Embry-Riddle Aeronautical University

From the SelectedWorks of Darryl Chamberlain

March, 2023

How one instructor can teach a large-scale, mastery-based College Algebra course online

Darryl Chamberlain, *Embry-Riddle Aeronautical University*

Available at: https://works.bepress.com/darryl-chamberlain/15/

How one instructor can teach a large-scale, mastery-based College Algebra course online

Abstract: Mastery-based testing is an assessment scheme that encourages students to learn from their mistakes and develop an understanding of material before moving on. For an entry-level course with large enrollment like College Algebra, this could allow some students to move at an appropriate pace for themselves. This paper outlines the changes made to make College Algebra at the University of Florida an asynchronous mastery-based course for the 500+ students who take the course per year and concludes with a series of lessons learned from the endeavor.

Keywords: Mastery-Based Testing, Asynchronous, Automated Feedback

1 INTRODUCTION

Designing and implementing online courses has never been more important. While COVID-19 forced the immediate need for accessible courses, online courses have gained popularity in recent years due to their ability to increase access to quality learning (Trenholm et al., 2015). Mathematics courses are noted to be especially challenging to teach online due to issues such as quickly communicating mathematical notation and diagrams (Smith et al., 2003; Trenholm et al., 2019). These types of difficulties lead to online mathematics courses requiring more of an instructor's time and energy - difficulties that are compounded when the size of the course increases. Even still, the context of a mathematics course makes it difficult for instructors to easily take and implement what is being done at another university.

College Algebra is a prime example of a mathematics course with broad appeal and enrollment across higher education. For some majors (e.g., Nursing and Psychology) it is one of the highest-level math courses needed. For others (e.g., Accounting and Business) it is a stepping stone to Survey of Calculus - a flavor of applied Calculus. Still others (e.g., Biology and Chemisty) use College Algebra as the start of their journey to Calculus and Analytical Geometry. Even then some students who do not need it for their major take the course to satisfy their general mathematics requirement. Colleges and universities commonly adjust their College Algebra course to fit the particular demographics of their school, resulting in similar content with different foci (Herriott and Dunbar, 2009).

Given the general difficulties of large, online mathematics courses as well as the broad needs for the varying populations of students, the author was hired at a large R1 university in the southeastern US to modernize their College Algebra course. This paper presents the fruits of this labor – a large-scale, mastery-based online College Algebra course – in the hopes that other colleges and universities may use the course outright or similarly consider how to modernize their own College Algebra course. To this end it will outline the various challenges and solutions the author encountered as well as summarize the overall lessons learned in the process.

2 DESIGN OF THE COURSE AT UNIVERSITY OF FLORIDA

Designing and implementing College Algebra at the University of Florida took several semesters and went through two major modifications. The first began in Spring 2018 and the second in Spring 2020. This section presents an overview of how the various components of the course changed through the two major modifications.

2.1 Course type

The course was hybrid (meet once a week in person, otherwise online) before modification and during the first modification period. Between 500-750 students take the course annually and include:

- On-campus traditional students: *students who went to college directly after high school and are taking some courses in-person*;
- Distance-learning traditional students: *students who went to college directly after high school but are taking all of their courses online*;
- Nontraditional students: *students who are returning to college after some break while simultaneously working full-time*; and
- Dual enrollment high school students: *especially those without a nearby community college to earn college credit.*

In addition to background, a variety of majors and years are represented within the course. The unifying theme for students was a need to take the course to continue toward Calculus[1]. Thus the course was designed to be both a gentle introduction and a quick review to functions.

In-person meetings consisted of a graduate teaching assistant (GTA) meeting with 20-30 students once a week for 50 minutes. Before modification, these meetings were question-and-answer sessions where the GTA would either solicit and answer questions from the class or work out chosen problems in class. From Spring 2018 through Fall 2019, these meetings shifted to group-based activities where students engaged with material beyond what they would see in the homework. From Spring 2020 on, these activities were moved online to Canvas and completed asynchronously.

[1] Alternative math courses were available for students who did not need to take Calculus.

2.2 Curriculum

Curriculum change occured in Spring 2018 and remained static through the second major modification period. Curriculum was designed for students who planned to take Calculus. Literature suggests that potential Calculus students should be prepared for: function notation, exponential and logarithmic growth, rates of change, Riemann sums, function as a process, covariation, and limits and approximations (Boersma et al., 2017; Carlson et al., 2010). Given that people consolidate knowledge around concepts (Verschaffel et al., 2017) the curriculum was split to first focus on classes of functions individually that are later compared and acted upon. With a robust understanding of functions, students can better understand rates of change, covariation, and limits. Some textbook authors have moved in this direction, such as the Active Prelude to Calculus by Boelkins and Keller (`https://activecalculus.org/`).

The first set of content is referred to as the Core Modules and (mostly) represents the major classes of functions a student would encounter in a traditional College Algebra course. These are:

1. Real and Complex Numbers
2. Linear Functions
3. Linear Inequalities
4. Quadratic Functions
5. Radical Functions
6. Polynomial Functions
7. Rational Functions
8. Logarithmic and Exponential Functions

Within each Module, students focus on (i) the salient characteristics of the function, (ii) different representations of the function, and (iii) how to solve equalities including the function. For example, the Radical Functions module has three learning objectives:

- Identify the domain of a radical function.
- Translate between representations *(equation, graph, description, etc)* of a radical function.
- Solve radical equations.

A point of emphasis is that the major change in this College Algebra curriculum is the *organization*, not the content. Many of the traditional learning objectives can be found in the course organized around either an individual class of functions or in operating on one or more functions.

The second set of content is referred to as Advanced Modules and represents ways in which functions can be compared or acted upon. This content was split into two paths to emulate the two routes a student may be taking the course for: to complete Survey of Calculus or Calculus I.

Preparation for Survey of Calculus

9. Modeling with Linear Functions
10. Modeling with Power Functions
11. Modeling with Log or Exp Functions
12. Solving Real-world Modeling Word Problems

Preparation for Calculus I

9. Operations on Functions
10. Synthetic Division
11. Introduction to Limits
12. Graphing Rational Functions

In the preparation for Survey of Calculus, students begin to compare and categorize the functions they learned in the Core Modules by considering the rate of growth the model requires. These modules are meant to provide an introduction to rates of change, function as a process, and covariation. In the preparation for Calculus I, students begin to act on functions through two main operators: division and limits. These modules are meant to provide an introduction to rates of change, function as a process, and limits and approximations. As before, many of the learning objectives associated to these modules can be found in a traditional College Algebra textbook. The learning objectives associated with each module are provided in the Appendix.

2.3 Assessment Method

As with the curriculum shift, the assessment modifications began in Spring 2018. Throughout the course a student was expected to work through 12 modules of content: 8 core modules and 4 advanced modules from one of the two tracks. Moreover, students should develop a strong understanding of earlier functions (e.g., linear and quadratic) that can be used to leverage understanding of later functions[2], which would then be acted upon in some way in the advanced modules. Thus the course utilized a mastery-based testing scheme that allowed students to progress at their own pace and ensured a foundation of understanding was developed before progressing.

Mastery-based testing is an alternative assessment scheme that focuses on proficiency-based rather than percentage-based grading (Harsy, 2020). Mastery-based testing is similar to *standards-based grading*, which focuses on students showing proficiency in a subset of specific learning objectives and these proficiencies are the only factors on final grade (Knight and Cooper, 2019). According to Elsinger and Lewis (2019), both grading schemes share four main features:

1. Students are provided with a clear list of course learning objectives.
2. Final course grades are based primarily on how many standards/concepts the student masters.
3. Students are provided ample opportunities to reassess mastery of any standard/concept.
4. Attempts are graded using a form of pass/fail grading, and only the best mark is used toward the course final grade (p. 2).

For this course, learning objectives are explicitly provided and organized around modules. Students are graded pass/fail on individual modules during each *Progress Quiz*. From Spring 2018 through Fall 2019, Progress Quizzes covered 4 modules and each module contained 5 questions linked to learning objectives. From Spring 2020 on, Progress Quizzes covered 2 modules and each module contained 10 questions. An individual module was graded as *passed* if the student answered 80%+ of the question from that module.

[2]For example, the vertex form of quadratics may be abstracted into a general form for translations of rational, radical, logarithmic, and exponential functions

A passing score means the student continues to the next module and does not retest on the passed module. Progress Quizzes are thus different for each student based on the student's progress (hence the name) and students choose which modules they would test on each time. From Spring 2018 through Fall 2019, students took four Progress Quizzes (approximately one every two weeks) and thus had a total of 16 attempts at passing 12 modules. From Spring 2020 on, students take ten Progress Quizzes (approximately one a week) and thus have a total of 20 attempts at passing 12 modules, giving them even more opportunities to reassess the concepts. Progress Quizzes were administered online through Canvas, had a 2-hour time limit, were open for a 48-hour time window, and were proctored by a third party (either ProctorU or Honorlock).

Grading pass/fail by module kept the focus on learning concepts and not amassing some specified number of potentially unrelated standards. Their final grade in the course is primarily[3] based on the number of modules they have passed and is summarized below.

- **A**: 8 core modules and 4 advanced modules;
- **B**: 8 core modules and 2 advanced modules;
- **C**: 8 core modules;
- **D**: 6-7 core modules;
- **F**: 0-5 core modules.

The grading scheme again emphasizes that students who pass the course with a C have a strong foundation in functions while students who earn an A or B have made some progress toward comparing and acting on these functions. Moreover, the Progress Quizzes act as both formative (informing the student what they know so that they can develop) and summative (summary of their current knowledge) assessments. Note that students could not see their grade until the end of the semester based on this grading scheme. However, starting after Progress Quiz 5, students were provided weekly projections of their final grade in the course based on how many modules they had passed up to that point.

Homework is graded upon completion and due at the end of the term. Students have unlimited opportunities to answer each question. The goal of the homework is to provide students ample opportunity to engage with the content in each Module. Homework was written in LaTeX and SageMath to create dynamic questions that students could repeat for practice. The homework system was hosted online through Ximera https://ximera.osu.edu/ and can be found at `https://xronos.clas.ufl.edu/ufmac1105`.

2.4 Overview of Course Design

An overview of the course design, including when changes were made to the course, is provided in Table 1. The course is now an asynchronous, mastery-based College Algebra course with a single instructor. Asynchronous as all assignments (homework, discussion activities, Progress Quizzes, and Final Exam) are

[3] A detailed breakdown of how a student's grade is determined is provided in the Appendix.

completed in specific time windows rather than at synchronous, specific times. Mastery-based as the primary assessment method is mastery-based testing.

Table 1. Changes to College Algebra by term.

Term	Course Type	Assessments	Grading Scheme
Fall 2017	- Online lectures - Weekly 50-minute discussions with GTAs	- Weekly HW and Quiz (100 pts) - Weekly Discussion activities (100 pts) - 3 Exams: 17 MC, 3 FR (225 pts) - 1 Final: 25 MC (75 pts)	**A**: 435-500 pts **B**: 385-434 pts **C**: 335-384 pts **D**: 285-334 pts **F**: 0-284 pts
Spring 2018 - Fall 2019	- Online lectures - Weekly 50-minute discussions with GTAs	- Homework, due date end-of-semester[1] - Weekly Discussion activities[1] - 4 Exams: 5 MC/module (pass 80%+), 4 modules/exam - 1 Final: 25 MC[1]	**A**: 8 core, 4 advanced passed **B**: 8 core, 2 advanced passed **C**: 8 core passed **D**: 6-7 core passed **F**: 0-5 core passed
Spring 2020 - Present	- Online lectures - Asynchronous discussions	- Homework, due date end-of-semester[1] - Discussion activities - 10 Exams: 10 MC/module (pass 80%+), 2 modules/exam - 1 Final: 25 MC[1]	**A**: 8 core, 4 advanced passed **B**: 8 core, 2 advanced passed **C**: 8 core passed **D**: 6-7 core passed **F**: 0-5 core passed

[1] *Homework, Discussions and Final Exam determined +/- for letter grade as detailed in Appendix.*

3 CHALLENGES AND SOLUTIONS

There were five major challenges to implementing and running the large-scale, mastery-based College Algebra course as an individual instructor:

- Creating numerous assessments needed for mastery-based testing;
- Grading numerous assessments and providing students grade estimates;
- Providing detailed, individualized feedback to students;
- Implementing synchronous discussion sessions with teaching assistants; and
- Countering student unfamiliarity with the new course structure.

This section describes why these challenges appeared and how they were overcome. They may prove enlightening for those implementing parts of the course design.

3.1 Creating Numerous Assessments

The most daunting challenge to employing mastery-based testing in a course is the sheer number of assessments that are required (Armacost and Pet-Armacost, 2003). Large-enrollment courses, with hundreds of students to a single instructor, make it a near-impossible task to grade free-response work for mastery. I instead opted to use an all multiple-choice assessment. Yet, with sixteen modules and ten questions per module, some Progress Quizzes required the generation of 160 questions per version! Adding in the challenge of creating multiple-choice questions that students cannot work backwards from options to the correct answer made it impossible to create these assessments by hand.

Working with a colleague, Dr. Russell Jeter, we utilized automatic item generation[4] (Gierl et al., 2015) to procedurally-generate multiple-choice assessments for mathematics. We named our program Auto-DIG (**Auto**matic **D**iagnostic **I**tem **G**eneration) as it focuses on associating item options to common student conceptions and misconceptions in order to diagnose why a student may have answered an item incorrectly (Chamberlain Jr. and Jeter, 2019, 2020). By answering a series of questions on a certain concept, Auto-DIG could help a student drill-down into a concept and uncover any misconceptions or under-developed conceptions they may hold. The method consists of three to four steps: (1) Creating an item model, (2) Identifying the item content, (3) using computer-based algorithms to procedurally solve both the problem and "nearby" problems, and, if necessary, (4) disguising numeric options in intervals. A short summary of each step is provided below and will be followed by an example question students answered in the course.

Step 1. Creating the item model

This step shares many similarities with writing a free-response question. In order to procedurally-generate versions of the question, elements of the stem and problem that can be modified must be identified at this point. A 1-layer model should be developed if only some small number of elements in the model can be modified. For example, generating the equation $a(x-b)=c$ where a, b, and c are integers would be a 1-layer model as a small number of elements can be modified all at once. An n-layer model should be developed if many elements at multiple levels in a model can be modified (Gierl and Lai, 2013). For example, generating a system of two linear functions to be solved could require an 2-layer model if the model first considers whether there will be 0, 1, or infinitely many solutions (layer 1), then generates coefficients to ensure this condition holds (layer 2). To be clear - this step requires careful **human** planning to create a free-response question that can be procedurally-generated. We are not suggesting a computer would be able to develop and create the stem and mathematical problem of the question.

Step 2. Identifying the item content

[4]Automatic in that items are generated with little to no human interaction.

In this step, the content knowledge required to solve the problem is determined. To accommodate the development of plausible distractors, any common errors or misconceptions associated to the problem should also be determined here. This can be collected by content specialists recalling common errors or misconceptions they are familiar with, recording any common errors identified in educational research experiments, or theoretically-predicted errors or misconceptions according to published mathematics education theoretical perspectives (Chamberlain Jr. and Jeter, 2019).

Step 3. Procedurally solve the problem and nearby problems

Here we start to distinguish hand-written multiple-choice items and the automatic item generation. In this step, the content knowledge collected in step 2 is utilized to procedurally solve the problem. In addition, distractor solutions are be generated by:

1. Isolating common conceptual misunderstandings or common errors related to the topic assessed by the problem.
2. Using these misunderstandings and/or errors to construct "nearby problems".
3. Algorithmically solving these nearby problems to create a list of distractor solutions.

As a simple example of what a "nearby problem" looks like, consider solving the equation $a(x - b) = c$ where a, b, and c are integers. A common error students make when solving a problem requiring distribution is to not distribute (Booth et al., 2014). To emulate this error and make a "nearby problem", we algorithmically solve the equation $ax - b = c$. Procedural generation of the stem, mathematical problem, and solution is becoming more common in online homework systems such as MyLab and WeBWork. The additional step of identifying, creating, and solving nearby problems however is not.

Step 4. Disguise numeric options in intervals

Numeric answers are an issue for multiple-choice question integrity when a student can plug in options and circumvent the targeted mathematical content being assessed. One method to disguise numeric options is to generate disjoint intervals for each answer. This may increase the cognitive load for students when choosing an option but will prevent working backwards from options to answer.

Example question implemented in Spring 2020

The question in Table 2 was taken from the first exam administered in Spring 2020 that corresponded to the content objective "Solve a linear equation with rational coefficients." The question was initially designed with two common student errors in mind: not distributing the negative in front of the second fraction and considering the fraction as only dividing the first term in the numerator. Through random generation and meeting with students, a more nuanced thought process emerged: if a numerator term is divisible by the denominator, divide those terms consistently throughout the rest of fractions. This lead to a new distractor (considering the fraction as only dividing the first term in the numerator) and controlling the generation of the equation problem to allow either the first terms or second terms to be divisible by at least one of their denominators. For evidence that students' responses on an assessment aligned with this specific thought process, consider the response rates in Table 2. We see that the most common error was not distributing the negative in front of the second fraction (22% response rate) and aligns with results from research literature (Booth et al., 2014). The next most common error was consistently dividing the second term's coefficient in the numerator by the denominator (13% response rate). Note that no student chose the option associated to dividing only the first term's coefficient. These results suggest the students' responses to the multiple-choice question emulated their thought process during the assessment.

A single question from a single exam by no means provides definitive statistical evidence that responses corresponded to students' thought processes. A robust qualitative and quantitative analysis of the questions implemented throughout the development of the course is currently under review while the actual code used for the course can be found in the Appendix. The goal of the previous example is to illustrate that careful algorithmic design of a stem, problem, and options can provide additional insight into student thinking and procedure. A future goal is to implement these insights into an online homework system with free-response questions to provide targeted feedback based on *how* a student answered and not just whether the answer was right or wrong.

3.2 Progress Quiz Difficulties and Estimating End-of-Semester Grades

One difficulty introduced by the Progress Quizzes was the "uniqueness" of each quiz to the individual student's progress in the course. For example, Student A could have Modules 4 and 6 on their Progress Quiz while Student B could have Modules 1 and 3 on the same Progress Quiz. To address the varied content on the same quiz, I created a Qualtrics survey with the following flow:

- Ask what Module the student was testing on first (no time limit).
- Provide 1 of three versions of this Module and 10 multiple-choice sets[5] for the student to submit their answers for this Module (60 minute limit).
- Ask what Module the student was testing on second (no time limit).
- Provide 1 of three versions of this Module and 10 multiple-choice sets[6] for the student to submit their answers for this Module (60 minute limit).
- Ask students to attach their scratchwork (no time limit).
- Ask students to digitally sign an academic integrity statement (no time limit).

[5] 5 were provided from Spring 2018 through Fall 2019.
[6] 5 were provided from Spring 2018 through Fall 2019.

Table 2. Exam 1 question from Spring 2020 that tests on the course objective "Solve a linear equation with rational coefficients."

	Spring 2020 Q10
Problem	$\frac{4x-3}{3} - \frac{-3x+6}{5} = \frac{3x+4}{2}$
Options	A. $x \in [2, 5]$ B. $x \in [-4, 2]$ *C. $x \in [8, 11]$ D. $x \in [28, 31]$ E. There are no real solutions
Explanation of Distractors	A. Not distributing the negative in front of the second fraction to both terms in the numerator. B. Dividing only the second coefficient in each numerator by the denominator. C. Correct option. D. Dividing only the first coefficient in each numerator by the denominator. E. Believing it was not possible to solve the equation or that a non-integer solution was not valid.
Percentage of Responses	22%, 13%, **59%**, 0%, 6%

After downloading the student responses, I used an Excel workbook with a series of sheets to organize and grade the individual Modules. This allowed the data to be pasted into the first sheet and, with a small number of manual edits to account for student error, return the organized and graded results of the quiz. These results were copied into another Excel workbook meant to track students' modules passed throughout the semester and provide their up-to-date passing report in Canvas. Templates of both Excel workbooks are provided in the Appendix.

Another difficulty introduced by the Progress Quizzes was confusion on what modules each student should be testing on in any given Progress Quiz. Even with weekly updates explaining how to generally determine which modules should be taken, some students were either unsure and asked for confirmation or unable to determine their progress. I then created a Javascript that asked students to input their grades for each module they tested on and responded with the specific two modules they would test on next. The transition to this tech-enabled, personalized response greatly reduced the number of confused emails about which modules to test on. The webpage where this Javascript is hosted is https://sites.google.com/view/mac1105uf/home and is provided in the Appendix.

One difficulty introduced by the mastery-based testing scheme was the estimation of final grades. To counter this difficulty, I first provided students weekly estimates that used an average number of modules passed as shown in Table 3.

Table 3. Pacing of Modules Passed for Final Grade Estimate in Spring 2020.

	A	**B**	**C**	**D**
Progress Quiz 1	1	1	0	0
Progress Quiz 2	2	2	1	1
Progress Quiz 3	3	3	3	2
Progress Quiz 4	5	4	3	2
Progress Quiz 5	6	5	4	3
Progress Quiz 6	7	6	5	4
Progress Quiz 7	8	7	6	4
Progress Quiz 8	10	8	6	5
Progress Quiz 9	11	9	7	6
Progress Quiz 10	12	10	8	6

While some students found the chart helpful, others lamented that they didn't know how it applied to *their* specific situation. I then used historical data from Fall 2020 to create a linear model using the number of quizzes taken and number of modules passed to estimate a student's final grade. Like the Javascript to help students determine which two modules they should be testing on in any given Progress Quiz, the estimated final grade was hosted on a webpage at https://sites.google.com/view/mac1105uf/home and is provided in the Appendix. Like before, this tech-enabled, personalized response greatly reduced the number of emails asking for final grade estimates and was eventually included in the Canvas gradebook.

3.3 Providing Detailed, Individualized Feedback

Mastery-based testing requires substantial, individualized feedback to help students who did not show mastery on their first attempt. Individualized feedback goes beyond correct/incorrect or a worked-out solution – it provides feedback based on the response a student makes on the assessment (Robinson et al., 2015). A byproduct of developing quality options for the multiple-choice assessments was the ability to provide targeted feedback at-scale. An example of the feedback provided after a Progress Quiz is provided in Figure 1. Note that each option has an explanation of the underlying thought process and there is a general comment to provide additional hints or techniques to avoid issues.

Solve the equation below. Then, choose the interval that contains the solution.

$$-7(-10x - 16) = -19(-14x - 15)$$

The solution is $x = -0.883$, which is option D.

A. $x \in [-3.12, -1.99]$

$x = -2.026$, which corresponds to not distributing the negative in front of the first parentheses correctly.

B. $x \in [1.49, 2.62]$

$x = 2.026$, which corresponds to not distributing the negative in front of the second parentheses correctly.

C. $x \in [-1.19, -0.91]$

$x = -1.182$, which corresponds to getting the negative of the actual solution.

D. $x \in [-0.89, -0.78]$

* $x = -0.883$, which is the correct option.

E. There are no real solutions.

Corresponds to students thinking a fraction means there is no solution to the equation.

General Comment: The most common mistake on this question is to not distribute the negative in front of the second fraction correctly. The best way to avoid this is putting the numerator in parentheses, which will help you remember to distribute the negative correctly.

Figure 1. Feedback for a question aligned with solving linear equations by distributing.

These keys were beneficial to the students and instructor alike. Students could use the explanations to diagnose their own work at their own pace. The instructor could use the key as a crystal ball, telling students where to check in their scratchwork for a potential issue. Every novel misconception or error was added to the automated generation and thus the keys became more effective over time.

For example, a frustrated student emailed after not passing a section on quadratic functions. The student sent a picture of their work and the homework system question (Figure 2) that just was not accepting their "correct" answer and no automated feedback was provided for their particular answer. Based on how the graph was presented as touching the bottom of the grid, the student understood the y-value of the vertex to be 0. The student's difficulty was with reading the graph and not the content objective! After reviewing how to read a graph, the student passed the module and continued on to successfully complete the course. Automating a response to this particular way to read the graph was also added to system.

The previous example illustrates the power of associating students' responses to potential misconceptions and errors. Were the online homework system able to recognize this programmable response, it could provide the necessary feedback to help the student without instructor intervention. This frees up the instructor to engage students in more open-form tasks that encourage abstraction and allows for a single instructor to provide students the necessary procedural feedback they need.

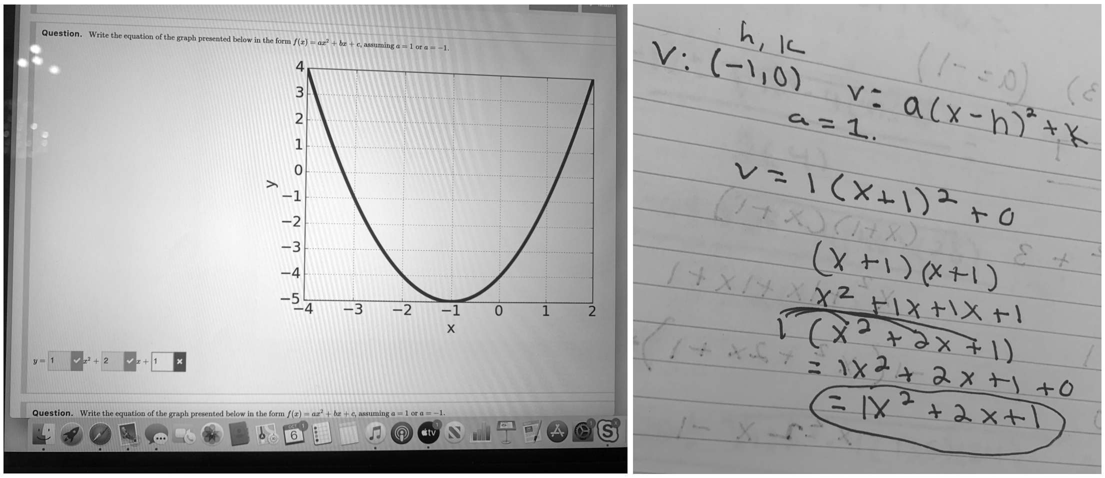

Figure 2. A question asking students to convert from graphical to functional representation for a quadratic function and a student's work.

3.4 Removing Synchronous Discussion Sessions

Before converting to mastery-based assessments, the College Algebra course had numerous in-person discussion sessions where students met with a graduate teaching assistant once a week for an hour. During these sessions students normally asked the TA to either lecture or answer specific homework questions. Lecture and Q&A sessions did not work with the asynchronous, mastery-based nature of Progress Quizzes and so these sessions were converted to open-form assignments meant to extend the concepts students saw in the homework. Students were resistant to engaging in this type of work. Moreover, some graduate teaching assistants did not have the training to teach the open-form assignments effectively. These aspects were eventually scrapped to provide a completely asynchronous version of the course. The author is still working on how to incorporate open-form assignments that are not directly tested on into the course. One positive aspect to making the course completely asynchronous was that it removed the need for additional teaching assistants, furthering the solo-instructor design of the course.

3.5 Countering Unfamiliarity of Course Design with Weekly Discussions

Synchronous discussion sessions were replaced with weekly asynchronous online discussions where the instructor acted as a mediator and students were the primary drivers of knowledge. Example discussion prompts now include discussing various aspects of the course, presenting lessons on topics from the course, and providing emotional support to other students. These discussions occurred asynchronously in the learning management system.

Aspects of the course were discussed for two reasons: (1) some students were resistant to the mastery-based assessments because they did not understand they had the opportunity for productive failure on exams *without it severely affecting their final grade* and (2) illustrate to the class that students who normally had difficulty with math were thriving since they could move through the material at their own pace. To be clear - students were asked to provide their thoughts about a particular aspect of the course and argue for or against its implementation in future iterations. For example, Week 6 focused on mastery-based testing. Students were presented 4 non-technical articles describing mastery-based testing and then were asked to argue for either traditional testing or mastery-based testing. While these assignments improved buy-in from students, it was due to their peers' arguments for mastery-based testing.

Students also needed assistance learning how to use the diagnostic keys. While these keys provided the high-information feedback that research suggests has the largest cognitive benefit (Wisniewski et al., 2020), some students had trouble parsing a hint or feedback that was not directly linked (through marking scratchwork or physical pointing) to some work they completed. The differences between instructor high-information feedback and automated high-information feedback needs further study.

4 REFLECTION ON REDESIGN EFFORTS

Overall, the course redesign was a success in numerous ways. The introduction of mastery-based testing allowed students to work through content at their own pace, satisfying the need for a single course to serve a wide variety of students at different levels. The increase in number of Progress Quizzes from 4 to 10 allowed students more opportunities to attempt to pass a module as well as reduce the number of modules they focused on in a given week. The automated item generation improved the link between student responses on multiple-choice questions and their potential thought processes. The cumulative effect of these boons led to higher student affect in the course. The following subsections provide some evidence for these claims of success.

4.1 Students Needed Multiple Attempts to Pass

The most prevalent student request in the first modification period was to increase the number of Progress Quizzes. Arguments could be summarized in one of two ways: (1) 4 Progress Quizzes didn't give them enough time to practice and retest and (2) testing on four different modules was too much at one time. Starting in Fall 2020, the Progress Quizzes were thus split to only cover two modules (rather than four) and increased to 10 (from four). This resulted in students being able to test on modules up to 20 times rather than the 16 times previously. To determine whether students needed these extra attempts, I analyzed the average number of attempts and the average pass rate per attempt based on the total number of modules passed by the end of the course. The results for Fall 2020 are provided in Table 4.

Double bars were used to denote grade cutoffs. The total percent of students who made 17+ attempts by the end of the course was 45%. Students in the F range of 0-5 total modules passed did not commonly take

Table 4. Fall 2020 Progress Quiz Pass Rate Data (n=288).

Total Modules Passed	% of Total Students	Average % Passed/Taken	Average Attempts	Median Attempts	% of 17+ Attempts per Total Passed
0	9%	0%	5.0	4	0%
1	3%	20%	14.6	8	10%
2	3%	22%	16.9	10	0%
3	3%	36%	16.4	10	13%
4	1%	35%	18.7	13	25%
5	6%	36%	14.8	15	38%
6	3%	40%	15.3	16	40%
7	5%	40%	17.2	18	64%
8	12%	50%	14.9	18	59%
9	3%	53%	15.1	19	70%
10	12%	60%	13.8	18	63%
11	6%	61%	14.3	19	89%
12	33%	76%	15.4	16	46%

advantage of the increased number of module attempts in Fall 2020. However, the final column shows that D, C, B, and A students all made use of the increased number of attempts. Despite low average number of attempts for students who passed a total of 8, 9, 10, and 11 modules, the medians of 18 and 19 illustrate a majority of students needed the additional attempts provided by additional Progress Quizzes to pass the modules. Increasing the number of attempts helped students at all levels.

Moreover, an unexpected boon of the increase in Progress Quizzes was the ease in administering makeups. Since the semester was 16 weeks long, 3 makeup Progress Quiz times were built in to the schedule. This allowed more students to ensure they took 10 Progress Quizzes (assuming they needed all 10) while providing students who were out sick ample opportunity to catch up in the course. Given the data presented in Table 4, a majority of students **at each grade level** took at least 9 Progress Quizzes and thus it was critical to make sure there were ample opportunities for students to make up any missed Progress Quizzes. This proved even more important with the outbreak of COVID-19 and considerations for extended, excused absence from the course.

4.2 Improved Link to Student Thinking

Pre-redesign exams had a large disconnect between the multiple-choice average and free-response average. For example, Fall 2017 averages for Exams 1-3 are presented in Table 5.

Exam	MC AVG	FR AVG
1	78%	43%
2	82%	47%
3	88%	46%

Table 5. Average scores on multiple-choice and free-response portions of exams in Fall 2017.

This large discrepancy suggested the multiple-choice questions may have been too "easy" while the free-response questions may have been too "hard". By redesigning the multiple-choice questions and options, we can improve the connection between students' responses and their potential thinking. Consider the Final Exam questions from Fall 2017, 2018, and 2019 testing on the course objective "Solve linear equations with rational coefficients" in Table 6. Each question in Table 6 fits the description of the content objective "Solve linear functions with rational coefficients" – they are all linear equations with rational coefficients after all. However, there are two issues with the pre-redesign Fall 2017 question that correlate with a student not **solving** the linear equation:

(a) A student could simply check whether each fraction makes the equality true. If a student does this, the associated content objective becomes "reduce compound fractions to show equivalence", which is distinct from the tagged content objective.

(b) A student could take an educated guess that the correct answer, D, is the one that shares either a numerator or denominator with each of the other options. If a student does this, there is no associated content objective.

These alternative explanations for why a student may choose the correct answer bring the content validity of the Fall 2017 item into question. A distractor chosen more than 5% of the time by students suggests it is associated to some thought process and was not chosen by chance (Haladyna and Downing, 1993). The response rates of 4% and 3% suggest they were chosen by chance and not associated to student thinking. Thus, this pre-redesign question does not give the instructor insight into student thinking. Combined with the high percent of correct answers and alternative methods to arrive at the correct answer, the pre-redesign question does not assess student knowledge.

Contrast this with the post-redesign final exam questions linked to the same learning objective. Both questions had at least one distractor chosen more than 5% of the time. Moreover, there is no way to work backgrounds from options to the solution and there is no discernible pattern to guess at the correct answer.

Table 6. Final Exam question from Fall 2017-2019 that tests on the course objective "Solve linear equations with rational coefficients."

	Fall 2017 Q9	Fall 2018 Q9	Fall 2019 Q5
Problem	$\frac{x-2}{9}=\frac{x-4}{2}$	$\frac{-4x-6}{2}-\frac{-4x+6}{5}=\frac{3x+7}{4}$	$\frac{-8x+7}{4}-\frac{-3x+4}{5}=\frac{-3x+6}{2}$
Options	A. $\left\{\frac{40}{7}\right\}$ B. $\left\{-\frac{32}{11}\right\}$ C. $\left\{\frac{34}{7}\right\}$ *D. $\left\{\frac{32}{7}\right\}$	A. $x \in [-1.9, -0.7]$ *B. $x \in [-5.1, -2.1]$ C. $x \in [-11.6, -8.8]$ D. $x \in [-3, -1.9]$ E. There are no Real solutions.	A. $x \in [2, 5]$ B. $x \in [27, 31]$ C. $x \in [-5, 0]$ *D. $x \in [18, 25]$ E. There are no Real solutions.
Percentage of Responses	4%, 3%, 3%, **91%**	20%, **68%**, 3%, 4%, 4%	10%, 3%, 17%, **66%**, 4%

This suggests the options students chose corresponded to their thought process. A more detailed quantitative analysis of the final exams pre- and post-redesign has been completed and the associated technical paper is currently under review.

4.3 Increased Student Affect

While student evaluations of courses are a dubious at best way to evaluate a course (Stroebe, 2020), they can provide insight into student satisfaction with the course. Student satisfaction with the course improved as seen in Table 7.

	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019	Spring 2020	Fall 2020	Spring 2021
Course AVG	3.49	3.17	3.12	3.66	3.39	4.01	3.78	3.76
Dept AVG	4.08	4.28	3.95	4.23	4.02	4.22	4.22	4.20

Table 7. Course Evaluation scores (out of 5) for College Algebra by semester between Fall 2017 and Spring 2021.

There was an initial backlash as the course switched from a traditional grading scheme in Fall 2017 to a mastery-based testing scheme in Spring 2018. After some growing pains that were addressed with the Weekly Assignments and use of technology to provide specific feedback to students (automated keys and Javascript response to questions), students were more satisfied with the revised course than the traditional course. Student evaluations commonly included comments such as:

> [...] the large range of skill in this course requires an option to learn at a different pace. This prevents the advanced from becoming bored and the people first seeing the material from getting weighed

down. I like the design.

The mastery–based set up for the course was very instrumental in my ability to pass the course despite being dyslexic and struggling with math.

I thought the weekly assignments he created were beneficial for our learning and for understanding to the structure of the course. He made sure we knew it was okay to have to retake quizzes and that we would still have time to recieve [sic] an A in the class.

Again, setting aside the questionable value of student evaluations, the increased average course evaluation illustrates that course satisfaction increased after the re-design. This is especially important to note as the ratio of instructor/teaching assistant to student increased with the removal of teaching assistants from the course.

5 LESSONS LEARNED

A series of lessons were learned from the process to converting the College Algebra course to mastery-based testing run by a single instructor. For other instructors looking to make the switch, I've outlined the most important lessons below.

Technology is necessary for one instructor to run the course

A yearly enrollment of 500 to 750 students would be difficult for one instructor to manage in a traditional classroom. Given the increased needs for assessment creation and grading in mastery-based testing schemes, it would be impossible for a single instructor to run the course without technology. After the initial time invested to create various technological supports (primarily Bash/Python scripts), the instructor could complete any of the following tasks in minutes:

- Create individualized assessments of 160 dynamically-generated questions;
- Create detailed keys relating responses on assessments to specific student conceptions and errors;
- Grade and upload hundreds of different assessments at once;
- Track individual student progress and generate a list of students who need an academic intervention.

Completing the tasks quickly was required to run the course effectively as Progress Quizzes were administered on a weekly basis. That meant students needed to have their results and feedback on how to improve early in the week so they could prepare for the next Progress Quiz. To put this in perspective, lets consider the process of creating and grading Progress Quiz 6.

1. Create 480 questions: 10 questions per Modules, 16 Modules total, 3 versions per Module.
2. Create keys for each version of each Module.

3. Load all questions into a repository the quiz will draw from.
4. Download responses from quiz.
5. Match student responses to Module/version for the 20 questions they answered.
6. Update student progress (pass/fail) on each Module to date.
7. Present results to students.

Creating 480 questions aligned with learning objectives within a Module would be a massive undertaking without automation. While creating multiple-choice keys *for grading* would be relatively simple, creating keys *with feedback for each of the 5 response options* would require hand-writing feedback for 2800 potential responses. Beyond the creation of the quiz, mastery-based testing requires dealing with different students testing on different questions. Matching these responses to their quiz taken would be extremely time-consuming by hand, as would manually updating their progress. Completing this entire process on a weekly basis without automation would be impossible. The Bash/Python scripts were necessary to offload tedious, time-consuming work to allow one instructor to complete the tasks that could not be automated, such as recording conceptual videos.

Automated feedback helps with procedural knowledge

The automated feedback accurately targeted common errors and misconceptions students encountered in the course. In early iterations, students would bring their scratchwork to office hours and ask the instructor to figure out why they got a question wrong. Instead the instructor pulled up the associated automated feedback and asked the student to look at specific lines in their work for a specific error (e.g., check line 2 to make sure you distributed the negative in front of the parentheses to both terms). Since feedback for conceptual questions was lacking, the automated feedback primarily targeted procedural understanding. This is consistent with the literature on computer-assisted mathematics learning (Trenholm et al., 2019). Moreover, automated feedback reduced the amount of time the instructor interacted with students to enhance their procedural knowledge.

...but some students need assistance parsing the automated feedback

Some students needed explicit instruction to understand how to use the automated feedback provided after the exam. In future iterations, it may be useful to create a discussion assignment asking students to reflect on the procedural/conceptual errors made during the exam and if these were present in the automated feedback. This type of activity would provide a structured form of self-reflection that is often times missing from a course and can build students' metacognition (Rhem, 2013). Moreover, spending time helping students understand the automated feedback during one assignment early in the semester would reduce the amount of time the instructor needs to individually help students parse the automated feedback.

Frequent communication with students is necessary

The mastery-based assessments allowed students to test and retest on topics they still needed to develop. While some students thrived in this flexible environment, others struggled to stay on-track without rigid dates to complete tasks by. Students appreciated weekly reminders of what they should be doing in the course based on the amount of modules they had passed to that point. Words of encouragement and motivation were also helpful for students who were not used to productive failure and the idea that not every assignment would affect their final grade.

Additional instructor-student time is necessary for a personal touch

In early iterations, students lamented that the course was self-taught as:

- There were no traditional lectures that highlighted information to appear on an exam;
- Feedback was automated and thus not provided by the instructor; and
- Students were responsible for reviewing their own work to determine how to improve.

To addresss some of these complaints, the instructor provided video and/or audio recorded feedback to add a personal touch to automated feedback. In addition, a weekly discussion was devoted to outlining the work the instructor put in to the course ahead of time to provide immediate feedback and asked for ways to improve. Another weekly discussion was devoted to the students' role in learning and how it differed from secondary school. These discussions raised student awareness of the backstage role the instructor plays in their learning and that traditional lectures were not the only method to teach a course. Moreover, it highlighted the use of automation to ensure that one instructor could sufficiently interact with all of the students needs on procedures, concepts, and non-content material.

6 Future Work

Future work on the course would include polishing the homework content, improving the groupwork aspects of the discussions, and providing a quiz feedback per student based on just the questions they did not answer correctly. More work could also be done on making the assessment generation accessible to instructors as it currently requires a Linux machine and a minor amount of setup to run properly. Moreover, the assessments primarily targeted procedural knowledge. Developing automated feedback for multiple-choice assessments, such as Concept Inventories (O'Shea et al., 2016), would provide ways for instructors to assess conceptual knowledge.

APPENDIX

All resources developed for the course are open-source and available in the authors Github at https://github.com/Darryl-Chamberlain-Jr/ufmac1105. Links to individual resources are provided below.

- Syllabus: https://github.com/Darryl-Chamberlain-Jr/ufmac1105/blob/master/MAC1105%20Spring%202021%20Syllabus.pdf
- Course Learning Objectives: https://github.com/Darryl-Chamberlain-Jr/ufmac1105/blob/master/Course-goals-and-objectives-Summer-2020.pdf
- Canvas Course Shell: https://lor.instructure.com/resources/fa28b2fd6fad46a5a9c89f5f907cf47f?shared
- Progress Card: https://github.com/Darryl-Chamberlain-Jr/ufmac1105/blob/master/Progress%20Card%20-%20Fall%202020.pdf
- Automated Assessment Generation: https://github.com/Darryl-Chamberlain-Jr/AAG-College-Algebra
- Open-Source Homework System: https://xronos.clas.ufl.edu/ufmac1105
- Excel Assessment Grading Template: https://github.com/Darryl-Chamberlain-Jr/ufmac1105/blob/master/Progress%20Quiz%20Cleaning%20and%20Analysis%20Template.xlsx
- Javascript to Determine Progress and Estimate Final Grade: https://sites.google.com/view/mac1105uf/home

REFERENCES

Armacost, R. L. and Pet-Armacost, J. (2003). Using mastery-based grading to facilitate learning. In *33rd Annual Frontiers in Education, 2003. FIE 2003.*, volume 1, pages T3A–20. IEEE.

Boersma, S., Hartzler, R., and Savina, F. (2017). Driving the design of the stem-prep pathway. unpublished, retrieved from https://dcmathpathways.org/sites/default/files/resources/2017-06/Driving%20the%20design%20of%20the%20stem-prep%20pathway.pdf.

Booth, J. L., Barbieri, C., Eyer, F., and Paré-Blagoev, E. J. (2014). Persistent and pernicious errors in algebraic problem solving. *The Journal of Problem Solving*, 7(1):3.

Carlson, M., Oehrtman, M., and Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction*, 28(2):113–145.

Chamberlain Jr., D. and Jeter, R. (2019). Leveraging cognitive theory to create large-scale learning tools. In *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education*, pages 741–747.

Chamberlain Jr., D. and Jeter, R. (2020). Creating diagnostic assessments: Automated distractor generation with integrity. *Journal of Assessment in Higher Education*, 1:30–49.

Elsinger, J. and Lewis, D. (2019). Applying a standards-based grading framework across lower level mathematics courses. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, pages 1–23.

Gierl, M. and Lai, H. (2013). Using automated processes to generate test items. *Educational Measurement: Issues and Practice*, 32:36–50.

Gierl, M., Lai, H., Hogan, J., and Matovinovic, D. (2015). A method for generating test items that are aligned to the common core state standards. *Journal of Applied Testing Technology*, 16:1–18.

Haladyna, T. M. and Downing, S. M. (1993). How many options is enough for a multiple-choice test item? *Educational and psychological measurement*, 53(4):999–1010.

Harsy, A. (2020). Variations in mastery-based testing. *Primus*, 30(8-10):849–868.

Herriott, S. R. and Dunbar, S. R. (2009). Who takes college algebra? *PRIMUS*, 19(1):74–87.

Knight, M. and Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, 103(1):65–92.

O'Shea, A., Breen, S., and Jaworski, B. (2016). The development of a function concept inventory. *International Journal of Research in Undergraduate Mathematics Education*, 2(3):279–296.

Rhem, J. (2013). *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy*. Stylus Publishing, LLC.

Robinson, M., Loch, B., and Croft, T. (2015). Student perceptions of screencast feedback on mathematics assessment. *International Journal on Research in Undergraduate Mathematics Education*, pages 363–385.

Smith, G., Ferguson, D., and Caris, M. (2003). The web versus the classroom: Instructor experiences in discussion-based and mathematics-based disciplines. *Journal of Educational Computing Research*, 29:29–59.

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42(4):276–294.

Trenholm, S., Alcock, L., and Robinson, C. (2015). An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology*, 46:1197–1221.

Trenholm, S., Peschke, J., and Chinnappan, M. (2019). A review of fully online undergraduate mathematics instruction through the lens of large-scale research (2000-2015). *PRIMUS*, 29(10):1080–1100.

Verschaffel, L., Van Dooren, W., and Star, J. (2017). Applying cognitive psychology based instructional design principles in mathematics teaching and learning: introduction. *ZDM*, 49(4):491–496.

Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Frontiers in Psychology*, 10:3087.