The Impact of Formative Assessment and Problem-Solving on Course Completion and Grades in higher Mathematics Education

Dr. Ilana Ram, Technion Israel Institute of Technology

Ilana is a research associate at the Technion-Israel Institute of Technology. Her research focuses on higher STEM education.

Adi Wolf
Dr. Lisa Bosman, Purdue University (West Lafayette, IN, USA)

Dr. Lisa Bosman, PhD in Industrial Engineering, is an Associate Professor in the Purdue Polytechnic Institute at Purdue University. Her engineering education research interests include entrepreneurially minded learning, interdisciplinary education, and faculty professional development.

The Impact of Formative Assessment and Problem-Solving on Course Completion and Grades in higher Mathematics Education

Abstract: Large-scale Mathematics courses for engineering students pose challenges for instructors and students alike. These include encouraging students to stay engaged throughout the term and providing meaningful feedback to minimize attrition and improve learning outcomes at scale. Formative assessments and problem-solving activities were suggested as ways to address these challenges. However, their optimal quantity, timing, and context in higher Mathematics education remains an open question. To address it, we analyzed learning activities and outcomes in four runs of a Modern Algebra course (n=1564) for Computer Science and Electrical Engineering undergraduate students. We found that formative assessments increased students' likelihood of taking the final exam, but not necessarily passing it. Focusing on students' clickstream data of problem-solving activities in one course run (n=352) revealed similar patterns. We discuss our findings in the context of the formative assessment and problem-solving literature and offer design implications.

Keywords: Formative assessment, problem-solving, advanced algebra education, learning analytics.

Introduction and background

Higher Sciences, Technology, Engineering, and Mathematics (STEM) education plays a key role in preparing today's students for tomorrow's changing world and industry. Within STEM, special attention is given to Mathematics education, in preparing students for new challenges and developing their 21st-century skills [1]. These skills include critical thinking, reasoning, and complex problem-solving [2]. Supporting students in problem-solving skill development requires instructional support and feedback [3]. However, large scale higher STEM courses are limited in the possibility to give such support [4].

One way to address these challenges of scale is using educational technologies in and outside the classroom [5]. Such technologies can support learning by digitalizing and personalizing content, feedback or assessment. However, the impact of digital formative assessment on students' problem-solving skills is understudied [6].

To address this gap, we study the association between the timing and context of digital formative assessments and problem-solving activities and between learning outcomes and attrition in large enrollment undergraduate mathematics courses. In the next sections, we review the literature on formative assessments and problem-solving, present the current work and our results, discuss our findings in the context of the formative assessment and problem-solving literature and bring instructional design recommendations.

Literature review

Learning assessment and feedback

Assessments of learning and feedback are integral to learning processes and academic education and can be designed to support different learning goals [7]. Formative assessments (FA) can act as Assessment for Learning (AfL) and Assessment as Learning (AaL). AfL are assessments conducted by instructors to evaluate instruction quality and students standings, while AaL allow students to self-evaluate and serve as learning instances on their own. Summative assessments on the other hand usually serve as Assessment of Learning (AoL), evaluating students' performances against a criterion and for certification [7], [8]. The different goals are reflected by the types of feedback provided and by the weight assigned to the assessment (e.g., points in final grade).

Considering the scale of most STEM courses, providing optimal formative feedback is a challenge. Notably, recent work points to the potential of automation and personalization of feedback to address these issues. In [9], the authors found that semi-personalized automated feedback improve students' learning outcomes and feedback satisfaction in a large enrollment computer sciences course. Similarly, [10] found that automated feedback shortened students' path to correct solution. However, only 36% of students' resubmissions based on the feedback. These studies show that feedback can be automated and personalized at scale, but not all students make optimal use of it, leaving open questions regarding its implementation in large enrollment mathematics courses.

Problem-solving Activities: Timing and context

Similarly to FAs, much work points to the importance of problem-solving (PS) for learning in general and math in particular. However, the PS literature is mixed regarding the optimal timing and context of PS activities. Prior work shows that allowing learners to attempt PS before instruction benefits constructive reasoning and transfer of learning, but not procedural knowledge [11]. On the other hand, other studies point to the importance of designing PS activities as worked examples with solutions [12]. The literature further suggests that students' prior knowledge and type of knowledge required for the solution (i.e., procedural, conceptual) is key for understanding the impact PS has on learning outcomes in different contexts [13]. Thus, there are still open questions regarding the integration of these nuances into large enrollment mathematics courses considering variations in prior knowledge and types of knowledge required.

The Current work

The current work is set to address these gaps and evaluate the association between different types of instructional designs on two meaningful outcomes in higher mathematics education: course completion and final exam scores. Our research questions are what is the optimal combination of formative assessments to support course completion (RQ1) and improve final exam scores (RQ2). We hypothesize that in course runs with more FAs, course completion rates will be higher, and final exam grades will improve. We are also interested to see if students' decision to work with different types of PS activities contribute to explained variance in these two outcomes (RQ3). To address these questions, we compared four runs of an advanced Algebra course in a research intensive technological university in [blinded for review].

Method

Context and Participants

The course *Modern Algebra* is an advanced course covering topics such as groups, rings, and fields (see Appendix A). The course serves Computer Science and Electrical Engineering undergraduates in the 3^{rd} or 4^{th} term at a research-intensive technological university in (country). Here, we report the results of 1564 students in total (300-400 students every term; See Table 1). The course made systematical instructional changes over a few years while keeping the staff and material constant.

Research Design

We used a between-subjects design comparing four runs of the course over a few years. Table 1 shows the assessment components of each course run. All assessments were low-stakes not exceeding 20% of the final grade. The additional 80% were based on the final exam. The combination of assessments was basis for determining the overall activity level in each course run. The course also included face-to-face weekly lectures and tutorials, and a course site on a Learning Management System (LMS). As preparation for tutorials, the LMS included PS activities in two formats. Once before tutorials as open problems to encourage students to attempt and solve before instruction. After tutorials, the same problems were uploaded again as worked-examples with the solution as presented in tutorial.

Table 1: Formative assessments type, number of enrolled students, and course activity level

Course run	Assessment types, % of course grade, frequency	Students enrolled	Activity level
2018-19	Graded homework (10%) – six times a term	413	No activity
2020-21	Graded homework (8%) – six times a term Digital quizzes (12%) – twice a term Group projects (10%) – twice a term	417	High activity
2021-22	Graded homework (8%) – six times a term Digital quiz (7%) – once a term Group project (5%) – once a term	382	Low activity
2022 - 23	Graded homework (8%)– six times a term Digital quiz (2%) – once a term Group project (5%) – once a term Digital graded homework (5%) – five times a term	352	Medium activity

Graded homework.

Available every other week on the LMS, submitted alone or in pairs, and included five to six open-ended questions. One question was graded and received feedback as errors marking or checked-up correct solutions. No process feedback was provided. Grades were assigned based on submission (yes/no) and the correctness of the one graded question (*AoL*)

Graded digital homework.

Weekly web-based multiple-choice questions available on the LMS. Questions varied in difficulty, from basic to exam-level questions. The questions allowed multiple attempts over a few days with two phases of feedback. Before the final submission, feedback indicated if there was an error anywhere in the question. Immediately after response submission, students received automated grade feedback. After the final submission, the feedback pointed at the error and at the correct answer. Feedback was specific to the error made (*AfL* and *Aal*).

Digital Quizzes.

Multiple-choice, exam-level items delivered remotely on a specific timeslot with one submission allowed. To maintain academic honesty, the quiz included different items that were randomly selected from a pool of questions. Distractor order was randomized within each question and between students. Immediate grades and feedback were available when the timeslot ended for the entire class *(AoL)*. Teaching staff was available online to answer questions.

Group Projects.

Projects were three-hour synchronous hackathon-like meetings. The activity included groups of 4-5 students each, solving questions on extra-curricular subjects, such as RSA encryption systems demonstrating Euler's theorem and modular arithmetic, or Permutation Games (Rubik cube, 15 Puzzle) demonstrating properties of the symmetric group S_n (See Appendix A). The questions were implemented as an interactive assignment in the LMS, or as a Phyton code assignment. Activity was designed to emphasize the role mathematics plays in engineering applications. Students attended face-to-face or remotely. Project preparation included readings, short videos, and worked examples. Feedback was non-formal and given during the session (*AaL* and *AfL*). Teaching staff (2-5 persons) answered questions on-site and remotely (~200 students participating each term).

Data Preparation and Analysis

For comparability, we computed each term's total FA grade to be out of 100 percent. Other variables used in the analysis were students clickstream data on the course's LMS for two types of PS (before and after tutorial) and number of clicks made during the quiz, which is also a PS activity. This data was only available for the 2022-2023 term (Medium activity).

To answer RQ1, regarding course completion, we used a Chi-Square test comparing the percent of students who took the final exam in different course runs with post-hoc analyses. To answer RQ2, regarding final exam scores, we explored the grades spread and correlations between course activity level and exam grade. To answer RQ3, regarding students' PS activity, we focused on the Medium Activity term and conducted two Multiple Linear Regression (MLRs; stepwise inclusion criterion) exploring the factors explaining variance in both outcomes.

Results

RQ1: Likelihood of taking the final exam and completing the course

The Chi-Square test was significant for the overall table: X^2 (3) = 15.59, p = 0.001. Post-hoc analysis revealed that the *No Activity* term had the lowest percentage of exam takers (adjusted alpha=0.0063 for multiple comparisons). The *Medium Activity* term had the highest percent of exam takers. The difference between the *Medium* and the *Low* terms was ~10%, suggesting that 30 to 40 students who took the final exam in the Medium activity term would otherwise not finish the course. This supports H1. The Low or High Activity terms were not statistically different from the others (See Table 2).

Table 2. Course completion rates by course activity level: Chi square results

Course run	Percent students not taking the exam	Course completers	Students
No activity	28.8%, p = 0.007	**71.2%**, p = 0.004**	413
Low activity	27.2%, p = 0.09	72.8%, p = 0.09	382
Medium activity	**18.5%***, p = 0.006**	**81.5%***, p = 0.006**	352
High activity	20.9%, p = 0.08	79.1%, p = 0.08	417

RQ2: Formative assessments and exam grade

Figure 1 presents exam grade distribution by FA grades (standardized to 100%) in the four terms. The figure shows that the data is not normally distributed with a cluster of students in the upper right corner in all terms. Notably, the number of students enrolled each term is higher than presented in Figure 1 as not all students took the final exam, generating missing data. The No Activity term (top-left), shows a ceiling effect with a cluster on the x-axis at the maximum FA grade and exam grades distributed on the y-axis ($r = 0.21$, $p < 0.001$). The other terms show a similar pattern but a greater spread of FA grades with a cluster of cases in the top-right corner (Low Activity: $r = 0.37$, $p < 0.001$; Medium Activity: $r = 0.44$, $p < 0.001$; High Activity: $r = 0.4$, $p < 0.001$). While the correlations are statistically significant, the grade spread highlights the need for a more nuanced approach. We therefore focused on the Medium Activity term, where course completion and correlation were highest, to study explained variance in course completion and exam grade by FAs and PS.

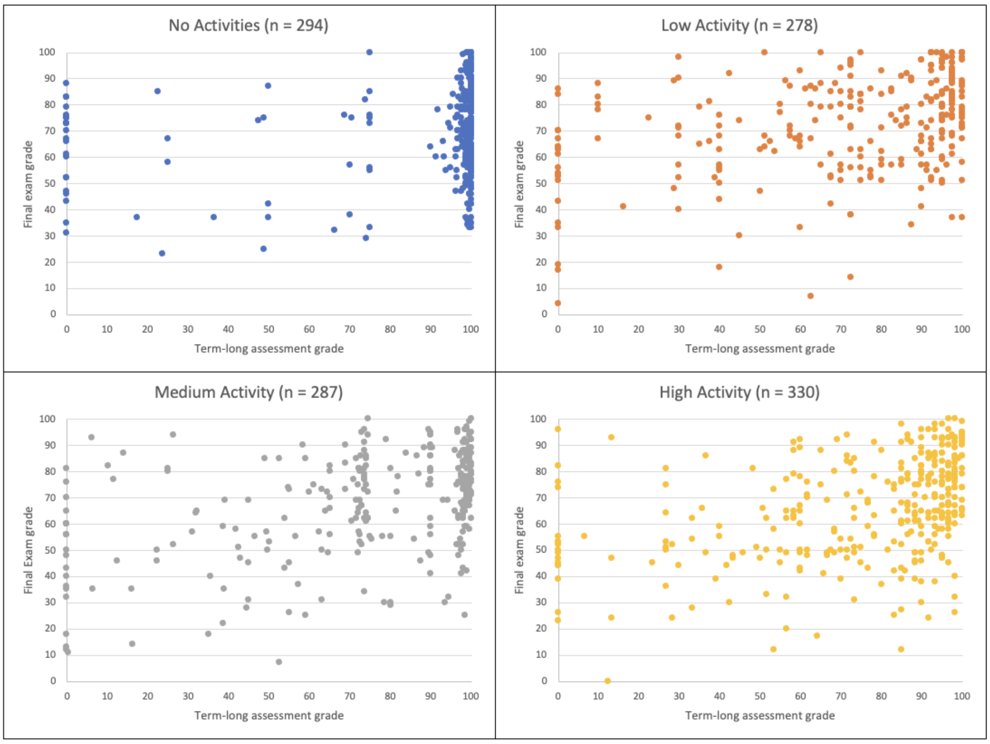


Figure 1. Final exam grade by FA grades (out of 100%)

RQ3: Explaining variance in learning outcomes with FAs and PS

We summarized *students'* weekly clickstream data on PS before and after tutorials on the *Medium activity* term (n=352). For example, a count of 1006 clicks for PS before tutorial on week one suggest that, on average, each student entered this resource three times that week. As seen in Figure 2 (heatmap) at the beginning of the term, students used both resources, with a preference for PS before tutorials. On week three, the use of PS without solutions dropped dramatically and remained low. The use of PS with examples continued until week nine. Then students showed a steep drop until the end of the term in the use of both resources, likely due to the overall workload in the program.

	Problem-solving before Instruction	Worked examples
Week 1	1006	749
Week 2	835	717
Week 3	352	803
Week 4	240	583
Week 5	305	794
Week 6	502	756
Week 7	529	833
Week 8	479	756
Week 9	165	563
Week 10	280	391
Week 11	274	368
Week 12	243	321
Week 13	174	282

Figure. 2. LMS click summary for PS before and after tutorials (2022-2023)

Table 3 shows the results of two MLRs for course completion and exam grade (RQ3) based on different FA *components* and PS click count. Students' engagement with process-oriented course components (PS without solutions, digital homework with formative feedback; AfL and AaL) explains 16.5% of variance in course completion. 18.7% of variance in final exam scores is explained by performances-oriented components (quiz and graded homework grades, AoL).

Table 3. Predictive models for course completion and exam grade

		Course completion (0=no; 1=yes)	**Final exam grade (0-100 scale)**
# of students in model	N	351	288
Percent explained variance	R^2	16.5%	18.7%
Digital formative assessment grades	Digital quiz grade (AoL)	β = 0.03	**β = 4.05****
	Digital homework (AfL & AaL)	**β = 0.08*****	β = 0.11
	Project participation (0=n, 1=y) (AfL & AaL)	β = 0.09	β = 0.07
	Traditional homework (AoL)	β = 0.05	**β = 2.24*****
Problem-solving (clickstream)	PS - # of clicks before tutorial	**β = 0.003****	β = - 0.02
	PS - # of clicks with solutions	β = 0.05	β = - 0.01
	Digital quiz: # of clicks	β = 0.09	β = 0.01

Discussion

We studied the relationship between digital FAs and PS on course completion and exam scores by comparing four runs of an Algebra course and by focusing on one course run with two PS formats. We found that providing students with meaningful, just-in-time formative feedback increases course completion rates. This echoes prior work highlighting the benefits active learning has for learning outcomes in STEM education [14] and the role assessment can have *As* and *For* Learning [7]. Interestingly, students' engagement with PS before instruction is associated with taking the exam, albeit not necessarily passing it. This is in line with work showing that passing but low grades can indicate college dropout early [15]. Moreover, students show a preference for one type of PS format over the other throughout the term. This pattern possibly reflects prior knowledge differences, as suggested by prior work [13], [16]. Our findings thus show how different course components relate to different outcomes pointing to the possibility to personalize instruction in challenging topic areas in math and at scale.

Several design suggestions can be based on our findings. First, the optimal quantity of FA includes at least two stand-alone events with formative feedback. There was no advantage to adding more workload (i.e. another group project or another quiz) for course completion rates. This points to the optimal quantity of these activities in terms of personnel, course resources, and learning outcomes. Finally, while students prefer PS with solutions, accessing PS without solutions is the resource associated with course completion. Therefore it is important to provide access to both types of PS formats to address attrition at scale.

Notably, this work also presents several limitations. We report correlations, explain variance and contingencies, and refer to clickstream data. We do not however analyze the content of these processes, limiting process related conclusions. Further, while the research design minimizes noise, it may also introduces biases (e.g., same teaching staff). Replications and extensions will therefore help establish the findings and evaluate the design suggestions.

In summary, across four course runs we found a consistent pattern: adding low-stakes, feedback-rich formative assessment is associated with higher course completion (i.e., more students actually reach and take the final exam), but it does not automatically translate into higher passing rates or uniformly higher exam performance. When we zoomed into the term with detailed LMS traces, students' engagement with process-oriented components, especially attempting problem-solving before tutorials and completing digital homework that provided formative feedback, was most strongly associated with persistence to the end of the course, whereas performance-oriented measures (e.g., quiz and traditional homework grades) were more closely tied to final exam scores.

Taken together, our results suggest that "what helps students stay" is not identical to "what helps students score," and that the most efficient design for large-scale mathematics courses may be a focused set of well-timed formative feedback opportunities alongside access to both unsupported problem-solving and worked examples, rather than simply increasing the overall volume of assessed activity. Future work should combine trace data with finer-grained process evidence (e.g., solution strategies, error types, and prior knowledge measures) to identify which learners benefit from which sequencing and feedback designs, and to support more targeted personalization at scale.

References

[1] Y. Li and A. H. Schoenfeld, "Problematizing teaching and learning mathematics as 'given' in STEM education," *Int. J. STEM Educ.*, vol. 6, no. 1, 2019, doi: 10.1186/s40594-019-0197-9.

[2] T. J. Kennedy and C. W. Sundberg, "21st Century Skills BT - Science Education in Theory and Practice: An Introductory Guide to Learning Theory," B. Akpan and T. J. Kennedy, Eds. Cham: Springer International Publishing, 2020, pp. 479–496.

[3] Yengoyan, "L. S. VYGOTSKY: 'Mind in society: The development of higher psychological processes,'" *Lang. Soc.*, vol. 9, p. 249, 1980.

[4] R. Pettit and J. Prather, "Automated assessment tools: too many cooks, not enough collaboration," *J. Comput. Sci. Coll.*, vol. 32, no. 4, pp. 113–121, 2017.

[5] S. Beaudry, "Scaling Active Learning Classrooms Adopt 11 Best Practices to Transform Existing Spaces to Support Student Success," no. June, pp. 15–25, 2022, [Online]. Available: www.scup.org/phe.

[6] K. K. Bhagat and J. M. Spector, "Formative Assessment in Complex Problem-Solving Domains: The Emerging Role of Assessment Technologies," *J. Educ. Technol. Soc.*, vol. 20, no. 4, pp. 312–317, 2017.

[7] N. J. Rao and S. Banerjee, "Classroom Assessment in Higher Education," *High. Educ. Futur.*, vol. 10, no. 1, pp. 11–30, 2023, doi: 10.1177/23476311221143231.

[8] P. Black and D. Wiliam, "Developing the theory of formative assessment," *Educ. Assessment, Eval. Account.*, vol. 21, no. 1, pp. 5–31, 2009, doi: 10.1007/s11092-008-9068-5.

[9] A. Pardo, "Using learning analytics to scale the provision of personalised feedback," vol. 50, no. 1, pp. 128–138, 2019, doi: 10.1111/bjet.12592.

[10] H. Gabbay and A. Cohen, *Exploring the Connections Between the Use of an Automated Feedback System and Learning Behavior in a MOOC for Programming*, vol. 13450 LNCS. Springer International Publishing, 2022.

[11] T. Sinha and M. Kapur, "Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension," *Learn. Instr.*, vol. 75, no. May, p. 101488, 2021, doi: 10.1016/j.learninstruc.2021.101488.

[12] R. Fedler and R. Brend, *Teaching and Learning STEM: A Practical Guide*, vol. 70, no. 5. 2016.

[13] O. Chen, G. Woolcott, and S. Kalyuga, "Comparing alternative sequences of examples and problem-solving tasks: the case of conceptual knowledge," *Educ. Dev. Psychol.*, vol. 38, no. 1, pp. 158–170, 2021, doi: 10.1080/20590776.2021.1915098.

[14] S. Freeman *et al.*, "Active learning increases student performance in science, engineering, and mathematics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 23, pp. 8410–8415, 2014, doi: 10.1073/pnas.1319030111.

[15] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 17–29, 2017, doi: 10.1109/TLT.2016.2616312.

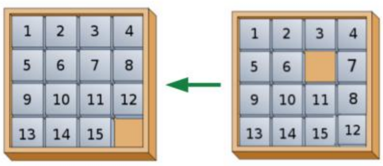
[16] T. van Gog, L. Kester, and F. Paas, "Effects of worked examples, example-problem, and problem-example pairs on novices' learning," *Contemp. Educ. Psychol.*, vol. 36, no. 3, pp. 212–218, 2011, doi: 10.1016/j.cedpsych.2010.10.004.

[17] J. Conklin, "A Taxonomy for Learning , Teaching , and Assessing : A Revision of Bloom ' s Taxonomy of Educational Objectives Complete Edition by Lorin W . Anderson , David Krathwohl , Peter Airasian , Kathleen A . Cruikshank , Richard E . M," *JSTOR*, 2019, [Online]. Available: https://www.jstor.org/stable/42926529.

Appendix A – Course subjects, learning goals and example activities

Subject	Learning goals (referring to revised Bloom's taxonomy; Conklin, 2019)	Learning activities Examples

Properties of Integers and modular arithmetic	**Remember** definitions (e.g., a prime number)
	Understand properties of integers
	Apply the definitions and properties to solve technical exercises
		Example from a project on RSA:
		Question: implement *modular_exponent* function
		`def modular_exponent(a, d, n):`
		`"""`
		`Returns a to the power of d modulo n`
		`Parameters`
		`----------`
		`a : The exponential's base.`
		`d : The exponential's exponent.`
		`n : The exponential's modulus.`
		`Returns`
		`-------`
		`b: such that b == (a**d) % n`
		`"""`
		Tip given to the students during the activity: recall that if $d = \sum_{i=0}^{m} b_i 2^i$ is the binary representation of d, then $a^d \equiv_n \prod_{i=0}^{m} a^{b_i 2^i} (mod\ n)$

Groups, Sub-groups, Cyclic Groups	**Remember** definitions (e.g., a cyclic group) **Understand** properties of groups (e.g,. "the unit element of a group is unique") **Apply** the definitions and properties to solve technical and theoretical exercises	Example from a digital graded homework: Find the order of the following groups: $G = U_{48}$ $G = U_{27}$ $G = U_{270}$ $G = U_{18} \times \mathbb{Z}_{18}$
The Symmetric group S_n	**Remember** definitions (e.g., a transposition) **Understand** properties (e.g., order, inverse etc of an element $\sigma \in S_n$) **Apply** the definitions and properties to solve technical exercises	Example from a project on permutation games: Find a permutation which describes the following move in the 15-Puzzle  Tip given to the students during the activity: use transpositions
Normal sub-groups, cosets, Lagrange's Theorem, Quotient Groups	**Remember** the definition of normal sub-group and the formulation of Lagrange's theorem. **Understand** the concepts of cosets, quotient groups.	Example from a digital quiz: Let G be a group and H a subgroup. For each claim choose if it True or False: • If $[G:H]=3$, $\|G\|=9$ then H is normal in G • If $[G:H]=3$, $\|G\|=6$ then H is normal in G

	Apply the above to solve technical and theoretical exercises.	
Homomorphisms of groups, Isomorphism theorems	**Remember** the definitions of homomorphism, kernel, etc. **Understand** the concept of isomorphism **Apply** the above to solve technical and theoretical exercises.	Example from a project on classifying groups of small order: Show using your code that there are only 2 multiplication tables for groups of order 6 (up to renaming the elements of the group) The students were advised to use backtracking rather than brute force to implement their code.
Rings and Fields: Definitions, Examples and Basic Properties	**Remember** the definition of a ring, an ideal, etc. Understand properties (e.g., a **commutative** division ring with a unity is a field) **Apply** the above to solve technical and theoretical exercises.	Example from a graded homework: Show that if R is a finite integral domain (commutative with unity) then R is a field.
Polynomial Rings, Quotient Rings and Finite Fields	**Remember** the definition of a ring, an ideal, etc. **Understand** properties (e.g.,) **Synthesize** principles from previous subjects in the context of rings **Apply** the above to solve technical and theoretical exercises.	Example from a digital graded homework For each of the following rings, determine if $R_i \cong R_j$ (isomorphism of rings) $R_1 = \mathbb{Z}_{25}$, $R_2 = \mathbb{Z}_5 \times \mathbb{Z}_5$ $R_3 = \mathbb{Z}_5[x]/\langle (x-1)(x-2) \rangle$ $R_4 = \mathbb{Z}_5[x]/\langle x^2+x+1 \rangle$