Changes in assessment practices of calculus instructors while piloting research-based curricular activities

We report our analysis of changes in assessment practices of introductory calculus instructors piloting weekly labs designed to enhance the coherence, rigor, and accessibility of central concepts in their classroom activity. Our analysis compared all items on midterm and final exams created by six instructors prior to their participation in the program (355 items) with those they created during their participation (417 items). Prior exams of the six instructors were similar to the national profile, but during the pilot program increased from 11.3% of items requiring demonstration of understanding to 31.7%. Their questions involving representations other than symbolic expressions changed from 36.7% to 58.5% of the items. The frequency of exam questions requiring explanations grew from 4% to 15.1%, and they shifted from 0.8% to 4.1% of items requiring an open-ended response. We examine qualitative data to explore instructors' attributions for these changes.

Key words: [Calculus, Assessment, Cognitive Level, Representations, Problem-Solving]

One component of the recent national study of calculus programs in the United States (Bressoud, Mesa, & Rasmussen, 2015) examined the assessment practices of instructors of these courses. Tallman & Carlson (2012) analyzed the content of 150 Calculus 1 final exams sampled from a variety of post-secondary institutions in the larger study along three dimensions in their Exam Characterization Framework (ECF) detailing the cognitive orientation, mathematical representations, and answer format of each item. The study demonstrated that few final exam items required a demonstration or application of understanding of the material, primarily involved only symbolic representations, and rarely required explanation or involved open-ended responses. One explanation of these results may be that faculty assessment practices simply reflect the expectations of institutionally adopted curriculum. Lithner (2004), for example, found that a majority of exercises in calculus textbooks could be solved by choosing examples or theorems elsewhere in the text based on surface-level features and mimicking the demonstrated procedures.

We examined the assessment practices of pilot instructors implementing activities in their calculus courses designed to simultaneously enhance the coherence, rigor, and accessibility of student learning throughout the course. Project CLEAR Calculus provided weekly labs in which students participated in group problem-solving activities to scaffold the development of central concepts in the course along with instructor training and support to implement the labs. While the project did not address student assessment through exams, we hypothesized the conceptual focus in the labs and requirements of student write-ups would significantly impact the instructors' assessment practices. We address the following research questions:

1. How do the pilot instructors' exam questions compare to their previous exams along the three ECF dimensions?
2. What factors do the pilot instructors attribute for any shifts in their assessment practices?

Background

Limit concepts are at the core of mathematics curriculum for STEM majors, but decades of research have revealed numerous misconceptions and barriers to students' understanding. Building off of work by Williams (1991, 2001), Oehrtman (2009) identified several cognitive models employed by students that met criteria for emphasis across limit concepts and for sufficient depth to influence students' reasoning. Williams noted that frequently students attempt to reason about limits using intuitive ideas associated with boundaries, motion, and approximation. Oehrtman found that, unlike most other cognitive models employed by students, the structure of students' spontaneous reasoning about approximations shares significant parallels with the logic of formal limit definitions while being simultaneously conceptually accessible and supporting students' productive exploration of concepts in calculus defined in terms of limits. With this in mind, we contend that a false dichotomy exists between a formally sound, structurally robust treatment of calculus on the one hand and a conceptually accessible and applicable approach on the other. By adopting an instructional framework utilizing approximation and error analyses, we designed labs based on criteria listed in Figure 1 intended for weekly use in an introductory calculus sequence.

Design Criteria 1.	Language, notation, and constructs used in the labs should be conceptually accessible to introductory calculus students.
Design Criteria 2.	The structure of students' activity should reflect rigorous limit definitions and arguments without the language and symbolism of formal ε-δ and ε-N notation that is a barrier to most calculus students' understanding.
Design Criteria 3.	The labs should present a coherent approach across all concepts defined in terms of limits and effectively support students' exploration into these concepts.
Design Criteria 4.	The central quantities and relationships developed in all labs should be coherent across representational systems (especially contextual, graphical, algebraic, and numerical representations)
Design Criteria 5.	All labs should foster quantitative reasoning and modeling skills required for STEM fields.
Design Criteria 6.	The sequence of labs should establish a strong conceptual foundation for subsequent rigorous development of real analysis.
Design Criteria 7.	All labs should be implemented following instructional techniques based on a constructivist theory of concept development.

Figure 1. Design criteria for the labs.

When left unguided, students' applications of intuitive ideas about approximations are highly idiosyncratic (Martin & Oehrtman, 2010a, 2010b; Oehrtman, 2009). To systematize students' reasoning concerning approximation ideas and support an accessible yet rigorous approach to calculus instruction, throughout the labs students are engaged in contextualized versions of the questions in Figure 2. These questions develop coherence between structural components, reveal operations performed on these components, and highlight relationships among the operations, foundational for the generation of new understandings.

Question 1.	Explain why the unknown quantity cannot be computed directly.
Question 2.	Approximate the unknown quantity and determine, if possible, whether your approximation is an underestimate or overestimate
Question 3.	Represent the error in your approximation and determine if there is a way to make the error smaller.
Question 4.	Given an approximation, find a useful bound on the error.
Question 5.	Given an error bound, find a sufficiently accurate approximation.
Question 6.	Explain how to find an approximation within any predetermined bound.

Figure 2. Approximation questions consistent across most labs.

Exam Characterization Framework

Tallman and Carlson (2012) developed a three-dimensional framework to analyze a large sample of post-secondary Calculus 1 final exams and generate a snapshot of the skills and understandings that are currently being emphasized in college calculus. Their *Exam Characterization Framework* (ECF) characterizes exam items according to three distinct item attributes: (a) *item orientation*, (b) *item representation*, and (c) *item format*.

Item Orientation

Tallman and Carlson adapted the six intellectual behaviors in the conceptual knowledge dimension of a modification of Bloom's taxonomy (Anderson & Krathwohl, 2001) to characterize the cognitive demand of exam items. The six categories of item orientation are hierarchical with the lowest level requiring students to remember information and the highest level requiring students to make connections (see Table 1).

Table 1
Item orientation codes (adapted from Tallman & Carlson, 2012)

Cognitive Behavior	Description
Remember	Students are prompted to retrieve knowledge from long-term memory.
Recall and apply procedure	Students must recognize what procedures to recall and apply when directly prompted to do so.
Understand	Students are prompted to make interpretations, provide explanations, make comparisons or make inferences that require an understanding of a mathematics concept.
Apply understanding	Students must recognize recognize the need to use a concept and apply it in a way that requires an understanding of the concept.
Analyze	Students are prompted to break material into constituent parts and determine how parts relate to one another and to an overall structure or purpose.
Evaluate	Students are prompted to make judgments based on criteria and standards.
Create	Students are prompted to put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure.

Item Representation

The item representation domain of the ECF involves classification of both the representation of mathematical information in the task as well as the representation the task solicits in a solution (see Table 2). A task statement or solution may involve multiple representations. Since many tasks can be solved in a variety of ways and with consideration of multiple representations, we observed Tallman and Carlson's recommendation of considering only the representation the task requires.

Table 2
Item representation codes (Tallman & Carlson 2012)

Representation	Task statement	Solicited solution
Applied/ modeling	The task presents a physical or contextual situation.	The task requires students to define relationships between quantities or use a mathematical model to describe a physical or contextual situation.
Symbolic	The task conveys information in the form of symbols.	The task requires the manipulation, interpretation, or representation of symbols.
Tabular	The task provides information in the form of a table.	The task requires students to organize data in a table.
Graphical	The task presents a graph.	The task requires students to generate a graph or illustrate a concept graphically.
Definition/ theorem	The task asks the student to state or interpret a definition or theorem.	The task requires a statement or interpretation of a definition or theorem.

Proof	The task presents a conjecture or proposition.	The task requires students to demonstrate the truth of a conjecture or proposition.
Example/ counterexample	The task presents a proposition or statement.	The task requires students to produce an example or counterexample.
Explanation	Not applicable.	The task requires students to explain the meaning of a statement.

Item Format

The third and final dimension of the ECF is item format. The most general distinction of an item's format is whether it is multiple-choice or open-ended. However, there is variation in how open-ended tasks are posed. For this reason, Tallman and Carlson define three subcategories of open-ended tasks: *short answer*, *broad open-ended*, and *word problem*. A short answer item is similar in form to a multiple-choice item, but without the choices. A student can anticipate the form of the solution of a short answer item upon reading the item. In contrast, the form of the solution of a broad open-ended item is not recognizable upon immediate inspection of the item. Broad open-ended items therefore elicit various responses, with each response typically supported by some explanation. Word problems can be of a short answer or broad open-ended format, but prompt students to create an algebraic, tabular and/or graphical model to relate specified quantities in the problem, and may also prompt students to make inferences about the quantities in the context using the model. Also, tasks that require students to explain their reasoning or justify their solution can be supplements of short answer or broad open-ended items.

Exam Characterization Results of the National Sample

Tallman and Carlson coded 14.83% of items in their randomly-selected sample of 150 post-secondary calculus I final exams, collectively containing 3,735 items, at the "Understand" level of the item orientation taxonomy or higher. Their coding also revealed that 34.55% of items in their sample were not stated symbolically and required a symbolic representation in the solution. Also, Tallman and Carlson found that only 1.34% of items in their sample were broad open-ended questions.

Methods

Twelve instructors piloted up to 30 labs in 24 different first and second semester calculus classrooms at eight different institutions from Fall 2013 to Spring 2015. Training began with in-person and online meetings with pilot instructors before the start of the Fall semesters, and most of the instructors attended a three-day workshop outlining the goals, strategies, and activities of the project. We supported their implementation of the labs throughout the fall and spring semesters with online meetings with project personnel. The project website provided instructors with student materials, instructor notes for each lab, solutions, grading rubrics, and supporting handouts and virtual manipulatives. Support meetings frequently included discussions of assessing lab write-ups but did not include discussions of creating or grading exams.

To document changes in the pilot instructors' assessment practices, we collected mid-term and final exams from the calculus classes the instructors taught prior to implementing CLEAR Calculus labs and from the classes in which they were implementing the labs. Five of the instructors either had not previously taught calculus or were required to give exams that were created by other faculty, so these all exams from these instructors were removed from the comparative sample.

A lead researcher in the development of the ECF and its application in the national study trained two members of our team to code with the framework resulting in 89% agreement between coding the training sample. Subsequent training focused on discrepancies. One member of our team has coded 355 items from 21 exams given by six instructors prior to using CLEAR Calculus labs and 417 items from 22 exams given by the same instructors while implementing the labs. A small number of exams remain to be coded, and we will choose a random sample of items to be coded by the second team member and the trainer to determine agreement and resolve discrepancies.

We collected self-reported characterizations on the impact of pilot instructors' teaching and exams through their implementation of CLEAR Calculus labs. We are currently analyzing this data for themes in and for shifts in assessment priorities.

Preliminary Results

Our analysis of exams given by our pilot instructors prior to participating in the project revealed a pattern very similar to the national profile found by Tallman & Carlson (2012) as shown in Table 3. In contrast, while implementing the labs the instructors nearly tripled the frequency at which they asked questions requiring a demonstration or application of understanding (from 11.3% to 31.7%) and included representations other than symbolic expressions at over 1.5 times the previous frequency (36.7% to 58.5%). They asked for explanations nearly 4 times as often (4% to 15.1%) and included broad open-ended items over 5 times as often (0.8% to 4.1%).

Table 3
Shifts in CLEAR Calculus pilot instructor's assessment practices.

	Tallman & Carlson National Sample (3735 items)	Pilot instructors prior to CLEAR Calculus (355 items)	Pilot instructors with CLEAR Calculus (417 items)
Items requiring understanding or higher level reasoning	14.83%	11.3%	31.7%
Items involving representations other than symbolic	34.55%	36.7%	58.5%
Items requiring explanation	2.36%	4.0%	15.1%
Broad open-ended items	1.34%	0.8%	4.1%

Discussion Questions

As our coding is nearly complete our presentation of the complete analysis will be very close to the data shown above. In addition, we will present themes from our qualitative data on instructors' attributions for these changes as well as interesting patterns in the differences of the individual instructors. We will seek a discussion with the audience on questions of additional ways to analyze the ECF data to reveal other insights, potential follow-up questions to pursue with the instructors represented in this data, and additional data we may collect as we work with our third round of pilot instructors.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Bressoud, D., V. Mesa, C. Rasmussen (2015). *Insights and Recommendations from the MAA National Study of College Calculus*. MAA Press.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *Journal of Mathematical Behavior*, *23*, 405-427.

Martin, J., & Oehrtman, M. (2010a). Part / whole metaphor for the concept of convergence of Taylor series. In P. Brosnan, E. Diana, & L. Flevares (Eds.) *Proceedings of the Thirty Second Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 97-104). Columbus, Ohio: The Ohio State University.

Martin, J., & Oehrtman, M. (2010b). Strong metaphors for the concept of convergence of Taylor series. In *Proceedings of the Thirteenth Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC: North Carolina State University.

Oehrtman, M. (2009). Collapsing dimensions, physical limitations, and other student metaphors for limit concepts. *Journal for Research in Mathematics Education*, 40(4), 396-426.

Tallman, M. & Carlson, M. P. (2012). A characterization of calculus I final exams in U.S. colleges and universities. *Proceedings of the 15th Annual Conference on Research in Undergraduate Mathematics Education* (p. 217-226). Portland, OR: Portland State University.

Williams, S. (1991). Models of limit held by college calculus students. *Journal for Research in Mathematics Education, 22*(3), 219-236.

Williams, S. (2001). Predications of the limit concept: An application of repertory grids. *Journal for Research in Mathematics Education, 32,* 341-367.