ARTICLE TEMPLATE

Graph Theoretic Reflection to Foster Alignment in Coordinated Courses

Haile Gilroy[a], Sean Grate[a], Melinda Lanius[a]

[a]Auburn University Department of Mathematics & Statistics, 221 Parker Hall, Auburn, Alabama 36849, USA

ARTICLE HISTORY
Compiled April 15, 2024

ABSTRACT
Despite online homework's growing prevalence as a uniform component in coordinated mathematics courses, few studies have considered the connection, or lack thereof, between instructors of record and fixed online homework sets. In this mixed-methods study, we examined how 10 university mathematics educators working in a coordinated setting judged the quality of a sampling of online Calculus I homework assignments. Following an initial review of the homework sets, we introduced the educators to a novel instrument called the Course Alignment Analysis Tool (CAAT), which leverages graph theory to assess the alignment between the learning outcomes that an instructor feels should be prioritised and the learning outcomes most emphasised by an assignment or assessment. We analyzed the impact of engaging with the CAAT on participants' consideration of uniform homework. We found that interacting with the CAAT affected coordinated instructors' definitions of homework quality and that the CAAT is a promising professional development tool for novice instructors in particular.

KEYWORDS
alignment; calculus; online homework; graph theory; course coordination

1. Introduction

Uniform course components, or course elements that are fixed across all sections of a course, are extremely popular amongst Calculus I courses in Masters or Ph.D. granting mathematics departments in the United States (Rasmussen et al., 2019). Unfortunately, at many of these universities, Calculus I is also a notorious 'weed-out' course that operates in such a way that only the highest-performing students pass, contributing to a leaky pipeline where large numbers of STEM majors do not complete a STEM degree; this loss has been partially attributed to a lack of alignment in the content covered between various course components, such as in-class instruction, homework assignments, and assessments (Weston et al., 2019). Alignment is the degree to which cohesion exists between course components including learning objectives, in-class materials, homework, and assessments (Wiggins & McTighe, 1998). In this paper, we explore the role of a uniform course component in instructors' considerations of course alignment.

CONTACT Haile Gilroy. Email: hmg0038@auburn.edu

1.1. Relevant Literature & Theoretical Perspectives

In the setting of a large university, sections of calculus are commonly taught by educators with a broad variety of teaching experience and career orientations. Establishing some coordination between the sections is a logical solution to control for variation in the content covered and to promote effective instruction (Rasmussen & Ellis, 2015). The role of the course coordinator can be significant in this coordinated system. For example, coordinators can serve as change agents, fostering the widespread implementation of active learning strategies by the instructors they work with (Rasmussen & Ellis, 2015; Mesa et al., 2020; Williams et al., 2022).

However, there is great variation in how coordinators approach their role. Martinez et al (2021) identified two common coordination orientations: Humanistic-Growth and Resource-Managerial. The first mindset is community-based and focuses on the growth of the instructors teaching the coordinated course, while the latter is material-based and focuses more on the logistics of managing the course and course resources. Compared to coordinated chemistry and physics courses, US-based postsecondary mathematics instructors are more likely to experience *controlled coordination* (which we classify as strict Resource-Managerial type coordination), meaning they are mandated to utilise common course elements and are provided no opportunity to give their input on those items or to participate in decision making (Couch et al., 2023). The most effective coordination likely requires a balance between the two orientations. For example, institutions that implement uniform course components in conjunction with regular instructor meetings have been found to be more successful concerning both overall passing rates and student persistence in Calculus II (Rasmussen et al., 2019).

An additional promising strategy to increase the efficacy of coordinated Calculus instruction is the development and use of a set of common learning outcomes within the coordinated community (Zazkis & Nuñez, 2015; Williams et al., 2022). In their 2015 work, Rasmussen & Ellis identified a Calculus I coordinator who had mapped the uniform homework problems to a set of learning objectives. An instructor working with this particular coordinator found the mapping useful because it communicated what their department thought was important for students to learn and helped them align their in-class instruction to the homework. We consider this an example of a Resource-Managerial approach to coordination. This paper aims to introduce and evaluate a Humanistic-Growth approach to fostering course alignment through learning objective mapping.

1.2. Research Questions

Within our study, we consider the role of common learning objectives in supporting overall course alignment by surveying our participants before and after they have a set of common learning outcomes and they reflect on alignment using a novel instrument called the *Course Alignment Analysis Tool (CAAT).* CAAT provides a quantitative measurement of alignment, capturing the difference between the learning outcomes that an instructor feels should be prioritised and the learning outcomes most emphasised by an assignment. The goal of CAAT is to provide instructors with a new instrument for reflecting on the design of existing assignments or assessments. In this paper, we will first explain the graph theoretic underpinnings of CAAT and how to use the instrument. Then, we will answer the following research questions concerning our participants' perceptions of coordinated online homework and CAAT as a reflective tool.

(1) How do university mathematics instructors working in a coordinated calculus environment define the ‘quality’ of uniform homework assignments?
(2) Does the use of the CAAT in conjunction with a set of common learning objectives support a Humanistic-Growth coordination style (i.e. did our instructors evolve through participation)?
(3) Does CAAT provide evidence of agreement among university mathematics instructors as to the alignment of homework assignments to a common set of objectives?

2. Development of CAAT

2.1. Backward (Re)Design

The design of CAAT is partly inspired by an instructional design concept called Backward Design. This approach begins with developing a list of learning objectives that students are expected to achieve throughout a course and then constructing the rest of the course around this list of learning objectives, taking care that all course components are aligned (Fink, 2013). Centers for teaching and learning at universities around the United States have developed tools to aid instructors in creating courses using the Backward Design approach (Duke University Learning Innovation, 2020; Scott, 2023; University of Michigan Center for Academic Innovation, 2020; Yale Poorvu Center for Teaching and Learning, 2017). However, these resources imply that the user is creating their course components from scratch without utilizing any existing resources for assistance. On the other hand, we designed CAAT for use in the Backward *re*-design of mathematics courses, where the alignment of existing course components is analyzed.

2.2. The Math Behind the CAAT

Some concepts from Combinatorics and Graph Theory are central to the CAAT’s design. We explain these concepts in the context of an instructor using the CAAT to analyze an assignment. For reference, a blank copy of the CAAT has been included in Appendix A.

A ranking where ties are allowed is called a *weak ordering*. For instance, horse races allow for horses to tie for placement so that two horses might both finish in second place. In voting systems where voters rank candidates by preference, voters can rank multiple candidates the same rank indicating that they have no preference for one candidate over another. Because an instructor may find two learning objectives equally important, the user begins the CAAT by ranking a list of learning objectives they have deemed should be assessed, using a weak ordering.

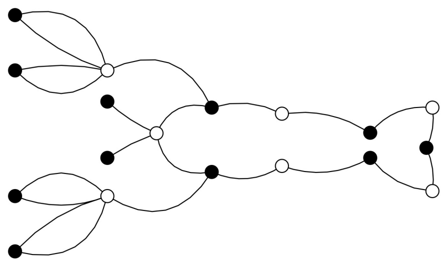
Suppose there are n learning objectives an instructor wants to cover. The instructor orders the objectives and assigns a ‘desired’ ranking according to what objectives they hope are emphasised the most. For instance, if $n = 4$, then the instructor might form the ranking $r_{\mathrm{de}} = (2, 1, 2, 4)$ indicating that the second objective should be emphasised the most, the first and third objectives should receive second priority equally, and the fourth objective should be emphasised the least.

After initially ranking their desired emphasis of each objective, the user completes the Problem-Objective Grid by associating each problem with one or more learning objectives and weighting this association according to the number of times an objective is assessed in a problem. For example, if a problem contains three parts with the same set of instructions for each part, then each objective assessed by the problem would receive a weight of '3'[1].

The Problem-Objective Grid can be viewed as a weighted association matrix of a bipartite multigraph that models the homework assignment in question. An example of a bipartite multigraph is shown in Figure 2.2. The Lobster is bipartite because no

Figure 1.

An example of a bipartite multigraph, which we call 'The Lobster'



two black vertices share an edge and no two white vertices share an edge, and it is a multigraph because multiple edges may exist between pairs of vertices (e.g. the claws).

After the user has completed the Problem-Objective Grid, they sum the columns of the grid. The column sums are then ranked from greatest to least, with the greatest column sum receiving rank 1. Returning to our example with 4 learning objectives, an 'observed' ranking with $n = 4$ might be $r_{\text{ob}} = (4, 1, 2, 3)$ indicating that the second objective appeared the most, and so forth.

Lastly, the user compares the observed and desired rankings by computing the standard Euclidean distance[2] (or root squared error) between the two rankings. Keeping with the previous examples, the measured disparity between the two rankings is

$$d(r_{\text{de}}, r_{\text{ob}}) = \sqrt{(2-4)^2 + (1-1)^2 + (2-2)^2 + (4-3)^2} = \sqrt{5} \approx 2.24.$$

In this way, the instructor can measure how aligned an assignment is to their desired emphasis of learning objectives using a mathematically defined distance, where a smaller distance between ranking vectors indicates better alignment.

2.3. The Rating Scale

To help instructors interpret the CAAT's numerical score, we developed a rating scale that translates the numerical alignment score obtained by the CAAT into one of three ratings: 'Excellent', 'Acceptable', or 'Poor'.

---

[1] A video of a fully worked example CAAT is available at `https://aub.ie/o3BsDL`

[2] Note that since the rankings are regarded as vectors in $\mathbb{N}^n$, the choice of metric (e.g., standard Euclidean distance) can be changed. Using the L^∞ distance would measure the maximum discrepancy between the ground truth and observed rankings, while the L^1 distance (or absolute error) is less sensitive to large discrepancies. For completeness, the discrepancy from our example would be

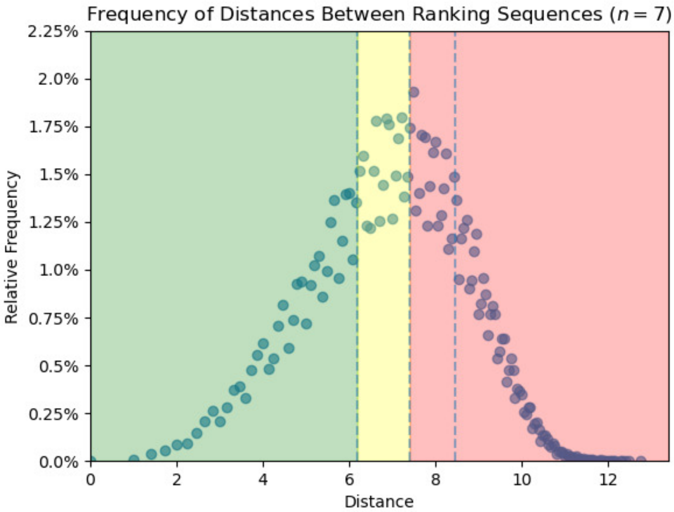
$$d_1(r_{\text{de}}, r_{\text{ob}}) = |2-4| + |1-1| + |2-2| + |4-3| = 3, \text{ and}$$
$$d_\infty(r_{\text{de}}, r_{\text{ob}}) = \max\{|2-4|, |1-1|, |2-2|, |4-3|\} = 2,$$

where d_1 is the L^1 distance and d_∞ is the L^∞ distance.

To construct the Rating Scale, we first found the distribution of all possible CAAT scores for each number of objectives. To do this, we developed Python code[3] that calculates the Euclidean distance between all pairs of weak orders of length n and outputs the relative frequency distribution of the distances along with quartile values; see Figure 2.3 for $n = 7$ objectives.

Figure 2.

The relative frequency of distances between weak orderings on 7 objectives



Note. The quartiles are delimited by the dashed lines.

The quartile ranges correspond to our final alignment rating categories; the lowest 25% of distances receive a rating of ‘Excellent’, the distances between the first and second quartiles receive a rating of ‘Acceptable’, and the highest 50% of distances receive a rating of ‘Poor’. Using quartiles to delineate rating categories ensures that each rating category has an equal probability of being used — in our rating scheme, an assignment has a 50% chance of a positive alignment rating (Excellent/Acceptable) and a 50% chance of a negative alignment rating (Poor).

We were able to construct rating scales for $n \in \{4, 5, 6, 7\}$, after which the number of pairs of weak orders of length n becomes computationally intractable[4]. However, we are currently working on improving the code to handle $n = 8$ and possibly more, which could extend the use of CAAT to analyze course components other than homework (see Section 7).

[3]GitHub: `https://github.com/TheGrateSalmon/CAAT`

[4]For $n = 7$, there are $\binom{64}{2}$ weak orders to compare, amounting to $1,161,841,311$ distances to compute. Moving to $n = 8$, there are $\binom{128}{2}$ weak orders to compare, amounting to $152,377,145,815$ distances to compute, a substantial increase over $n = 7$.

3. Study Design

After securing institutional review board approval, we invited all Graduate Teaching Assistants and faculty employed by a large R1 public university's Department of Mathematics & Statistics to participate in our study. 10 people meeting this criteria expressed interest.

The 10 participants in this study consisted of three faculty members, (who we call Lasairfhíona, Siobhán, and Méabh) and seven graduate student instructors or GSIs (who we call Séamus, Alastar, Mícheál, Eoin, Íde, Máire, and Ciarán). Note that these are Irish pseudonyms. The participants reported varying levels of experience with serving as instructor of record, recitation leader, and/or tutor for a Calculus I course; three participants (Alastar, Mícheál, Lasairfhíona) had no experience with Calculus I, three (Séamus, Siobhán, Ciarán) had 1-2 semesters of experience, two (Eoin, Íde) had 3-4 semesters, 1 (Méabh) had 5-6 semesters, and one (Máire) had 7-8 semesters.

Participants completed two phases of data collection. In Phase 1, participants were provided copies of five Calculus I homework assignments and the textbook section(s) corresponding to each assignment. This sample of assignments was chosen to provide variety in length, content, and question formatting. The participants then completed a Pre-CAAT survey that asked them to rate the quality of each assignment on a 5-point Likert scale and explain the rationale behind each of their Likert ratings.

In Phase 2, participants were provided a CAAT with a list of six to seven preselected objectives for each homework assignment. Since the homework assignments were not created using a list of learning objectives, the preselected objectives were adapted from the College Board's AP Calculus AB content objectives (Mui & Tully, 2020), which have been thoroughly vetted by a combination of both university educators and high school teachers. After analyzing the five homework assignments using CAAT, the participants completed a Post-CAAT survey with questions that mirrored those on the Pre-CAAT survey and, additionally, asked if the CAAT influenced their Post-CAAT quality ratings of the homework assignments.

4. Qualitative Results

We answered Research Questions 1 and 2 using qualitative methods. Subsection 4.1 contains our analysis, and Subsection 4.2 discusses our answers to Research Questions 1 and 2.

4.1. Data Analysis

To analyze participants' free responses to their Likert Scale ratings, we used open coding (Thornberg & Charmaz, 2014). During analysis of the Pre-CAAT surveys, the following codes emerged, each containing both a positive and negative sub-code.

Coverage of Topics references the degree to which an assignment includes questions that are discussed in the textbook. Specifically, this code includes broadly stated comments that do not mention specific topics or objectives.

Narrative/Progression is attention to either situating a topic within a broader mathematical context or the ordering of topics or questions on an assignment. For example, in the context of assigning computationally dense problems, Íde noted 'We lose the forest for the trees and in the process break the storyline of the course for tedium and busy work,' which implies that Calculus is a small part of a larger mathematical narrative. Máire specifically focused on the progression of topics from Precalculus to Calculus stating, 'It is unnecessary to mention names of curves (cardioid) because conic sections and polar graphs are not covered in Precalc here.'

Question Format was used for comments referencing either the phrasing of homework questions, the appearance of student answers, or the anticipated 'hacking behaviors' of students. For example, Alastar commented, 'The multiple choice format is even worse then [sic] the regular fill in the box format because it incentivizes not doing any work and getting through the assignment by blind guessing.'

Difficulty manifested through three foci: (1) **Computational Complexity**, (2) **Very procedural, lacking conceptual**, and (3) **Lack of higher-order thinking**. Regarding computational complexity, Íde noted, 'We are asking incredibly tedious questions to the students', and on the procedural nature of the homework, she remarked, 'I think this one is very repetitive, but I would call this section the methodical backbone of calculus.' Finally, regarding the lack of higher-order thinking, Lasairfhíona said, 'No assessment is done that requires a bit higher cognitive thinking.'

Discussion of a specific problem. This code was used for instances in which participants indicated a like or dislike for specific problems by their number. For example, Eoin noted, 'a few of the problems are unnecessarily hard, the biggest culprits being problems #8 and #9.'

Assignment Length. This code was used for instances in which participants commented on the amount of time they anticipated an assignment would take students to complete or the number of questions present on the assignment.

Table 4.1 gives the frequency of each of the homework quality codes, disaggregated by positive and negative subcodes.

Table 1.

Frequencies of homework quality codes pre- and Post-CAAT

No.	Code Title	Freq. (Pre)	Freq. (Post)
1	Coverage of Topics (-)	13	5
2	Narrative/Problem Progression (-)	17	5
3	Question Format (-)	37	18
4	Very Procedural, Lacking Conceptual (-)	13	1
5	Computational Complexity (-)	22	4
6	Lack of higher-level thinking (-)	5	1
7	Disgruntled with one specific problem	26	1
8	Assignment Length (-)	16	10
9	Coverage of Topics (+)	21	13
10	Narrative/Problem Progression (+)	2	2
11	Question Format (+)	0	0
12	Very Procedural, Lacking Conceptual (+)	4	3
13	Computational Complexity (+)	3	0
14	Lack of higher-level thinking (+)	0	0
15	Loved one specific problem	1	0

In our analysis of the Post-CAAT surveys, the following additional codes emerged concerning homework quality.

Discussion of a specific objective. Similar to 'Coverage of Topics' before the use of CAAT, this code was used to denote comments that mentioned the material covered by an assignment. However, in 'specific objective' comments, participants used language that specifically referenced an objective listed on the CAATs rather than the more general language denoted by the 'Coverage of Topics' code. For example, before using CAAT, Ciarán commented about one of the assignments, 'I found questions 13 and 27 quite surprising ... I don't recall teaching the normal line in Calculus I ... but it is also not present in the textbook section,' which was coded as 'Coverage of Topics'. After using CAAT, he said the following about the same assignment

> I've now realized that the derivatives homework really tested a ton of things that weren't really focused on in the textbook, and it almost completely ignored exponential functions except in the context of other problems. I still liked how much it drilled the power rule[5].

Fostering careful reflection. This code was used for responses containing language with contemplative connotations. A few different phrases indicating this code are 'I've now realized...', 'the homework did' or 'did not give significant attention to...', and 'the homework gave a pretty good balance'.

Gut-check refers to responses in which CAAT scores reaffirmed the user's instinctive opinion about an assignment's quality.

Complimentary to other measures of homework quality. This code is best characterised by a response from Méabh

> In my Pre-CAAT assessment ... my primary consideration was the mechanics of completing a homework problem ... The CAAT provided me a complimentary way of thinking about homework quality, one that focused on the overall composition of problems and which objectives were covered.

Table 4.1 gives the frequency of each of the CAAT Influence codes that emerged after participants interacted with the homework assignments using CAAT.

Table 2.

Frequencies of CAAT influence codes Post-CAAT

No.	Code Title	Freq. (Post)
16	Coverage of Specific Objective (-)	27
17	Coverage of Specific Objective (+)	9
18	Complementary measure of quality	1
19	Fostered Careful Reflection	28
20	Gut Check	1

To disaggregate the codes by participant, we constructed a radar chart for each participant's Pre-CAAT and Post-CAAT survey responses. We coloured regions of the radar charts to denote negative subcodes (red), positive subcodes (green), and CAAT Influence codes (yellow). Codes are numbered the same as in Tables 4.1 and 4.1. All 10 participants' radar charts showed changes from Pre-CAAT to Post-CAAT. We

[5]Finding the derivatives of exponential and power functions were listed as objectives on this assignment's CAAT. Additionally, Ciarán did not mention the power rule in his pre-survey response.

highlight these changes through three participants, Mícheál, Lasairfhíona, and Máire.

Mícheál. Recall that Mícheál was a GSI with no experience as an instructor of record, recitation leader, or tutor for Calculus I. His Pre-CAAT and Post-CAAT radar charts are given in Figure 4.1. In his initial ratings of homework quality, Mícheál focused largely on specific problems that he disliked. However, in his Post-CAAT ratings, his outlook on the homework became slightly more positive (there is 1 point more in the Post-CAAT green region than the Pre-CAAT green region). Furthermore, his Post-CAAT chart contains two more non-zero points than his Pre-CAAT chart, suggesting that overall, his definition of homework quality became more multi-faceted.

Figure 3.

Mícheál's Pre-CAAT and Post-CAAT radar charts

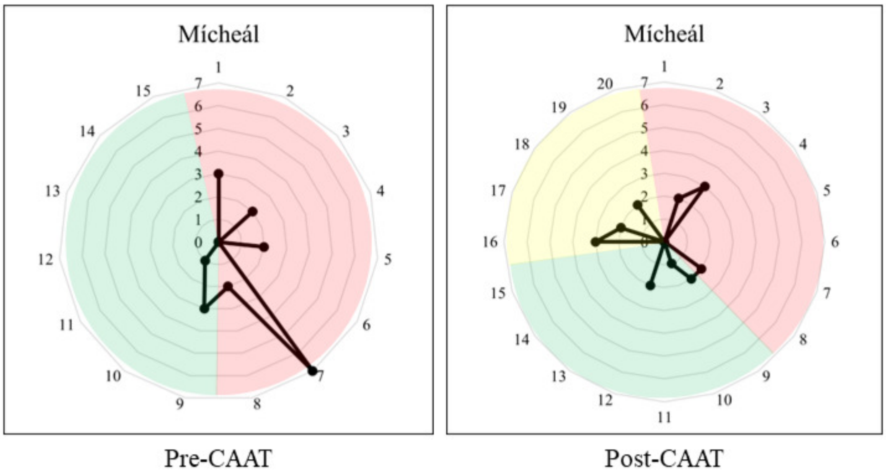
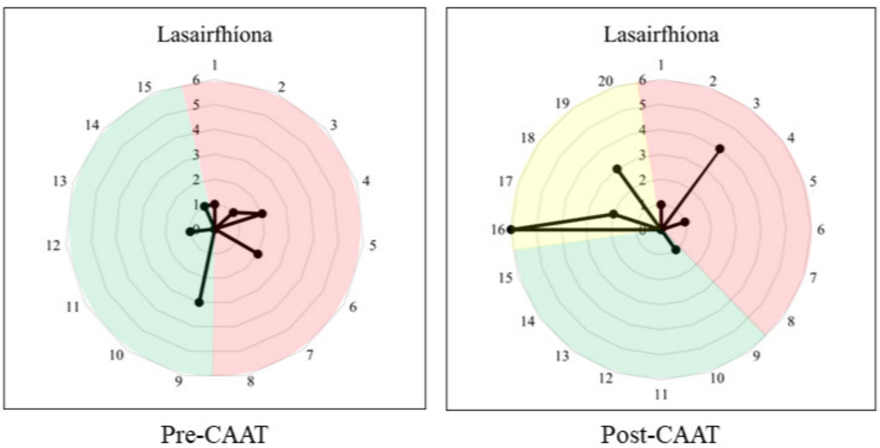


Figure 4.

Lasairfhíona's Pre-CAAT and Post-CAAT radar charts

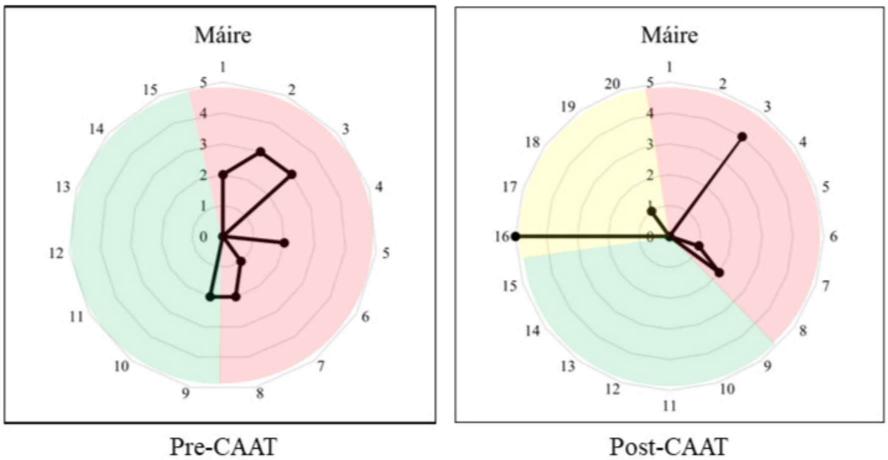


Lasairfhíona. Recall that Lasairfhíona was a faculty member with no Calculus I experience. However, her case was unique because, as a faculty member, she had much more total experience teaching other courses than the GSIs with no Calculus I experience. Her Pre-CAAT and Post-CAAT radar charts are given in Figure 4.1. Unlike Mícheál, Lasairfhíona's definition of quality did not become increasingly multi-dimensional. Rather, her two charts contained the same number of non-zero points, with some of her negative sentiment shifting to deeper reflection and a heightened focus on objectives. Additionally, her Post-CAAT chart contained far more total instances of codes than her Pre-CAAT chart, suggesting she did experience growth.

Máire. Recall that Máire was a GSI with 7-8 semesters of Calculus I experience, the most of any participant. Her Pre-CAAT and Post-CAAT radar charts are given in Figure 4.1. Given that Máire's Calculus I experience was on the opposite end of our experience spectrum to Mícheál's and Lasairfhíona's experience, it was interesting that her charts were also opposite to the other two. Like Lasairfhíona, Máire's Pre-CAAT chart showed a preexisting multi-dimensional definition of homework quality. However, in her Post-CAAT chart, her attention narrowly focused on the negative aspects of Question Formatting and specific objectives.

Figure 5.

Máire's Pre-CAAT and Post-CAAT radar charts



4.2. Discussion

RQ 1. How do university mathematics instructors working in a coordinated calculus environment define the 'quality' of uniform homework assignments?

Though the CAAT specifically focuses on alignment, the results of our qualitative analysis of the rationale behind participants' Likert scale quality ratings suggest that instructors working in a coordinated calculus environment have a multi-dimensional definition of homework quality which grows with experience.

Furthermore, by taking the five codes with the highest frequency between Pre-CAAT and Post-CAAT ratings, we have determined that the most important factors instructors consider when defining homework quality are:

(1) *Alignment.* Coordinated instructors believe that homework should reflect the content students are exposed to in their textbooks, which should also be reflected in lectures.
(2) *Task Difficulty.* Coordinated instructors believe that one aspect of good homework is including both procedural and conceptual tasks that students will find approachable to build their confidence before progressing to more difficult tasks.
(3) *Question Format.* Coordinated instructors believe that the way questions are formatted is related to the theme of alignment but in the vein of general mathematical practices rather than content specifics. A prominent issue that emerged among our participants was that online homework in particular does not require students to show work, which is the opposite of what is expected on an exam.
(4) *Narrative.* Coordinated instructors view mathematics as a story, and it follows that assignments should tap into this instinct by also telling a story, whether through a gradual increase in task difficulty or by situating assignments within a broader curricular context.
(5) *Specific Problems.* Coordinated instructors believe that specific problems they don't like can make or break a good homework assignment.

RQ 2. Does the use of the CAAT in conjunction with a set of common learning objectives support a Humanistic-Growth coordination style (i.e. did our instructors evolve through participation)?

In short, yes. Every participant was influenced by CAAT whether they realised it or not. Even Séamus, who significantly disagreed with most other participants and who stated that CAAT did not influence him at all, used the word 'objectives' rather than 'topics' in his Post-CAAT ratings. For the others, the influence was more pronounced; all nine other participants stated that CAAT influenced their Post-CAAT ratings, and this was reflected in their radar charts (all participants' Post-CAAT charts contained points in the yellow region).

Furthermore, the influence of CAAT manifested in different ways. More novice Calculus I instructors (such as Micheál) saw growth from their Pre-CAAT to Post-CAAT responses, both in their definition of 'quality' and their instances of coded phrases. On the other hand, more experienced Calculus I instructors (such as Máire) saw the opposite effect. From Pre-CAAT to Post-CAAT, they became more narrowly focused on specifics. The evidence of evolution in all participants suggests that using CAAT to onboard instructors to a coordinated setting could support a Humanistic-Growth coordination style in mathematics courses.

5. Quantitative Results

To answer Research Question 3, we used Quantitative Methods. Subsection 5.1 contains our analysis, and Subsection 5.2 discusses our answer to Research Question 3.

5.1. Data Analysis

To determine if there was any agreement among participants' CAAT scores, for each participant $p = 1, \ldots, 10$, we defined a CAAT Score Vector

$$\vec{C}_p = \langle s_1, s_2, s_3, s_4, s_5 \rangle$$

where s_i is the CAAT score obtained by participant p for homework assignment $i \in \{1, 2, 3, 4, 5\}$. Then, we computed the *(weighted) Jaccard index*, J, between all 45 pairs of participants,

$$J = \frac{\sum_i \min\{x_i, y_i\}}{\sum_i \max\{x_i, y_i\}}$$

where the x_i are the components of the first participant's CAAT score vector and the y_i are the components of the second participant's CAAT score vector. Values of J range from $[0, 1]$ with values closer to 1 indicating more agreement and values closer to 0 indicating less agreement. The Jaccard index has been used previously to validate educational instruments (Smith et al., 2013). The five highest and the five lowest Jaccard indices among our participants are presented in Table 5.1.

Table 3.

Participant pairs with the 5 highest and 5 lowest weighted Jaccard indices

Participant Pair	Five Highest J	Participant Pair	Five Lowest J
Mícheál, Íde	0.860	Séamus, Máire	0.504
Siobhán, Máire	0.853	Mícheál, Méabh	0.51
Siobhán, Ciarán	0.824	Séamus, Lasairfhíona	0.515
Lasairfhíona, Siobhán	0.793	Séamus, Siobhán	0.515
Lasairfhíona, Ciarán	0.790	Íde, Méabh	0.52

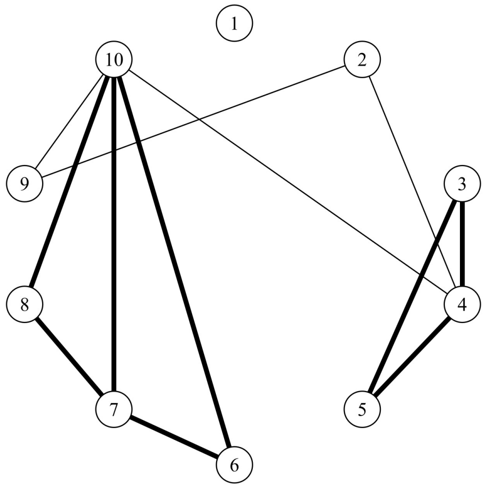
To contextualise the weighted Jaccard indices, we estimated a random mean Jaccard index by Monte Carlo method in Microsoft Excel (200 iterations). This yielded a value of 0.622 or 62%, while the mean of our computed Jaccard indices was 0.678 or 68%. Excluding Participant 1, who showed significant disagreement to the other participants (Participant 1 had less than 60% agreement with 5 out of 9 other participants), the mean of our computed Jaccard indices was 0.698 or 70%. This indicates that the 9 participants aside from Participant 1 showed moderate to high agreement in their CAAT scores.

To further explore agreement, we modeled the Jaccard indices indicating high agreement ($\geq 75\%$) with a graph (see Figure 5.1) and looked for sub-cliques (subgraphs with maximum edges between vertices), which could indicate some commonality among a group of participants. The largest sub-cliques in the graph were three different triangles: $(3, 4, 5)$, $(6, 7, 10)$, and $(7, 8, 10)$. These are bolded in Figure 5.1. Furthermore, the two triangles involving 7 and 10 are one edge away from forming a 4-clique (the edge between 6 and 8 is missing). This led us to ask if these high-agreement triangles might indicate the existence of common themes among participants' open-ended survey responses.

For the participants in each high-agreement triangle, we used open coding (Thornberg & Charmaz, 2014) on the participants' initial opinion of online homework (a ques-

Figure 6.

Pairs of participants with $J \geq 0.750$



tion on the Pre-CAAT survey) and quality ratings of the homework assignments. For quality ratings, we used the same codes defined in Subsection 4.1. However, when coupled with the participants' opinion of online homework, the following themes emerged.

In the first triangle $(3, 4, 5)$, all three participants' explanations of quality were detail-oriented, focusing on the complexity of procedures among the homework problems. They all referenced specific problem numbers, with one participant even noting that there was a more efficient way to solve one of the problems than was shown by the textbook.

In the second triangle $(6, 7, 10)$, all three participants focused on a lack of conceptual questions. They didn't all reference specific problem numbers like the first triangle but focused more on the assignments as a whole.

The third triangle $(7, 8, 10)$ was similar to the second triangle, which may explain why the four participants 6, 7, 8, and 10 nearly formed a 4-clique. The difference was that while all four participants noted a lack of conceptual questions, participants 7, 8, and 10 went further, noting that students both are cheated and cheat their way out of a deep conceptual understanding of the material due to question formatting. Furthermore, all three of these participants focused on an assignment's progression and how the assignment was situated in the overall course narrative as a whole.

The large difference in focus between the first triangle and the other triangles (procedural complexity versus conceptual understanding) may also help explain why these three triangles formed two distinct subgraphs with one edge connecting them in the original graph (the edge between 4 and 10). Participant 10 was the only member of the 'conceptual group' having any high agreement with the 'procedural complexity group'. This may have occurred because Participant 10 also made a few comments about procedural complexity in addition to focusing on conceptual understanding.

5.2. Discussion

RQ 3. Does CAAT provide evidence of agreement among university mathematics instructors as to the alignment of homework assignments to a common set of objectives?

In short, yes and no. From our quantitative analysis, most participants were split between a focus on the procedural versus conceptual nature of homework, which could indicate that similar teaching philosophies may be a factor in high-agreement groups. Additionally, most pairs of participants had a non-random agreement. However, instances of significant *disagreement* also existed among the participants. Again, we suspect differences in instructors' teaching philosophies may account for this.

The appearance of both agreement and disagreement provides further evidence for the case of using CAAT as a professional development tool for coordinated course instructional teams. By proactively identifying disagreement amongst team members, a Humanistic-Growth approach to coordination would seek to mediate these differences through group discussion and community-building.

6. Study Limitations

This study examined 10 instructors within the same department, five of whom only have experience with their current institution's instantiation of Calculus I. Additionally, this version of Calculus I only involves homework assigned through an online platform.

7. Additional Prospective Applications

The implications of our data open up many possible avenues of study including a deeper dive into homework modalities, new approaches to qualitative analysis, and improving our Python code.

Online vs. Paper Homework? Online homework has its own set of unique factors. When interacting with online homework assignments, students use online sources outside their course materials for assistance (Dorko, 2020) and are more likely to use solution methods shown in the homework software's help features than those demonstrated in lecture (Dorko & Cook, 2022). How might students' interactions with homework and course materials differ for homework that is not assigned online? The results of this study might differ (1) if the study was replicated so that instructors complete study procedures using the online homework platform instead of printed copies of online assignments (giving them a more realistic experience from a student point-of-view) or (2) if the study was replicated using homework assigned from a physical textbook rather than online software.

New Approach to Qualitative Analysis? From our graph-theoretic analysis of participants' CAAT score agreement, there is evidence to suggest that if two instructors have a high agreement in CAAT scores, then they likely focus on similar aspects of a homework assignment's design. Could this method be used more broadly in qualitative data analysis as a preliminary tool for detecting commonalities between study participants?

Improving the Code? If we can improve the Python code to handle larger objective lists, then we could extend the use of CAAT to allow instructors to reflect on the design of other course components that involve larger numbers of objectives such as exams, quizzes, lecture notes, or even recordings of live lectures. However, with the current format of CAAT, improving the Python code may not be feasible. Thus, this goal may require reformatting CAAT so that the score is a probability rather than a distance[6].

References

[1] Couch, B. A., Prevost, L. B., Stains, M., Whitt, B., Marcy, A. E., Apkarian, N., Dancy, M. H., Henderson, C., Johnson, E., Raker, J. R., Yik, B. J., Earl, B., Shadle, S. E., Skvoretz, J., & Ziker, J. P. (2023). Examining whether and how instructional coordination occurs within introductory undergraduate STEM courses. Frontiers in Education, 8, 1156781. https://doi.org/10.3389/feduc.2023.1156781

[2] Dorko, A. (2020). Red X's and Green Checks: A Model of How Students Engage with Online Homework. International Journal of Research in Undergraduate Mathematics Education, 6, 446–474. https://doi.org/10.1007/s40753-020-00113-w

[3] Dorko, A., & Cook, J. P. (2022). Why do students rely on online homework over lecture? In S.S. Karunakaran & A. Higgins (Eds.) Proceedings of the 24th Annual Conference on Research in Undergraduate Mathematics Education (pp. 951–957) Boston, MA.

[4] Duke University Learning Innovation (2020). Online Course Planning Worksheet. Duke University Flexible Teaching. Retrieved April 16, 2023, from https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fflexteaching.li.duke.edu%2Ffiles%2F2020%2F06%2FCourse-Design-Plan-F20-Self-Service-Guide.docx&wdOrigin=BROWSELINK

[5] Ellis, J., Hanson, K., Nuñez, G., & Rasmussen, C. (2015). Beyond plug and chug: An analysis of calculus I homework. International Journal of Research in Undergraduate Mathematics Education, 1(2), 268–287. https://doi.org/10.1007/s40753-015-0012-z

[6] Fink, L. D. (2013). Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses. John Wiley & Sons.

[7] Martinez, A. E., Gehrtz, J., Rasmussen, C., LaTona-Tequida, T., & Vroom, K. (2022). Course coordinator orientations toward their work and opportunities for professional development. Innovative Higher Education, 47(2), 327-346.

[8] Mesa, V., Shultz, M., & Jackson, A. (2020). Moving Away from Lecture in Undergraduate Mathematics: Managing Tensions within a Coordinated Inquiry-Based Linear Algebra Course. International Journal of Research in Undergraduate Mathematics Education, 6(2), 245–278. https://doi.org/10.1007/s40753-019-00109-1

[9] Mui, S., & Tully, B. (Eds.). (2020). AP Calculus AB and BC Course Description and Exam Guide. The College Board.

[10] Rasmussen, C., Apkarian, N., Hagman, J. E., Johnson, E., Larsen, S., & Bressoud, D. (2019). Brief report: characteristics of precalculus through calculus 2 programs: insights from a national census survey. Journal for Research in Mathematics Education, 50(1), 98-111.

[11] Rasmussen, C., & Ellis, J. (2015). Calculus coordination at PhD-granting universities: more than just using the same syllabus, textbook, and final exam. In D. Bressoud, V. Mesa, and C. Rasmussen (Eds.) Insights and recommendations from the MAA national study of calculus. MAA Press.

[12] Scott, C. (2023). Course Alignment Tool. Stony Brook University Center for Excellence in Learning and Teaching. Retrieved April 16, 2023, from https://www.stonybrook.edu/celt/_pdf/CourseAssessment_and_AlignmentTemplateTool.pdf

[6] The authors would like to acknowledge our colleagues, Jordan Eckert and Padmini Nukala, for this idea.

[13] Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. CBE-Life Sciences Education, 12(4), 618-627.
[14] Thornberg, R. & Charmaz, K. (2014). Grounded Theory and Theoretical Coding. In U. Flick (Ed.), The SAGE Handbook of Qualitative Data Analysis (pp. 153–169). SAGE.
[15] University of Michigan Center for Academic Innovation (2020, July 8). Online Course Blueprint Planning Guide - v 1.4. University of Michigan Online Teaching. Retrieved April 16, 2023, from https://docs.google.com/spreadsheets/d/1EHnNeHLn5JcjvCGo0_9uTQe9sBAYoO3fdH2gvYDvXdk/edit#gid=302108820
[16] Weston, T. J., Seymour, E., Koch, A. K., & Drake, B. M. (2019). Weed-Out Classes and Their Consequences. In E. Seymour & A.B. Hunter (Eds.), Talking about Leaving Revisited (pp. 197–243). Springer International Publishing. https://doi.org/10.1007/978-3-030-25304-2_7
[17] Wiggins, G. P., & McTighe, J. (1998). Understanding by Design. Association for Supervision & Curriculum Development; eBook Collection (EBSCOhost). http://spot.lib.auburn.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=49873&site=eds-live&scope=site
[18] Williams, M., Apkarian, N., Uhing, K., Martinez, A. E., Rasmussen, C., & Smith, W. M. (2022). In the driver's seat: Course coordinators as change agents for active learning in university Precalculus to Calculus 2. International Journal of Research in Undergraduate Mathematics Education, 8(1), p. 121-148.
[19] Yale Poorvu Center for Teaching and Learning (2017, January 1). Teaching and Learning Frameworks. Retrieved April 16, 2023, from https://poorvucenter.yale.edu/BackwardDesign
[20] Zazkis, D., & Nuñez, G. (2015). How departments use local data to inform and refine program improvements. In D. M. Bressoud, V. Mesa, & C. L. Rasmussen (Eds.). Insights and recommendations from the MAA national study of college calculus. (pp. 123-129.) MAA Press.

Appendix A. Blank Copy of the CAAT

width=!,height=!,pages=1-2