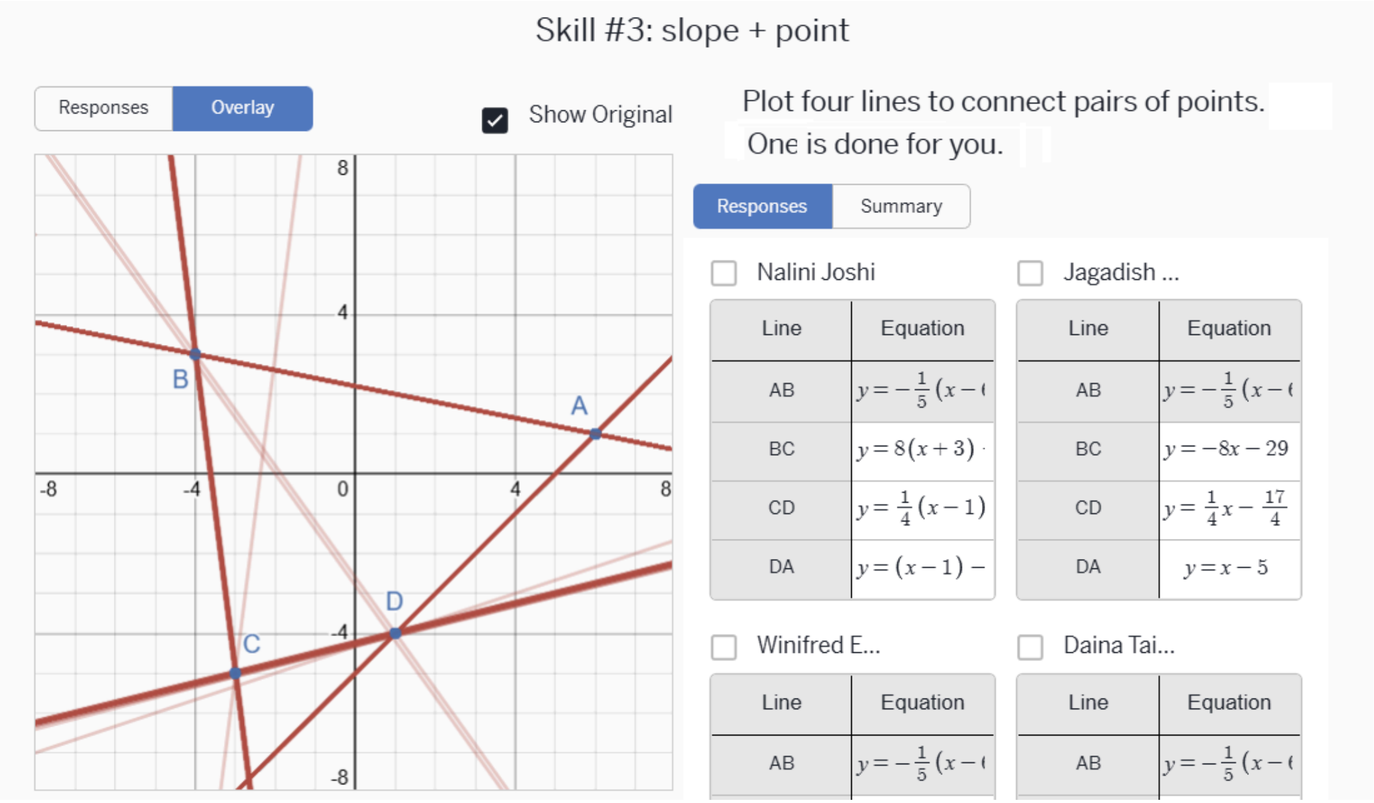
+

k-means Clustering

No cost to scale for course size

CBA Framework Overview

Assessment Design → **Thematic Analysis** → **Automated Classification** → **Insights and Supports** →



Customized and **individualized**

Digital Skill Lab

In-person workshop

THANK YOU TO MITACS, CROWDMARK, McMASTER, UBC, AND THE MAA!

THANK YOU FOR YOUR TIME!
Please ask away with any questions you may have!

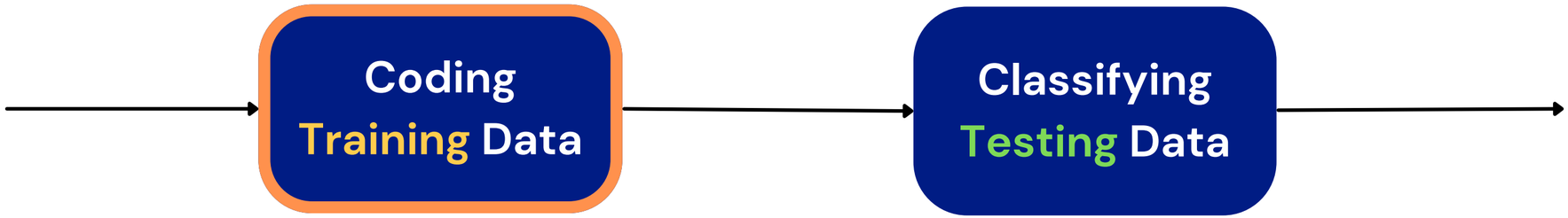
Visit our Github repo for more!

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

Additional Info

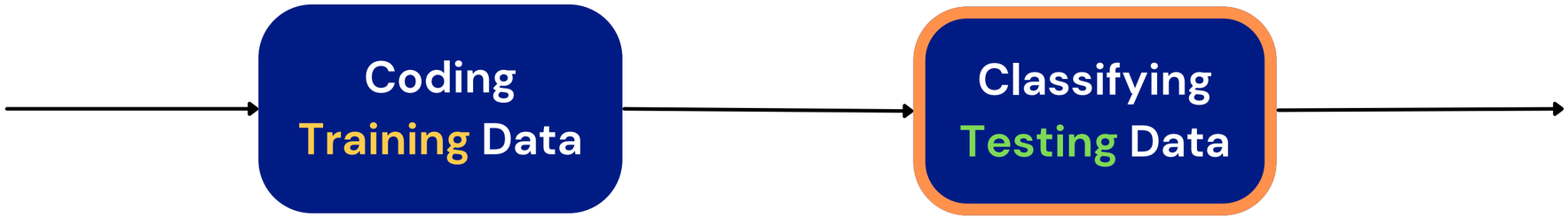
Our Methodology



- 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