On the Ground with Active and Collaborative Learning in STEM Gateway Courses: Findings from Classroom Observations, Year 6

George Mason University Gateway 2 STEM Project

Sandra Laursen
Ethnography & Evaluation Research, University of Colorado Boulder

1. Introduction and Objectives

Gateway 2 STEM, an NSF-supported project at George Mason University (GMU), seeks to incorporate active and collaborative learning (ACL) into gateway courses in math, physics and computer science (CS). All three departments are implementing ACL in recitations, offered as a separate hour of instruction to support a lecture, and in some blended lecture-recitation courses.

I visited campus to observe the extent and nature of ACL approaches used in these gateway courses, seeking to assess progress on the project's goals to implement ACL as a routine feature of gateway course recitations in math and physics and in CS labs. I had the following specific objectives for observation:

a) to characterize the typical forms and variability of ACL as implemented in these courses
b) to offer a judgment of the quality of these approaches in relation to established findings from education literature about effective ACL practices
c) to gather observation evidence that could be compared with evidence from student and instructor surveys that were also administered in fall 2023.

In this report, I address objectives (a) and (b), and will discuss (c) in conjunction with survey data analyses reported elsewhere.

2. Study Methods

In this section, I summarize terminology used in this report, sampling and procedures for the observation, and analysis methods. This is a primarily qualitative analysis, complemented by simple metrics to categorize and quantify the use of class time.

2.1. Terminology used here

Recitations in math and physics, and *labs* in CS, offer students review, practice and personal help with course material in smaller groups than the large lecture hall. *Blended classes* do not differentiate lecture from recitation periods; most use a mixture of lecture-based and interactive modes. Specific math and CS courses, and some physics sections, use blended formats.

I use *instructors* to name all those with a classroom teaching role. They include *faculty* (in tenure-stream or term appointments), *graduate teaching assistants* (GTAs), undergraduate *learning assistants* (LAs, used in math and physics) and *undergraduate teaching assistants* (UTAs, used in CS). The classroom roles of LAs and UTAs are here assumed to be the same, but difference in how they are hired, paid and trained may affect how they understand and execute their roles. All in-class group tasks are here called *worksheets*, whether described by instructors as a guided inquiry worksheet, tutorial, or programming assignment.

2.2. Sampling

Using GMU course schedule data, I identified the universe of weekly class meetings for the project-designated gateway courses: 107 sessions of recitations or blended classes (M 8 am to F 5 pm). When classes met twice weekly, I counted both (e.g., both M and W meetings of a MW class). The total of 107 includes a few online meetings to which I had no access.

My physical visit occurred from 5 PM on Monday 11/6 to 2:30 PM on Friday 11/10 at 2:30 PM. During that time, 67 face-to-face meetings occurred that did not conflict with other events on my schedule, such as PI team meetings. Many classes met at overlapping times, so I optimized my daily schedule by applying these parameters, in order:

a) Observe the most possible instruction in the available hours
b) Observe at least one meeting of each gateway course in math, physics and CS (Table 1)
c) Observe as many unique instructors as possible
d) Optimize transition time among buildings.

Table 1 summarizes the resulting observation sample, and is further discussed below. Courses listed in italics represented blended classes, and the others are true recitations.

Table 1: Summary of observation sample and class characteristics

	All	Math	Physics	CS
All weekly mtgs *% by dept*	107 100%	49 46%	39 36%	19 18%
Available in visit *% by dept*	67 100%	33 49%	19 28%	15 22%
Observed in visit (% of available) *% by dept*	22 (33%) 100%	13 (39%) 59%	5 (26%) 23%	4 (27%) 18%
Observed, by course, of available mtgs	11 courses in 3 depts	Math 113: 6 of 13 Math 114: 2 of 7 *Math 123: 4 of 9* *Math 124: 1 of 4*	Phys 160: 2 of 7 Phys 260: 1 of 4 Phys 243: 1 of 5 Phys 245: 1 of 3	*CS 108: 1 of 2* CS 109: 1 of 2 CS 112: 2 of 11
Lead instructors	25	8 GTAs 5 term faculty	3 term faculty 2 tenured faculty	3 GTA teams 1 term faculty
Sessions with LAs or UTAs	18 of 22	12 of 13	3 of 5	2 UTA +GTA lead 1 GTA +instr lead
Class length		50 min: 113,114 75 min: 123,124	50 min: all	50 min: 109,112 75 min: 108
Class size		up to 45: 113,114 up to 35: 123,124	up to 30	up to 60

In all I observed 22 class meetings, totaling 1250 minutes (20.8 hr) of instruction, meeting criteria (a) and (b). To meet criterion (b), I compromised once on criterion (c), observing one instructor twice. Otherwise I observed unique instructors or teams each time.

Eleven class meetings (of 29 possible) were taught by graduate teaching assistants (GTAs) working solo or in pairs, and eleven by faculty instructors (of 38 available). Math 113-114, CS 109 and CS 112 use GTAs to lead recitations (called labs in CS). Math 123-124 and CS 108 are taught by instructors with teaching-focused appointments, known as term faculty at Mason. In Physics, recitations are taught by tenured and term faculty. I identified undergraduate learning/teaching assistants (LAs, UTAs) in 18 of the 22 sections.

Sixteen class meetings were true recitations (including CS labs), distinct from lecture periods. Math recitations enroll up to 45 students, Physics recitations up to 30, and CS labs up to 60 students. Six class meetings were blended lecture-recitation courses, enrolling up to 35 students in Math and 60 students in CS.

On the whole, this is a robust sample of the project-targeted types of class meetings across 11 gateway courses in three departments. However, there are some limitations to generalizing from this sample. First, math sessions were modestly oversampled, relative to Physics and to CS 112. This arose from institutional scheduling parameters intended to maximize use of active learning spaces, and/or to sequence labs and recitations in relation to lecture, resulting in:

- high density of Physics recitations during the day on Monday, before I arrived
- high density of CS 112 labs on Friday, overlapping each other
- higher density of Math recitations in the late afternoon and early evening, when I could attend without other conflicts.

Moreover, my observations describe a single class meeting only. This is an appropriate sampling frame to characterize the suite of approaches used across the gateway courses as a group, but no conclusions can be drawn about the semester-long teaching that occurs in any course. I have no way to judge whether the specific sections I visited were typical or outlying examples.

2.3. Observation Procedures

Instructors were notified as a group by email that I might visit, but did not know if I would attend their particular class. I introduced myself to instructors at the start or end of class, or while students were working. I took copies of paper handouts so I could keep track of the ACL work, and sat where I could eavesdrop on 2-4 student groups and surveil the class as a whole.

I took field notes and did not use a formal observation protocol. In the field notes, I estimated overall attendance (including initial and late arrivals) and overall gender distribution. When students distributed into smaller groups, I recorded the numbers and visible characteristics of the groups, noting apparent gender presentation and some markers of appearance or clothing that I could use to distinguish students in my notes. I recorded the time at which different activities started and stopped (e.g., lecture, instructions, quiz, group work), using as base categories the same set of choices on surveys administered to students and instructors (see Section 3.1 and Table 2 for details). I noted instructor and student behaviors and interactions, tracked the flow of class activities, and recorded verbatim examples of instructor and student remarks. When possible, I spoke briefly with students, instructors, learning assistants (LAs) and undergraduate teaching assistants (UTAs), and graduate TAs (GTAs) during or immediately after class to ask about their practices and their experiences, introducing myself as a visitor from Colorado who was interested in active and collaborative learning and how it worked at Mason.

2.4. Data Analysis

To analyze the quantitative data, I converted field note time records into elapsed minutes and added these to a spreadsheet to compute the percentage of class time used for different activities, then conducted simple descriptive analyses of the mean and range. These analyses distinguish true recitations from blended sections because the purpose of these two types of classes is different and a different mix of activities is expected. Given small sub-sample sizes, I did not separate observations by department or course, in order to protect instructor identities.

For the qualitative analysis, I reviewed my field notes repeatedly to look for patterns linking instructor practices and student engagement, and to identify practices and conditions that align (or not) with effective ACL practices established in the literature. I sorted these practices based on how they contributed to the course, gathering them under themes such as facilitation tactics, grouping, accountability, and the physical environment. I selected examples of teacher prompts to initiate or probe elicit student thinking. These analyses (Section 3) describe the range and variation of what I saw, but they characterize ACL teaching in a somewhat atomistic way.

To offer a more holistic perspective on the quality of ACL teaching that I saw, I identified class sessions that combined particular practices in notable ways, and wrote these as short vignettes with brief case analyses, guided by a framework about collaborative learning from Karl Smith (1996). These vignettes (Section 4) are more analytical, drawing on research-based insights, and also more synthetic, in showing how ACL practices can combine in positively or negatively reinforcing ways. They may be useful as examples to share in professional development.

Finally, I spent time thinking how to identify a small set of core practices that could guide future work at GMU (Section 5). These practices could guide professional development for GTAs and faculty, inform LA/UTA training, or shape goals and discussion of ongoing faculty CCTs.

3. Findings and Interpretations

This section emphasizes descriptive analysis, first highlighting the class activities that I saw and the general use of class time, then some specific facilitation tactics and instructor behaviors that I observed during ACL work. I also characterize two structural features that offer essential supports for ACL at George Mason: the physical environments where these classes met and the use of peer instructors (LAs and UTAs). Their presence or absence can support or constrain class activities. In this section, I highlight comparisons with Liljedahl's 14 features of "thinking classrooms" (Liljedahl, 2020; n.d.; see also Boryga, 2023).

3.1. Description of Classes

Table 2 shows a quantitative summary of the general types of class activity I saw.

Table 2: Observed class activities for true recitation and blended class meetings

		true recitation/lab sections (16)		blended lec/rec sections (6)	
Definition	**Activity**	# meetings it occurred	mean % of class time spent (range)	# meetings it occurred	mean % of class time spent (range)
instructor gives lecture (formal + interactive)	**total lecture**	6	14% (0-80%)	6	47% (24-76%)
instructor summary, ≤5 min	**short explanations**	7	8% (0-28%)	1	1% (0-8%)
discussion in which students respond to peers	**whole class discussion**	0	0%	0	0%
students work on problems alone or in small groups	**total group+ individual work**	16	70% (20%-100%)	6	46% (24-76%)
students present their work to the whole class	**student presentations**	0	0%	0	0%
quiz, survey, business, announcements	**total other**	6	9% (0-30%)	1	10% (0-40%)

The categories were based on the same set of class activities that students and instructors reported on surveys. We have previously found these categories to be helpful in describing courses. For this analysis I modified the categories based on what I observed.

- Our survey instruments distinguish two main types of instructor talk by their function: lecture is pre-prepared instructor talk, and distinct from short, responsive explanations such as signposting or mini-lectures prompted by student questions or difficulties. Variants of lecture include formal lecture (involving little or no Q&A with students) and informal lecture (involving some Q&A). Most instructors used either formal or interactive lecture; a few used segments of each. Because the focus of observations was ACL, not lecture, I combined the two forms of lecture in Table 2.
- I combined small group work with individual seat or board work because several sections offered students a choice to work alone or together. The expectation to work in groups was explicit in 10 sections and optional in 5 sections. Four sections used individual seat work and 3 sections used individual board work.
- “Other” includes assessments (quizzes) and class business such as making announcements, handing back papers, and administering a student survey. “Other” time is unavailable for student interaction.

The data clearly show that using a substantial amount of class time for student work is common in both true recitations and blended classes. At least some of this work occurred in every class I visited, across math, physics, and CS courses. Both GTA-led and faculty-led true recitations had a strong norm of modest lecture time (mean 14%) and substantial time for group and/or individual problem-solving (mean 70%). At the high end of the lecture range, one first-year GTA used the worksheet as a structure for interactive lecture. Time for lecture was more prevalent in the blended class sessions (mean 47%) but generally well balanced against student work (mean 46%). While in previous studies of inquiry-based mathematics we have seen students presenting their own work to peers, often prompting whole-class discussion, these methods were not observed at GMU.

3.2. The Physical Environment

The physical environment can enhance or constrain how ACL work is facilitated and how students engage in it, so I discuss it before elucidating particular facilitation tactics. Nearly all classes (20 of 22) met in modern, well-equipped studio classrooms. These had wheeled chairs; movable tables that formed round or hexagonal spaces for 5-7 students; 2-3 walls of whiteboards and/or tabletop whiteboards; and multiple digital presentation screens. In some, labeled whiteboard sections (ABCD) made it easy for instructors to assign groups to specific whiteboard spaces.

One instructor pointed out that fixed digital displays (vs. pull-down screens) took up a lot of room where whiteboards could have been placed. With multiple screens, students had a view from any seat but were not always looking in the same direction; personally I found this disconcerting, but it does "defront" the classroom (Liljedahl, n.d.). Several students (~3-5%) wore headphones throughout class. Headphones may be an accommodation for neurodiverse students who need silence during tests or may serve as an assistive listening device for hearing-impaired students, if the instructor is mic'ed. It was unclear if headphones filled either of these roles in the classes I observed.

Two classes met in a room with smaller tables where pairs or trios could work side by side (PH 220). Some students did use the individual computer screens in this room, but these screens also obscured visibility and restricted interaction across rows of students. When the front screen timed out (fairly quickly), information the teacher had posted disappeared.

I noticed some accessibility issues. When instructors displayed handouts and other visuals not explicitly prepared for the large screen, they did not always enlarge the document or zoom in enough that it could be read from the back of the room. Often both instructors and students wrote too small on the whiteboards, and worn-out whiteboard pens further limited visibility. Finally, it was difficult to hear instructors in some rooms, especially when they faced the board as they wrote. Loud ventilation impeded hearing in Innovation Hall. Instructors should be alerted to these issues so that classes can be fully inclusive. Hearing impairment in particular is an invisible disability that is increasingly common in young people and military veterans.

Based on what I know about class sizes, attendance was pretty good, with rare exceptions. I did not detect a pattern of attendance related to the accountability mechanisms used in class. I did observe patterns of engagement, as discussed below.

3.3. Facilitation Tactics for ACL Work

Here I describe particular small-scale facilitation tactics that I saw instructors use as they interacted with students during ACL work. These specific examples may be useful as models for how instructors can probe and respond to student thinking. Some instructor behaviors supported group work by emphasizing "keep-thinking questions" (Liljedahl, n.d.):

- engaging with students in proactive ways—not just responding to raised hands but initiating interaction with groups through open-ended questions. I heard these examples of conversation starters:
  - How's it going up here in front? [listens] Show me what you've got so far.
  - What are we working on? ...Tell me what you've done so far. [listens] That's a good first step.
  - That looks pretty good. [without saying it is correct] Does it make sense?
  - I see you're looking at the lecture handout. I'm your walking notes, how can I help?
- asking open-ended questions after listening to students' initial response to a conversation starter. I heard these examples of follow-up probes:
  - What about this [pointing], what do you notice? [listens] Those are good observations—all three of those are good observations.
  - Why did you draw it that way? [listens] OK, why can you do that?
  - One small thing is missing, what is it? [waits for response] Bingo! What does that negative sign tell you?
  - That's a good first step. What do you need to do next?
  - Asking students to draw a diagram; add elements such as labels or arrows to their drawing; sketch a graph or number line; make a table; write down a sequence of steps; represent or highlight something with a physical gesture; give a real world example. These all encouraged sense-making and promoted good habits of mind.
- self-monitoring their engagement with each group. This improves equity of access to instructor time and benefits the instructor by providing a broader survey of how students are doing with the material. It is easy for teachers to favor chatty or interactive students, yet even more important to engage with students who are quiet and perhaps hesitant to ask for help. It can take some firmness to move on when a group wants more help. Here are some strategies I heard instructors use to offer formative assessment even as they disengaged from a group:

  - I see a couple of algebra errors, but your setup is OK. I'm not gonna stand here and spend 5 minutes looking for the specifics.
  - That's a neat way of looking at it. Now see if you can make it work.
  - I want to see some pictures when I come back, even if they are bad ones. Gotta commit!

- being patient for group work to play out. It takes time for students to generate and negotiate ideas, even on problems that seem easy to the instructor. This is an equity issue: People think at different paces and it is unfair to value only fast thinkers. It can be helpful to focus class time on a few "essential" problems with "stretch" problems to work on "if you have time." Label the worksheet directly, or post on the board to distinguish these.

- offering encouragement about specific student work or the difficulties of learning. Instructors can be constructive without assessing correctness. I heard these examples:
  - That's a great diagram, it's even color-coded!
  - Yay, this is a great response!
  - This problem may be a little past where you are, but don't let that throw you.
  - You said you didn't want to be here today, but you're here and you're doing it. Props to you, that's the spirit!
  - Seeing something for the first time is OK. Seeing in new ways is how we learn.
  - Intervening when a student asked a question about a topic the class hadn't studied. This avoids overvaluing prior knowledge that wasn't shared by everyone—particularly in CS, where it is a known deterrent to less experienced CS students.

I also saw examples of less supportive facilitation behaviors. These can have negative impacts, if they are repeated or typical.

- responding only to groups who have raised their hands, or who ask questions that interest the instructor
- checking in with students in a perfunctory way, e.g. asking "Are you good?" without waiting long enough for a response or staying to probe their thinking.
- jumping in to correct student work, rather than asking questions to draw out student ideas or help them spot and fix errors. "You need to do this, and fix that."
- criticizing student work overtly: "I have no idea what you are doing." "If you draw it like this, it is a mistake."
- referring to aspects of the work as 'horrible" and "tedious," such as computation or dimensional analysis. This can convey a negative attitude about the topic, or devalue some students' interests or confidence about that aspect of the work.
- interrupting student work time with general comments, without allowing students to reach a stopping place or calling their attention to the board. It was unclear if these remarks were meant for one group or all.

Taking student ideas seriously is at the core of all these facilitation tactics. Both formal and casual instructor behaviors can promote thinking in inclusive and encouraging ways. While ACL approaches inherently hand over some agency to students to manage and monitor their own learning, instructors have options to support student learning and achieve a positive classroom climate without relinquishing their authority.

3.4. Framing and Closing a Recitation

Framing or signposting ACL work is an important function for those who lead recitations or teach blended classes. Signposting helps students see the broader learning goals and how the day's work advances them. It answers key questions about the trajectory of the unit or course: Where have we been? Where are we going? Why are we doing *this* now? Signposting is beneficial at the start, and a recap can be very helpful at the end of class. Mini-lectures can also sometimes be useful midway, e.g. to redirect students or address a widespread difficulty.

In the true recitations that I saw, most instructors—GTAs and faculty alike—did this succinctly; it seemed to be a general norm to speak for 10 minutes or less at the start (often less, sometimes more). While faculty skill and attention to framing varied, GTAs more often visibly struggled to effectively frame the ACL work. They tended to introduce a general technique or concept, or comment on its utility, reflecting their own mathematical expertise and conceptual connections but not necessarily linking to student interests or understandings. It was clear that they were trying to place work in a bigger context, but I often found it difficult to understand their point. With some exceptions (e.g., those with prior high school teaching experience), GTAs will not have deep enough pedagogical content knowledge (PCK)—the specialized knowledge of how students learn particular content—to signpost how ACL work in recitation will link to or build on the ideas taught in lecture or covered in previous weeks. Thus, this is valuable information for lecture instructors to offer GTAs. However, math and CS GTAs reported high variability in how much they interacted with the lead lecturer, and what guidance they received. Building stronger linkages among instructional teams is one way to strengthen instruction generally and help make student experiences less uneven across sections.

The need for PCK not only emerges in signposting, but as instructors work with students and tasks. One example stood out to me, a math worksheet asking students to minimize the cost to make a cylindrical container. This was a difficult problem; students had to realize that minimizing cost meant minimizing the amount of material, in turn minimizing surface area and assuming a fixed thickness. To express the total surface area as the sum of two circles and a rectangle, they had to remember geometric formulae and visualize the shape in 3D to unroll the cylinder wall. These set-up steps obscured the mathematical goal to write the surface area function and set its derivative equal to zero to find the minimum. In trying to help students, one GTA focused on writing the function and taking the derivative, but students did not follow why the function was relevant, because they did not understand why cost related to surface area, and the GTA did not spot that this was the real issue.

In the long run, we want students to be able to unpack such details and draw on their other knowledge to set up a problem, but this was a hard place for everyone to start. This more advanced problem could be labeled for faster-moving groups to tackle as a challenge. This problem illustrates that PCK cannot be assumed, and that the choice of tasks matters.

In general, I did not analyze the worksheet tasks; I focused my observations on their deployment in use, rather than the task itself. Some tasks were clearly more routine—take this derivative, integrate that function—and others required more thought or analysis. I was however surprised at the extent of variation in the problems from section to section, even for recitations that were part of the same course and occurring in the same week of the term.

3.5. Grouping Strategies

I observed students working in groups from 2 to 8 students, based on the physical arrangement of the space, the task, and the instructions they were given. I saw a variety of strategies used to form groups, and likely missed others because these choices were already established

- informal one-time grouping, e.g. "work with a neighbor"
- informal habitual grouping, where students self-select into a group by choosing a table
- permanent or semi-permanent groups. For example, one instructor had formed groups based on students' majors.

I did not see examples of deliberate group remixing or of one-time grouping, such as using playing cards to assign groups randomly. I comment on student self-selection by gender presentation in Section 3.8. Liljedahl (n.d.) cites the literature in describing that visibly random groups strongly enhance students' mindset that they were not only going to think, but to contribute. Random groupings also break down social barriers, reduce stress, and increase enthusiasm for mathematics.

3.6. Whiteboards or Paper

Wall-mounted and/or mobile whiteboards were available in all rooms. When rooms were more crowded, some students would choose a mobile whiteboard, but most seemed to choose wall-mounted whiteboards when they could. Instructors said that having students work on whiteboards had multiple advantages. Helping 3-4 students together is more efficient than interacting with one student at a time, and instructors could easily survey the board work and target their interventions accordingly. They valued that group work on whiteboards pushed students to collaborate with each other and check their work with others. Turning in a photo of their joint work provided some group accountability to do and understand the work, and was an easy way to take attendance, as students' names were written on the board with their solutions.

Students told me that whiteboard work enabled all to see the work and discuss it. They liked using different colors to add their contributions, and they could talk about it, erase or refine it easily, and walk around and look at other groups' work. Group work helped them make friends and get to know people: "Even if you don't like math, it's fun." And it was good to stand up:

"Your brain works better when you move around." In general, I saw that whiteboards offer notable advantages over working on paper, corroborating Peter Liljedahl's findings (2020; n.d.; see also Boryga, 2023) about the "massive" merits of vertical non-permanent surfaces for group work. In particular, Liljedahl finds that non-permanence promotes risk-taking, and a vertical surface prevents students from disengaging, as students attested. He also notes that students think more when standing in loose formation around the teacher.

3.7. Accountability Mechanisms

Grouping and whiteboards were also related to how individual students were held accountable for class work. STEM faculty are often concerned about plagiarism and accountability for the recitation work. I was told more than once that Mason students needed to have work graded to stay on top of the coursework. I saw several accountability mechanisms in use:

- turn in all worksheet problems on paper (or via upload) immediately at the end of class
- turn in a subset of the worksheet problems later (e.g. by the next day)
- turn in problems later that are similar to, but not the same as the worksheet problems
- work individual warmup problems that are based on a pre-class assignment
- work individual problems at the board (e.g., after some group work)
- photograph and turn in the group solution on the whiteboard, usually soon after class.

In general, I observed a direct and negative relationship between the degree and liveliness of meaningful conversation among students during class, and the requirements to turn in work. A requirement to turn in a substantial body of individual work after class enhanced time pressure to complete the work and thus inhibited student conversation the most. Requiring turned-in work later (e.g., next day) fostered more conversation, likely because it reduced pressure on students to finish it in class. Turning in a subset of the work, or turning it in as a group, with everyone's name, were the practices associated with the highest level of on-task conversation.

Working on paper made it easier to turn in individual work but tended to inhibit conversation, while work on the whiteboards fostered conversation rather naturally. I saw only one notable exception to this pattern, described in Vignette 5 below.

There were also variations in whether the work was graded for effort or for a numerical score, and how it counted in the overall assessment and grading scheme. As far as I could tell, students always had other homework beyond the group tasks, as well as quizzes and tests—that is, each class had multiple ways to hold individuals accountable for course content. Collecting some work does encourage attendance and participation; however it need not be scored, or can be marked on a simple scale, such as 0-1-2 points for no/partial/full credit for the attempt. Concerns about plagiarism largely melt away if ACL work is not done for a high-stakes grade. Talbert (2021, 2023) offers some useful summaries of research about student motivation and its relationship to grading and feedback. Talbert's work with David Clark (Clark & Talbert, 2023; n.d.) may be very helpful as a resource for instructor conversations about these issues. Liljedahl (n.d.) also emphasizes the benefits of rebranding homework as "check-your-understanding" questions to emphasize purpose and thought over scoring, and of evaluating what we value.

3.8. Classroom Support from Peer Instructors (LAs and UTAs)

Eighteen of 22 sessions had LAs or UTAs who assisted instructors or GTAs; in one class a GTA assisted the instructor. A few were missing or arrived late. When I spoke with peer instructors, they described their work as helping people and answering questions, and encouraging people to work together in class or outside class. LAs and UTAs reported they held office hours but only the GTAs did grading, consistent with best practices for LA roles. A majority were proactive in how they circulated and interacted with groups, and some were quite skillful in interacting with students. Others (most often those who were new to this work) were more reactive, waiting in a corner and responding when a hand was raised. Lead instructors' expectations and instructions—not just modeling—had a clear impact on whether and how peer instructors interacted.

Overall I saw the LAs to be significant positive contributors to the quality of instruction and student support in these gateway recitation sections. The strongest LAs were as effective as GTAs, or more so, and their presence made it possible for one GTA to work with a section of up to 45 students. Serving as an LA or UTA is a also good campus job for students; it benefits their own learning and may contribute to persisting in the major and/or a future career in teaching.

3.9. Last but not Least, the Students

I very much enjoyed the GMU students. They were diverse, engaged, and down to earth. They were curious about my presence, keen to share their experiences with me when asked, and articulate about what was effective or not about this type of class work. They seemed very happy to be in class, and to be working with each other in face-to-face settings once again.

I was struck by some persistent gendered patterns in these courses. Students who presented as women were a minority in almost all the sections I observed, sometimes just 15-20% of those in attendance. Quantitative data from 2021 indicated that women were more numerous in the "stretch" calculus and algebra-based physics courses, a pattern that also held in my estimates.

In most sections, it was common for women students to gather in small groups. While in nearly every class a few women joined a group whose other members presented as men, most classes had 1-2 groups of all women, and/or a group that was majority women, such as 4 women with 1 or 2 men. Woman-centered groups thus accounted for the majority of women present. It seemed that women wearing hijab seemed to work together as well as with non-hijab-wearing women. Given the persistent issues for women in quantitative STEM fields, it would be very interesting to carry out a qualitative study of the experiences of women (and nonbinary) students in these gateway courses, with keen attention to intersections of gender and race/ethnicity. It would also be informative to track persistence patterns by gender.

4. Features of Effective Small Group Learning

Here I offer analysis of the quality of student engagement and its relationship to instructor practices. To analyze the quality of the ACL work I observed, I turned to an oldie but goodie from the literature. Karl Smith (1996) articulated five essential elements of well-structured cooperative learning groups, based on research by the Johnson brothers and others. Here I quote selectively from Smith's descriptions of the five elements (1996, pp. 4-5); the full paper elaborates on how instructors can design and facilitate group work to align with these elements.

E1. **Positive Interdependence:** The heart of cooperative learning is positive interdependence. Students must believe that they are linked with others in a way that one cannot succeed unless the other members of the group succeed (and vice versa). Students are working together to get the job done. In other words, students must perceive that they "sink or swim together."

E2. **Face-to-Face Promotive Interaction:** Once [an instructor] establishes positive interdependence, he or she must ensure that students interact to help each other accomplish the task and promote each other's success. Students are expected to explain orally to each other how to solve problems, discuss with each other the nature of the concepts and strategies being learned, teach their knowledge to classmates, explain to each other the connections between present and past learning, and help, encourage, and support each other's efforts to learn. Silent students are uninvolved students who are not contributing to the learning of others or themselves.

E3. **Individual Accountability/Personal Responsibility:** The purpose of cooperative learning groups is to make each member a stronger individual in his or her own right. Students learn together so that they can subsequently perform better as individuals. To ensure that each member is strengthened, students are held individually accountable to do their share of the work. The performance of each individual student is assessed and the results given back to the individual and perhaps to the group. The group needs to knows who needs more assistance in completing the assignment, and group members need to know they cannot "hitch-hike" on the work of others.

E4. **Teamwork Skills:** Contributing to the success of a cooperative effort requires teamwork skills. Students must have and use the needed leadership, decision-making, trust-building, communication, and conflict-management skills. These skills have to be taught just as purposefully and precisely as academic skills. Many students have never worked cooperatively in learning situations and, therefore, lack the needed teamwork skills for doing so effectively.

E5. **Group Processing:** Professors need to ensure that members of each cooperative learning group discuss how well they are achieving their goals and maintaining effective working relationships. Groups need to describe what member actions are helpful and unhelpful and make decisions about what to continue or change. Such processing enables learning groups to focus on group maintenance, facilitates the learning of collaborative skills, ensures that members receive feedback on their participation, and reminds students to practice collaborative skills consistently.

Below I use these elements as criteria to analyze the group work I observed. Elements E1-E3 are most pertinent for organizing everyday group work, and I saw numerous ways these elements manifested. Element E4 is important in setting and reinforcing good norms of behavior for group work and E5 is crucial when students work in stable groups yet also useful in strengthening behaviors among informal groups. I could infer when strong and positive norms had been established in a class, but saw few examples of E4-E5 in action (see Smith for more). In the vignettes, I also comment on instructor pedagogical content knowledge (PCK) and facilitation tactics that emphasize conceptual learning and equity among diverse learners.

4.1. Models of Effective Group Work

Vignettes 1 and 2 illustrate two different paths for generating positive interdependence through both promotive interactions and individual accountability. Vignette 1 comes from a GTA and represents skillful use of a basic structure for ACL work in a true recitation. This model of instruction offers a good model for any recitation instructor, given appropriate support.

Vignette 1	*Analysis 1*
After taking some time to frame the day's work and review one example, the GTA posts a set of problems on screen, assigns each table group a section of whiteboard, and hands out pens. "Turn in 2 of 5 problems in Gradescope" by tomorrow, he says. The students hop up to the board. Lively discussion ensues and continues for several minutes before most groups start writing on the whiteboards.	Working at the whiteboard, with a group, was clearly a class norm; working alone or staying at the table was not an option (E1). Working at the whiteboard emphasizes promotive interactions (E2), while writing up two problems to turn in gives individuals accountability (E3) for the group work yet still offers students agency and allows for groups to work at varying paces.

Vignette 2 comes from an experienced instructor teaching a blended class, and shows some others uses of whiteboards and collaborative work. Because some of these tactics depend on strong PCK, this vignette offers one model of practice for more experienced instructors leading blended classes, or those incorporating ACL into a lecture section.

Vignette 2	*Analysis 2*
Before class, students watched a video about solving a particular type of problem. To start class, the instructor posts three warmup problems and asks all students to go to the	The warmup problems set a norm of active engagement in class (E1). They alert students that they should arrive ready to work, having completed the pre-class assignment (E3). Pre-

board and work them. "Now find a solution that looks different from yours. Did everybody find one?" After brief conversation during which students compare notes and make adjustments, the instructor chooses one solution written in "beautiful orange." She comments on that work, asking the seated student specific questions about how she thought about the problem. She also asks all students questions about the pre-class video. Then she has them close their eyes and raise a hand if they watched the video before class. About 80% do raise their hand. For the next 15 minutes, she leads interactive example-solving, interspersed with student individual work on similar examples that she writes on the board. Students work on paper or use mobile whiteboards; they readily break into conversation as they do their own work. The instructor and two LAs circulate to observe student work. Noticing a common difficulty, she runs another eyes-closed, hand-raised poll: Who took trig online during the pandemic? Who never took it at all?	class work moves lower-level tasks to outside class and enables class time to focus on what is difficult or benefits from repetition and help. Checking work against others' work promotes interaction and communicates positive interdependence (E1, E2), and offers everyone a quick chance to fix easy errors. Selecting student work to discuss—taking it seriously—reinforces positive interdependence (E1). Being called on can feel higher in stakes for the student but, when done regularly in a supportive way, it is a way to hold individuals accountable without collecting or grading anything (E3). Likewise, informal formative assessment about the pre-class work reinforce norms of preparation and accountability for all (E3), and prompts some metacognition about one's own role in shared success (E5). The interactive lecture is interspersed with short spurts of student work. By circulating to look at student work, the instructor gains a sense of what to discuss or highlight next. Informal polls further help her understand student preparation, or lack of it, and provide moments of metacognition for the students themselves. Her polling method encourages candor by protecting their privacy.

4.2. Responding to Student Thinking in Real Time

Like Vignette 2, Vignette 3 highlights different features of excellent ACL work. It includes some simple ways instructors can signal the value of recitation work and connect it to the lecture content. To do this, they need to be in good communication with the lecture section instructor, whether they are a faculty peer or a GTA. This vignette also illustrates two uses of mini-lecture. This example, from an experienced instructor who had strong pedagogical content knowledge, is not necessarily something a novice can pull off.

Vignette 3	*Analysis 3*
The instructor opens class by posting new problems and asking students if some of the problems they had worked here [in recitation] helped on their recent exam. Yes, they agree, some were just like what we did. "How about our hard work on the right-hand rule?" Yes, that too, they agree. He gives a mini-lecture linking the new problems to the prior topic, using a computer tool to demonstrate a type of analysis common to both. "Pair up with someone to work. I will circulate and occasionally make comments, and we'll go back and forth like that," he says. Students work in pairs or alone for 5 minutes, at their tables (this room has no whiteboards). It's very quiet. The instructor moves around the room, observing and listening, then stops the class: "Give me a show of fingers: Where are you at with knowing where to start on #15?" Students point up, down or sideways to indicate high, medium or low confidence. Based on this feedback, the instructor takes about 6 minutes to work a new example, using drawings on the board, the simulation tool, and hand gestures. As students return to the worksheet, the noise level rises.	Connecting recitation work to exam success helps to show students that the group work is valuable, a mild type of group processing (E5) and demonstrating positive interdependence (E1). This also ties individual accountability on exams (E3) to taking an active part in recitation. Strong signposting further demonstrates how learning is cumulative and recitations support learning (E1), and helps students to make conceptual connections. The instructor encourages interaction (E2) but can tell by watching and listening that students were not working fruitfully yet. Using a quick self-assessment strategy that is already familiar to students, he fosters reflection and gathers information without making students be too public about what they don't know (mild E5) and reinforces the notion that we're all in this together (E1). By circulating and listening, the instructor detects a common issue impeding students. Strong PCK allows him to inject a mini-lecture to clarify the ideas of the problem, without using any exact example that students are to solve. This succeeds in re-launching students on the problem.

4.3. Finding the Right Balance

Vignette 4 illustrates some challenges that arise when the balance of positive interdependence is skewed toward individual accountability. In this case and others I saw, a recitation functions as supervised homework time, where students can work together or separately, as they prefer, with help available from instructors. This is not wasted time—students do value this work, as 2021 survey responses attest—but it is more like what goes on in a tutoring center. Use of this approach in recitation seems to emerge from the assumption that *hands-on* work—repetition or individual practice—is the main benefit of active classwork, rather than the *minds-on* work of collaborative sensemaking. Hands-on models were observed in both true recitations and blended classes. Instructor shyness, or discomfort in approaching students in ways that would have felt intrusive to them as students, may also play a role in shaping instructor choices.

Vignette 4	*Analysis 4*
There is no framing to start class, and students begin right away to work on a task. They are working quietly and must turn in their own solution at the end of class. About 15 minutes in, a student raises his hand. The GTA responds but doesn't answer his question as directly as the student wishes. The student repeats his question loudly and sounds frustrated; they continue discussing for several minutes. Other students at this table studiously ignore the interaction; one asks the GTA a question very politely, as if to distance himself from his classmate. Twice more during class, the same student raises his hand and one or two instructors respond. Each time he sounds rude and even belligerent, such that I consider what to do if he becomes more disruptive or even violent: Could I use my schoolmarm voice to calm things down, or where would I find more adult help if it is needed? Of the other student questions that I can hear, a good portion seem to be procedural, e.g. to clarify task instructions or locate a resource. Nonetheless, slowly over time, the level of conversation among students rises, and 35 minutes in, it has reached a pleasant coffee-shop level of buzz. I spoke only with a few students who left class early. Most were still on task trying to complete the work before the deadline.	This was a large section, yet one of the quieter class meetings I attended. Overall, positive interdependence was low (E1) and individual accountability was high (E3). Working together was an option, yet students did not seem to initiate conversation with each other until they were stuck (E2). In an earlier section of this course, a GTA told me that students "sometimes ask questions in ways that aren't very polite," and this incident seemed to exemplify that very issue. These GTAs did not feel they had standing to teach students how to work effectively in groups (E4), and other students seemed uncomfortable with the classmate's behavior (E5). At minimum, GTAs need to know what they can do to manage a rude student and have some tools for this. Moreover, setting more explicit norms and taking some time to develop teamwork skills may be time well invested, as students are coming off a long period of online remote learning and reduced socialization. Some skills that earlier cohorts may have built in school, on clubs or teams, may be frayed or under-developed compared to the past. These skills are likely to be useful in more advanced courses and future workplaces. This is an equity issue too: here one student consumed a sizable fraction of the instructor time available to all ~50 students, and distracted several classmates.

While Vignettes 1-2 emphasize the use of whiteboards, Vignette 5 illustrates that table-based group work on individual papers can work to foster positive interdependence. This approach has a higher degree of difficulty; students standing at whiteboards really does have some built-in advantages. This example comes from a GTA. It expands the range of possible approaches but should not be the standard model of instruction for GTAs.

Vignette 5	*Analysis 5*
The GTA reviews quiz results, frames the day's work and offers some related definitions. He emphasizes the big idea of the worksheet, and why we care—both the 'what' and the 'so what.' He makes some small errors but catches them. "Now break into your groups, start working on the packet, and we'll be around to help." Students work on their own paper at their tables yet still engage in lively conversations as a full table. One student jumps up to help his table mates, pointing at their papers and gesticulating. The GTA circulates, opening with "What are we working on?" as he sits down with each group and asks one person to explain the group's work. He listens as the student responds, follows up to clarify his understanding, then opens additional questions to the whole group. He offers encouraging comments—"You found the critical points, good start"—as well as tactical advice: "Ah, your trouble is you haven't made a number line. ...Start with a graph, don't try to do it in your head. OK, now from your graph, what number in this interval would you test?" He spends about 4-5 minutes before moving on to speak with the next group.	This was the only class where I saw very good student engagement when students were working on individual papers at their seats (E2). In general, engagement was stronger and conversation more lively when students were at the whiteboards. Here students were to turn in some or all of these problems by the next day (E3); they did not seem anxious to finish them. After class, the GTA explained his explicit practice to drop in at each table and call on a random student to share the group's work. This generates strong interdependence (E1) but does add some risk for shy or less prepared students to be publicly embarrassed. Nonetheless, I could not hear that students hesitated to speak, and I noted how quickly and keenly students engaged at all the tables. For this to work, positive group norms must be set early, reinforced often, and students supported to learn the needed skills (E4). The GTA's comments suggest that he had worked hard to establish these norms (E4). Strong PCK likely helped this GTA pull off this approach.

5. Overall Conclusions and Recommendations

Here I use these observation data to offer an assessment of the project's progress on its goals to implement ACL as a regular feature of recitations and blended gateway STEM courses.

I was impressed with the extent of ACL work that I saw. In both true recitations and blended classes, I saw high proportions of time used for active and often collaborative problem-solving during class time, and it seemed clear that this type of work was a general norm for teachers and students. I also saw many several high-quality examples of ACL work led by both GTAs and faculty. The active classrooms at GMU are impressive, and many instructors used these facilities to good effect. LAs and UTAs contributed in appropriate and essential ways. The commitment to ACL in these gateway courses has already moved the culture away from *hands-off* instruction, where the instructor does the thinking and students watch.

The foremost challenge for the project now is to reduce the variability in quality from section to section—a matter of both effectiveness and equity. I saw varying degrees of emphasis on the use of recitation time for *hands-on work* that supports student practice with course material in a setting where they can converse, compare answers and ask questions—and for *minds-on work* that aims to build conceptual understanding, requires students to explain their ideas aloud, and reinforces learning with metacognition. The research is pretty clear that hands-on work is good and minds-on work is better (e.g., Chi & Boucher, 2023). So course or department teams will need to decide what are their objectives for these courses, based on both what is desirable and what is feasible. The objectives may differ for blended classes and recitations, and strong support for instructors is important. For all three departments, I suggest attention to two aspects of instruction. I will discuss structural supports and department-specific concerns separately.

5.1. Recommendation: Providing Group-worthy Tasks for Classroom Work

One key support for minds-on work in gateway recitations is the curriculum, the quality of problems or student tasks themselves. All three departments have done substantial work to gather or develop such "thinking tasks" (Liljedahl, n.d.). Curriculum interests instructors and is thus a way to foster continued course development and instructor learning. It is not enough just to "build it and they will come;" worksheets on a website are not useful if people do not encourage, model or discuss their use. To develop instructors' understanding and practices, faculty may benefit from taking part in work to develop, review or analyze conceptually rich, group-worthy tasks, and GTAs from analyzing examples of student work. All will be more effective if they understand the goal of a task and some ways to guide it, develop insight about student difficulties, and build capacity to figure that out themselves.

5.2. Recommendation: Organizing and Facilitating Classroom Work

A second key support for minds-on work is how ACL activities are facilitated. Increasingly, the literature points to classroom structures for interaction as important to strengthen learning and increase equity (e.g., Hogan & Sathy, 2022; Sathy & Hogan, 2019/2023). Structures are repeated and regular practices that communicate to students the class norms of behavior and engagement, encourage their active participation, and make class time more impactful for learning. Thoughtful and explicit structures also help to reduce inequities that arise from differences in student circumstances and instructors' implicit biases. I identify three central structures for ACL work where stronger and more consistent choices could smooth variability in instruction:

- Whether students write on whiteboards vs. on paper at their table
- Whether and when students must hand in their work, alone or as a group
- Whether instructors circulate assertively and how they interact with students.

These structures can be understood and enacted by all instructors (GTAs, faculty, and LAs/UTAs). They will need a basic ACL teaching toolkit that includes explicit, research-based principles and tactics on structures for organizing group work, and a set of conversation starters and follow-up probes that they can use to interact with students, even if they aren't quite sure where the conversation will head next. Ideally, course teams will move toward course-wide agreements and norms about these practices, or at least minimum standards that all agree to implement. Next steps in this work may include

- equipping all instructors with good mental models of ACL teaching and a supportive facilitation toolkit that can guide their in-the-moment teaching decisions
- communicating expectations for instructors to use these models and tools, and provide feedback to help them develop as teachers
- educating instructors about what research says about student motivation, learning, and accountability, and identifying how classroom practices fruitfully balance these
- continuing to develop the teaching knowledge, skills and commitments of faculty who teach gateway courses, through ongoing engagement with each other, with ACL literature or speakers, and with other teaching-centered professional communities.

5.3. So What?

The classroom observations reported here provide evidence of good progress on the project's goals. They highlight practices that can further strengthen classroom ACL work and point to next steps for spreading and sustaining them. By working to shift and stabilize ACL practice further along the continuum from *hands-off* to *hands-on* to *minds-on* work, recitations and blended classes in GMU's gateway courses can provide stronger learning experiences and more positive learning environments. Prior research has shown that, in general, such features support student success and persistence (e.g., Freeman et al., 2014) and to reduce inequities among student experience and outcomes, acting as a rising tide that lifts all boats (e.g. Theobald et al., 2020). Moreover, data from this project already link student reports about the ACL practices used in their gateway STEM classes to their classroom experience and self-reported learning (Laursen & Archie, 2022; Archie & Laursen, forthcoming).

The project's strategic choice to implement ACL in recitations is a practical place to start but does have limits. We do not know whether and how desired outcomes, such as course success and persistence, respond to implementation of ACL in recitations alone, if other aspects of instruction and assessment are unaltered. Work from this project indicates that a coalition of empowered term faculty can make great strides in gateway courses when they are handed the reins (e.g., Bulancea et al., 2021), yet also reveals constraints that arise from larger institutional systems, such as power dynamics, institutional memory, lack of resources, and heavy teaching workloads. In my view, sustaining this work will require stable and strong course coordination and mechanisms for ongoing professional development of GTAs and new term faculty.

6. Acknowledgments

I thank the instructors and students who welcomed me to their class and spoke with me during my visit. Mark Snyder provided data about class schedules, instructors and enrollments, and Bob Sachs provided rides. Mark Snyder, Jess Rosenberg, and Bob Sachs answered questions and contacted instructors on my behalf. Melinda Ryan provided helpful coordination and logistical support, and several team members met with me for meals and good conversation. I remain fascinated with the army of autonomous robots that deliver food at GMU, and I wish I'd tried this while visiting campus, but doubt they would have been good conversationalists.

This work was supported by the National Science Foundation under award DUE-1821440. Design and implementation work at GMU was supported by the NSF under award DUE-1821589. Both awards are entitled *Collaborative Research: Building a Culture of Active Learning through Course-Based Communities of Transformation.* All findings and opinions are those of the author.

7. References Cited

Boryga, A. (2023, March 24). How to turn your math classroom into a 'thinking classroom.' *Edutopia*. https://www.edutopia.org/article/thinking-classroom-peter-liljedahl-math

Bulancea, G., Granfield, P., Jauchen, J., Love, J., Nelson, M., Sachs, R., & Sausville, C. (2021). A community of grassroots leaders: Leveraging faculty networks to create change. *PRIMUS*, *31*(3-5), 627-642.

Chi, M. T., & Boucher, N. S. (2023). Applying the ICAP framework to improve classroom learning. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply to science of learning in your academic setting* (pp. 94-110). Society for the Teaching of Psychology. https://teachpsych.org/ebooks/itow

Clark, D., & Talbert, R. (2023). *Grading for growth: A guide to alternative grading practices that promote authentic learning and student engagement in higher education*. Routledge.

Clark, D., & Talbert, R. (n.d.) *Grading for Growth*. https://gradingforgrowth.com/

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Hogan, K. A., & Sathy, V. (2022). *Inclusive teaching: Strategies for promoting equity in the college classroom*. West Virginia University Press.

Laursen, S., & Archie, T. (2022). Evaluation report: Findings from student surveys of GMU Gateway2STEM courses, Fall 2021. Boulder, CO: Ethnography & Evaluation Research. https://docs.google.com/document/d/1FFUMArVSenCoxMwKpXNId2sSru_Ajs76sMKpzdv0SYQ/edit?usp=sharing

Liljedahl, P. (n.d.) 14 practices. *Building Thinking Classrooms*, website to support his 2020 book. https://buildingthinkingclassrooms.com/14-practices/

Liljedahl, P. (2020). *Building thinking classrooms in mathematics, grades K-12: 14 teaching practices for enhancing learning*. Corwin Press.

Sathy, V., & Hogan, K. A. (2019, 2023). How to make your teaching more inclusive: Advice guide. *The Chronicle of Higher Education*. https://www.chronicle.com/article/how-to-make-your-teaching-more-inclusive/

Smith, K. A. (1996). Cooperative learning: Making “groupwork” work. In C. Bonwell & T. Sutherlund, Eds., *Active learning: Lessons from practice and emerging issues. New Directions for Teaching and Learning 67*, 71-82. San Francisco: Jossey-Bass. Available online at https://karlsmithmn.org/wp-content/uploads/2017/08/NDTLCL5.doc

Talbert, R. (2021, August 2). Words, numbers, or nothing? A classic study finds some eye-opening effects of different feedback on learning. *Grading for Growth*, https://gradingforgrowth.com/p/words-numbers-or-nothing

Talbert, R. (2023, March 13). Does alternative grading make cheating more likely? A helpful meta-analysis of research on academic motivation gives clarity. *Grading for Growth*, https://gradingforgrowth.com/p/does-alternative-grading-make-cheating

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., et al. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476-6483.