Chasing assessment: The faculty experience of trying to implement evidence based practices well

Dr. Todd M. Fernandez, Georgia Institute of Technology

Todd is a lecturer in the Wallace H. Coulter Department of Biomedical Engineering at Georgia Institute of Technology. His research interests are engineering students beliefs about knowledge and education and how those beliefs interact with learning engineering.

Mr. David Edward Torello, Georgia Institute of Technology

Dr. David Torello graduated with his B.S. in mechanical engineering from UC Berkeley and his M.S. and Ph.D. from the Georgia Institute of Technology. He is currently non-tenure track faculty in mechanical engineering at Georgia Tech, lecturing in mechanic

Chasing assessment:
The faculty experience of trying to implement evidence based practices well

A significant focus in engineering education research is encouraging faculty to improve the way they teach [1], [2], [3]. A research to implementation pipeline exists to achieve that goal - new approaches are developed and disseminated, faculty are trained or supported in implementation, and then results are evaluated [4], [5], [6]. Such efforts are often measured through increased use of evidence-based practices [7] as part of a process where faculty are introduced to new teaching techniques and implement them. Studies that evaluate change by measuring the amount and type of concrete changes to teaching are the norm [1], [2], [3]. However, two areas of alternative evaluation approaches have been called for, those that evaluate the fidelity of implementations [7] and those that evaluate the impact changes have on students [6], [8], [9]. Both shift away from solely measuring what faculty do. While measuring fidelity and learning are important improvements, they leave a gap in our understanding of course and curricular change.

Little work documents *how* faculty affect changes in their courses, especially change independent of major faculty development efforts and structures. In parallel, there is a "dissatisfaction with the rate of implementation, adoption, and scale-up of research-based instructional strategies (RBIS)" [2, p. 221] that is apparent in thought leaders and funding agencies. The lack of impact from RBIS driven change efforts creates a nearly circular effect - low impact suggests the need for more programs that have limited impact, which suggests the need for more programs. However, we know little about how faculty experience these calls for change or put them into practice. What we do know about the *results* of change efforts is illustrative. First, some faculty see coordinated efforts as disempowering when historically informal academic systems are formalized [10]. Second, many, if not most, teaching changes are not sustained over time [11], [12], [13]. Finally, when changes are made, they are often low fidelity - meaning they partially implement the instructional strategies they claim [7]. While these results are important, they merely reinforce what we know - current change process do not deliver the intended outcomes. However, current research does little to answer why. Is low fidelity an issue of applicability or practicality? Do faculty members not understand or not have the time to fully implement techniques? What causes changes to be sustained, improved, or dropped? Understanding these requires understanding *how* not *if* change is made.

This paper looks at the experience of one faculty in one course engaged in a common first course improvement project - changing course assessment. We chose this narrow focus because any efforts at large scale curricular change are a composite of such individual efforts to change a course[1]. A single type of change in a single course represents one of the most granular events that can be analyzed as part of curricular change work. In doing so this paper addresses two gaps in the existing threads of research described above - one of methodology and one of focus. Methodologically, we shift the focus from quantifying or analyzing what faculty *do* to an autoethnographic analysis of what faculty *experience* when they do so. Our focus shifts similarly - from an abstracted analysis of outcomes of change efforts to an analysis of the process by which change occurs. Our research questions are:

1) What pivotal events shaped the process of changing assessment in the course?
2) What themes exist in the instructor's experience of changing?

We focus on the faculty member's longitudinal experience of deciding to, changing, and continuing to revise their course assessment system. The results will be useful for both educators and researchers. We expect educators to gain language to articulate and make sense of the experience of change as a process as opposed to event. We expect that researchers will find value comparing and contrasting our results with the experience of others who live with the course changes that comprise curricular change efforts.

[1] We acknowledge this is a truism - but believe it to be a useful one to make visible for this paper.

Methodology

We used a qualitative approach based on collaborative autoethnographic methods [14], [15]. That method is appropriate for this study's focus on the experience of one faculty member (the lead author) in one course (a mechanical engineering dynamics of rigid bodies course). Data comes from interviews and course artifacts. Because of the researchers' roles (faculty) and the nature of the research methodology (i.e., the sole participant was also a researcher), our institutional IRB determined that no review or approval was necessary.

Relevant context

The study is grounded in the lead author's work to improve a 2nd year mechanical engineering rigid body dynamics course that includes the following major topics; kinematics, kinetics of particles, rigid bodies in one, two, and three dimensions, Newton-Euler equations, as well as Work-energy and impulse-momentum principles. The primary textbook is a custom edition of *Engineering Mechanics, an Introduction to Dynamics [16]*. A syllabus prepared for ABET accreditation purposes is hosted on the department website [17]. The course is a part of the required ME curriculum and a prerequisite for multiple later courses. The existing course used a flipped classroom design wherein students watch video lectures before coming to class and then use class time to complete problem worksheets.

The motivation for this project was the lead author's emergent dissatisfaction with grading in their course - with the experience of that emergence described in the results section. His dissatisfaction drove a process of changing assessments in the course. The second author served as a key resource for educational theory and best practices throughout the process. The portions of the class relevant to this paper are not the course content but rather course assignments and grading. Prior to assessment changes, assessment in class included in-class worksheets, review problems (similar to quizzes), and tests. The changes in assessment that are the focus of this paper are primarily grounded in the concept of *specifications grading [18]*. While a thorough review of specifications grading is beyond the scope of this paper, prior work in engineering education provides an overview as well as examples in a variety of courses[19]. Briefly, specifications grading is based on three key components:

1. Transparent and explicit assessment criteria (specifications) to evaluate assignments
2. Pass/fail grading of assignments instead of points using the specifications
3. The ability for students to resubmit work that does not initially meet the specifications

Some authors identify the alignment of specifications with course learning outcomes as a further component of specifications grading. For purposes of this paper we treat such alignment as a general principle of good assessment. Our reasoning for doing so is that alignment is identified as part of many proposed improvements for assessment and, as such, it is not differentiating as a unique indicator of specifications grading [20], [21].

Data collection

An interview serves as the primary data source for this study. The interview occurred in the first author's office and lasted approximately 1.5hrs. The interview was loosely structured based on the second authors' general knowledge of the history of the first author's course assessment development work and the process and timeline of changes in the course. The second author prepared a list of topics in advance but did not prepare a list of specific questions. The topics were also guided by the second author's experience with mentoring other faculty on course development. The interview was paired with two aspects of secondary data collection appropriate for our methodology. First, artifacts from the process of the course development. These included course syllabi, assignments, and grading rubrics. Second, was the first author's ongoing reflection and further discussion of his experiences during the analysis process (i.e., member checking and structured reflection).

Facilitated autoethnography

Autoethnographic methods, generally, focus on analyzing a researcher's own stories of their personal experiences [22]. Analyzing such experiences aids understanding of and connection to the broader world

(e.g., cultural, social) in which those stories exist. Doing so enables researchers to critically examine the dominant narratives and themes in which those experiences otherwise exist without question or analysis [15]. Autoethnographic research is useful for bringing deep and complex insights into fields - but relies on the capacity for reflexivity and connection of their reflection on experience to a field's existing knowledge base [22]. It is not, nor is it meant to be immediately generalizable - acknowledging and leveraging its grounding in one's own experience to garner unique insights.

Facilitated, sometimes called collaborative, autoethnography is a version of autoethnography in which the process is supported by someone with either/both expertise in ethnographic research or research in the field in question [14], [15]. In facilitated autoethnography, the outside researchers (usually called facilitators) and participant(s) (participant-researchers) interact to aid in the elicitation of and sensemaking. Facilitated autoethnography has previously proven useful in similar studies - i.e., understanding K-12 teacher's experience partnerships meant to change school practices to be more research based [14]. There, having university faculty with expertise on the pedagogical methods being introduced facilitated the autoethnographic process for K-12 teachers introducing them provided unique and clear insights into how curricular and classroom changes happened. That relationship specifically aided in naming, labeling, and describing teacher's experiences in ways that were useful to the field as well as empowering to the participant-researchers. The value of the facilitated process closely mirrors our study goals. However, we note that this method requires a strong relationship (and awareness of the risk of bias in that relationship) of trust and clear boundaries between the facilitator and participant-researcher.

Our perspective and positionality in this research are intertwined with how and why we choose this project. Our interest is not critiquing the conclusions of existing research. Instead, we are critical of the way in which existing approaches to curricular change research fail to consider aspects of faculty members' change experience that are useful to sustaining and grounding improved instruction. The first author (he/him) holds a PhD in mechanical engineering (focus: non destructive testing) and has been a non tenure track faculty member for approximately 7 years. His primary workload is divided between teaching and management of a scholarship program in the college of engineering. He has no formal training in engineering education. He has actively pursued opportunities to engage in learning about engineering education and teaching best practices throughout his time as a faculty member. The second author (he/him) holds a PhD in engineering education and has been a non tenure track faculty member for approximately 6 years. His role is divided between teaching responsibilities, engineering education research, and supporting other engineering faculty in improving their teaching and courses. Both authors see themselves as educational practitioners first and this research is informed by our observations of the gap between implemented and intended change to curricula and practice.

Results

The results are divided into two sections that mirror our two research questions. The first section describes pivotal events during the process of changing course assessment. The second describes themes of experience throughout the process. From here on, we refer to the first author and participant-researcher as the *instructor* and the second author as the *facilitator.* Rather than using headings to divide events and themes, we instead use a narrative reporting style and use **bold** to highlight key events and themes. All quotes are from the instructor unless noted

Critical events and insights

The initiating event for the entire process of changing the course was a **handoff** of the course from a previous faculty member, who was nearing retirement, to the instructor. As a newly minted faculty member, the instructor was given responsibility for teaching the course as well as established materials to do so. The nature of the course and his relationship with the previous faculty member defined his experience of this event. The previous faculty was someone that the instructor respected (Facilitator: "How did you feel about [faculty]? Instructor: "Reverence") and looked up to, and who had put significant effort into transitioning the course from a traditional lecture to flipped methodology. The result, to the instructor, was that the handoff came with the gravitas of a respected faculty member's

knowledge and a well established set of materials - i.e., a sense of **inertia**. That inertia both provided a stable starting point to learn to teach the course. During analysis, he termed the early years of teaching the course as "survival" which was supported by the inertia, which was enabled by his trust and respect for the previous faculty member. However, the inertia also affected later decisions about what could be changed.

The events that **sparked change** in course assessment were, indirectly, driven by educational changes during the pandemic. However, the pandemic itself did not lead to the changes, nor where the changes initially motivated by dissatisfaction with the existing course assessments. Instead, the instructor noted the increasing usage of flipped classroom techniques by others during the pandemic caused a moment of self reflection: "*That was good, until essentially, the pandemic happened, and everybody flipped their classrooms. And it caused kind of this cascade of thought about whether or not a flipped classroom was effective. And ultimately, through several conversations with you, actually, the idea of the fidelity of implementation really stuck in my head. And what I was seeing was that a lot of people were flipping classrooms kind of just haphazardly recording the lecture they would otherwise give, and then having students watch that, and then also doing stuff in the classroom, essentially doubling the classroom time. So it kind of started me on a spiral of is anything that we're doing done correctly. … . Watching people do it and give it a bad rap because they were doing it. So clumsily, really bothered me. So that was the first thing that kind of spurred this whole conversation about assessment was, are we doing kind of the things that we think we're doing well, actually, well."* We see two aspects of how the instructor experienced this event as notable. First, was the power of relative comparison - the instructor did not see any flaws in his course or assessment directly. Rather, seeing others do something he did in a way that he saw as poor was the root of his course re-evaluation and change. Paired with that is the **language** to describe the problem. The use of the phrase fidelity of implementation (c.f., [7]) was not organic, it came from a general discussion about educational change the instructor and facilitator had had previously. The instructor noted, in the interview and during analysis, that gaining labels, terms, and concepts from educational research was a powerful tool in making sense of the experiences that motivated, and guided, changes in his course.

The result was a **crisis of faith** for the instructor that led to a process of observation and decisions about how - in both abstract and concrete ways - assessment in his course needed to change. He noted course assignment was" *pretty traditional, in that sense to homeworks. And it was your traditional here's a system there's some forces on it drawn like what is the resultant force at point In a, and the idea here is that you would be able to create a coherent document that demonstrates the complete problem solving spectrum. So that's identify given information, write out more or less mathematically what you're trying to find. Come up with the plan or and like, draw your freebody diagram. And dynamics is the most common first step for a lot of problems. But really, it's figure out the correct suite of tools to apply to the problem, apply those and then justify that answer. That's the intent.*

While there were baseline characteristics of good assessment practices such as rubrics and scaffolding, he saw that students' approach to learning did not use those elements as he (and other faculty) perceived them. He described this as "*just diving right into the problem solving and kind of missing all the scaffolding work and just tripping their way through a lot of the assessment.* During our analysis process, the instructor further reflected that there was a gap between learning outcomes that he thinks resulted from a **gap between the way faculty and students thought about the course content/learning** - *"The rubric didn't feel relevant to the way students were approaching the problems - and disrupted rather than encouraged students to use other tools in the class (e.g., scaffolding and structures in the problem. Rubric elements didn't align with goals of learning in the lesson (or the overarching lessons in the class). It focused on check box steps (side box steps) but not the why they needed to do this or the plan - conceptual was missing and was presumed, through mathematical signaling, to show the student why they were incorrect."* While the assessments were aligned with the written learning goals of the class, the gap remained between how he understood those outcomes and how students did. In assessment in particular,

he noted multiple points of tension with his underlying educational goals goals, including everything from the message communicated by assignment choices to the method of assigning grades itself.

The result, as the, in his words, a feeling of being **completely overwhelmed** and "general panic, and internalized fault and blame" The facilitator noted that his description of events to this point seemed to reflect a language, goal, and pedagogical approaches characteristic of a shift towards student-centered learning, but his articulation of the instructors' underlying thinking suggested a need for control and making students' learning visible to him that was more characteristic of an instructor-centered mindset. The instructor reflected that at that point he consciously wondered whether he had the knowledge to understand and fix their experience. Both in having this realization and changing assessment in the course, he described **using his feelings as a proxy** for his students' feelings about the course because they were accessible. Specifically, he saw/sees his feelings as a synthesis of conversations and observations about student experience, but expressed through his sense of responsibility for the course.

At that point, the instructor had **made the decision to change** assessment and **sought input** from sources of engineering education knowledge he had - including the facilitator. Those sources provided a basis for *"my, like, for lack of better phrase crusade against numerical accuracy…I took a seminar reading seminar style class with [blinded], and a couple other people. … I was exposed some literature that helped me to understand even just this concept of a threshold concept, and how important that wasn't."* However, he noted that he had previously had trouble **translating those broad concepts of learning into specific implementations** in the classroom, that this had been a gap in course handoff, and that he thought such **knowledge should be inherent** in engineering faculty (*"They just gave me their exam questions. Nobody gave me rubrics to their homeworks, they would just give me what they asked in the homework"*), and was resistant to asking for help ("*But ultimately, I was also embarrassed to ask because I was given the keys to this class. And I was like, I should be able to figure this out."*). Specifications grading entered into this change when the instructor was given a book on specifications grading by the facilitator at the end of a discussion about assessment in his own course. Thus began the instructor's effort to use specifications grading in his dynamics course - which he sees as the key change in course assessment..

The **basic change occurred quickly but adjustments are still ongoing after four semesters**. The instructor was hopeful, and happy with his initial change to specifications grading. However, the initial change to specifications grading resulted in **untenable workloads** and frustration for both students and the instructional team. Partially that was a result of implementation decisions (e.g., unlimited resubmission - which is warned against in specifications grading literature [23]. Partially, the challenge of a change that *"touched every single work product [was] overwhelming because" I didn't have time to plan the work products together [and] was changing things on the fly"* did so as well. That was because the instructor had not realized how a seemingly simple change cascaded to adapt his class to an entirely new system. The **change made more work and changing made more work** and *"it actually felt out of control, literally out of control, I was hours ahead of the students ahead."* Happily, the instructor was able to separate the positive aspects of the change from how he experienced the process of that change. Over the following semesters, the implementation of specifications grading has continued to improve and change and he sees the course as reaching a point of "metastability - *I'm proud of this. I acknowledge it works but it's not ever fully right"*. He specifically invoked the ongoing process of changing the course as *"your almost Ship of Theseus"* and noted during analysis "*if I'm going to ask students to revise their work I need to be willing to revise my work."* Having reached a metastable process of ongoing change, i.e., repeating events, we switch to talking about key themes in the experience he had in changing his course.

Themes

We identified 4 major themes that occurred throughout the instructors' experience of changing the course. They range in the parts of the experience of change they focus on from how it is understood to how it is implemented to how it is evaluated. Observationally, most of the themes could be identified in the experience surrounding all of the critical events - but because we are at a first stage of reporting and

defining these themes we choose not to focus on their occurrence. The naming and description of the themes was a key part of the analysis process and involved both authors/researchers.

The first theme we call - **the language to see**. The instructor repeatedly referenced how having language and concepts to identify problems was as useful as it was in addressing them. He noted repeatedly that having exposure to educational concepts was not the *basis* for his observations about his course, but often it was the basis of making sense of those observations. *"And they said, You should read this book, because it has some, it has some good background, which is something that I distinctly lack. I could not write you a lit review for shit. Write a review on on the pedagogy behind or like the learning psychology or like the fundamental science behind these educational practices. Okay, I that is a that's a glaring black hole in my knowledge. ...And at least it gets me familiar with like, Perry's development of a graduate student or whatever of like a higher ed student, right? Or it gives me like these these core concepts that other people just take us take for granted. I'm some fucking acoustics guy who's like trying to do his best over here."* Gaining language and using that language of education to describe his observations was as empowering as the gaining of knowledge.

However, the flipside was that **language and understanding were assumed to be interchangeable**. The instructor invoked multiple concepts to describe what he did or intended to do - "T*hey were graded for correctness. Now you have this and you label it competency based. You also label it specifications based…I mean, it's actually specifications based the way I run it, but I think specifications kind of falls under the competency umbrella."* If compared to the basic literature on these concepts, scholars would note a spectrum of correctness. Some concepts (e.g., specifications grading) mostly aligned with definitions in primary sources, but others (e.g., competency based grading) seemed to be based on synonyms (e.g., his goal of grading students' broader competency was labeled competency based grading). Additionally, the **boundaries of key labels became flexible to accommodate other places within an understanding of education where the instructor's language was less developed**. This was especially common where abstract or generalized concepts of assessment (e.g., alignment [20]) became parts of rather than precursors for defined pedagogical techniques:

I, through the ample use of graders will have the students upload that work to grade scope, they will have that pass through essentially a rubric, which is just these core elements of the problem solving. And these are been as either conceptual mistakes or numerical mistakes. And if a student has more than one conceptual mistake, and like, if there's a couple of math mistakes, I'm not going to kick it back to them. But if it's like an embarrassing number of math mistakes or like something really bad, that will also get a kickback to them. Essentially, they know what this what this threshold is, and why What that means is either it passes, which is this is that which to me means this is a demonstration of competent work, which would not be embarrassing in a design review. Or it does not pass, at which point it goes back to the student and they can revise it. And we're talking about homework tests.

This description contains all three core characteristics of specifications grading that we listed above. But we see it as notable that the change in *how* to grade was concurrent with a broader change about what the rubric he used to grade contained. That focus of grading was something that had long bothered the instructor , and it seems as if the change to the fundamental system of assigning grades was a necessary

The second theme was **the Ship of Theseus -** a phrase the instructor used explicitly**.** The instructor grounded any effort at change as existing within the confines of an already extant course. The result was that any change was fundamentally **Interventions not designs**. The impact of this, both as the instructor experienced it and the facilitator saw it, cannot be overstated. "*But I also want to point out that this is the, my course is the quintessence quintessential Ship of Theseus. Like, there was never a moment where I was like, this is now the new course, right? It's always like, I'm gonna put this here and change this thing, and then observe this effect. And then I'm gonna change this thing, or walk this one back and observe this effect. It's like thing at a time, not like, what"*. The positive of this was that there was always a base course that the instructor knew worked acceptably well and that the result was often good getting better or that

small changes could be made in an ongoing way. The negatives, however, were that it made change difficult - *"Anything we change has secondary impacts that must be accounted for"*. Those secondary impacts, drove how changes worked in often unanticipated ways. During analysis we discussed how an assignment fit within a set of assignments might be obvious, but how that assignment existed within a course or a culture of education was also impactful - and harder to predict. During the interview the instructor described the pros and cons of this reality - "*cons you're building, sometimes you're building a good thing on top of a bad thing, right, so you're inheriting some of the problems of stuff that maybe you haven't even observed as a problem, that can be an issue. Pros, it's feasible to change things every semester, instead of waiting for a significant block of time to do a full redesign. So I can change things between spring and fall this approach, and then do bigger [stuff], between fall and spring with this approach because I have to summer to kind of play more, play harder. was the first thing he changed. The first thing I changed was about assessment about assessment was I built better rubrics with her assessing those problems in grade scope. And I made it because it was great scope, I made that transparent to everybody what was being assessed."* Fundamentally, one key negative - and the source of this themes name, was that the course was never redesigned, and as such any instructional decision that had been made in the past could have a very long tail of impact even after being removed from the course. This course inertia was heavily intertwined with the last two themes.

Theme three was that **time is free but priceless**. In the instructor's experience, the nature of changing one thing at a time in an existing course that he was responsible for running every semester was intertwined with the time. Time was always available if one made choices about priorities, but how that time was used was part of a complicated balance. The nature of that time was two fold - the instructor's time to plan and prepare changes as well as the time in class. He noted that especially in the early days of the course, he struggled with the relationship of classroom time and course planning: *"But Course Management is completely different. Like, what, what can I fit into an hour? What should I fit into an hour? Like? What what is important to fit into the class at all versus teaching it more thoroughly incorrectly? Like, is it worth it to punt some topics in favor of other times? I had no answers to any of these questions."* This aspect of the experience, which is difficult to disentangle from evidence based pedagogies. Changes often used more course time than what had existed previously for multiple reasons (e.g., explaining a new grading system). Given that his course, as are so many engineering courses, was perceived as 'full' even a seemingly course time neutral change like grading affected the timing of the course. One result was that the instructor experienced internalized pressure to ensure that the topics that had always been covered in the course remained covered - for a variety of reasons we plan to focus on in future work.

In parallel, when focusing on creating a better course using evidence based techniques, the instructor assumed (because specifications grading is advertised as doing so [23]) that specifications grading would save him time. That perception did not materialize - for two reasons. One, being unfamiliar with such course changes, the instructor underestimated the time required to develop the new materials. Two, he mapped his existing knowledge of how long it took to undertake an instruction task (e.g., write a test) onto the new system - failing to account for the fact that he was learning a new way of creating such assignments, which took longer. The fundamental result as noted during the 'first semester of change' event was that *"it actually felt out of control, literally out of control, I was hours ahead of the students ahead."* Part of that was a realization through the change of what is unmovable within the course, curriculum, and culture of education.

The final theme was the **perception of immovable objects,** which emerged mainly in the analytic discussions after the interview as the instructor and facilitator reflected on their prior conversation**.** The instructor, despite an appeal towards a more student and evidence centered teaching, retained certain assumptions and beliefs about teaching from what he termed 'traditional engineering education'. For example - the role of the faculty member in identifying what needs to be known and structuring the presentation of that information for students. His experience and efforts related to course change built on top of these beliefs, instead of changing them. He located these beliefs within a sense (discussed during

analysis) that technical courses are different from design courses. He saw classes about design as allowing students to *"flex a little more"* in what and how they learned because they deal with normative instead of objective truth. He saw that technical and design divide as being acculturated into engineering and questioned the truth of it while also admitting that it was a belief he held. In part it was because he noted he had *"not seen any good examples of the alternative."* The instructor summarized the discussion about embedded beliefs by noting *"One of the lies of the classroom is that we aren't training our students in constructivist thought, they aren't prepared for it, so we're limited in what we can get them to do - we know better but we do it anyways."* We see that statement as interesting for two reasons. First, it exists within the context of the instructor describing layering techniques based on constructivist philosophies on his own non-constructivist beliefs about education. Second, the philosophy discontinuity he sees as problematic is also reflected in the beliefs about the nature of education he had already described. That is not unheard of. Research on engineering faculty and teachers generally has found highly positivist beliefs about knowledge that exist in tension and/or in tandem with more constructivist beliefs about education [24], [25]. This thread of conversation ended with the instructor noting that "we are presuming the ability to abstract by students when we always do it for them" and noting that these barriers felt both inherent and in need of change for him to be able to consider fundamentally redesigning his course.

Discussion and Future work

We reiterate that this research only reflects one faculty member's experience. However, we see important connections to existing research findings and potential future frameworks to use to improve implementation of research based practices in engineering. As context for linking efforts at change to faculty experience, we note that the instructor has not expanded the use of EBIPs used in the course described above to his other courses. In that context, four connections between our results and extant literature seem salient:

First, the need, benefit, and challenge of sense making at the intersection of engineering and education fundamentally shaped the instructor's experience and deserves further study. While some research has identified engineers' struggles in understanding education research [26], we know little about how engineering faculty form their understanding of education. By extension, we know even less about how communication occurs across education-focused and non-education faculty to build shared understanding. We see possibilities in the use of the theory of communicative acts and related tools to specifically study how intersubjective, normative, and other ways of understanding interact (e.g., [27], [28]. Such an understanding would provide deeper insights into the observations from our work about *how* evidence based teaching techniques can become malformed and whether and how that process can be interrupted.

Second, we see the tradition of framing faculty as change makers instead of learners as limiting to what research can tell us about why course and curricular change are so lacking [1], [2]. The instructor's experience strikes us as a quintessential example of learning by doing, and similar to the constructive learning process espoused in education. In effect, that learning is inseparable from his perspective, beliefs, knowledge, and actions that create course change. We see the change based approach as likely informed by existing power hierarchies between students and faculty in higher education, and the implications of the student label. However, the difference between the approaches largely relates to the level of understanding of educational practices necessary for high quality implementation - a known limitation[7], [25]. If what a faculty member needs to succeed in creating change is to be approached as a learner and novice, and they are instead approached as an equally informed partner in change, that difference may limit the pace and quality of impact - and be demotivating to faculty members when efforts at change become frustrating.

Third, redesign and implementation of changes requires different supports, solves different problems, and generates different outcomes. Fundamentally, we see significant evidence in the instructor's experience that fundamental redesigns can seem daunting for a variety of reasons. However, smaller efforts have their impact limited by a variety of factors. Both are fundamentally framed by what assumptions about education can be relinquished and how they are embedded into a course, faculty member, and educational

culture. Surfacing, and understanding how those assumptions can create secondary change effects, seems important to making more and better change. The instructors' experience suggests a need to evaluate what happens when surface level characteristics of a constructivist approach to portions of a course (e.g., assessment) are implemented within other non-constructivist approaches to education. What seems to happen, and should be better understood, is the creation of (the facilitator's choice of term) *interstitial epistemologies and ontologies*. That is, places where seemingly contradictory portions of the way engineers think about knowledge and the world merge, without reflection or question, with the way educators do. Many evidence-based techniques embed a constructivist philosophy, but are implemented in engineering education by faculty (like the instructor) who hold or partially hold positivist views, something new is being created that may exist in an unobserved and uncanny valley between either .

Finally, as an extension of the third link, we think a critical (but broad) need is to understand the fundamental assumptions built into our existing systems and models of curricular change. Primarily, a discussion about what metrics are the most representative for measuring progress on the changes we want to see. For example, while a focus on number of changes or the fidelity of implementation seems facially logical, at what level does it matter to the broader goals of fundamental change? Is a partial implementation that is contextually appropriate the same or different as one driven by a misunderstanding of a pedagogy? Is the wrong technique that a faculty member understands and implements well better or worse than a better choice of technique that a faculty is still learning to understand and implement. We see the idea of fidelity as useful, but requiring expansion - i.e., looking beyond concrete observable characteristics of a course. Otherwise, current change efforts may be limited by the adage known as 'Goodhart's law' - when a measure becomes a target, it ceases to be a good measure' [29]. Fidelity as practiced may matter to the credibility of a researchers' work to ensure they evaluate good implementations to make valid claims. However, it may matter less to how students experience an individual implementation - or to the broader goal of overarching shifts in educational culture. Conversely, a chaotic and incomplete (or complete but not well thought integrated) version of those same pedagogical techniques may defeat the purpose of showing its impact but be critical to a faculty member's ongoing experience of and reflection on building a better course and better curricula. As we dug through the layers of this faculty member's experience and the known aspects of philosophical differences between engineering and engineering education [24], [26] we both began to question how we talk about change. We agree change is important, and more change is needed. But we are less convinced after writing this paper about the extent that specific changes to a single course (or curriculum) matter. Instead, we see understanding how a faculty member's perception of a technique's purpose and assumptions, their perception of their role as a teacher, and their perception of their role as a course designer change over time to be more useful. That is, course changes may be better treated as a reflective rather than formative measure lest our two ships stay stuck perpetually passing in the night [1].

References

[1] R. M. Felder, R. G. Hadgraft, and Others, "Educational practice and educational research in engineering: partners, antagonists, or ships passing in the night," *Journal of Engineering Education*, vol. 102, no. 3, pp. 339–345, 2013.

[2] M. Borrego and C. Henderson, "Increasing the use of evidence-based teaching in STEM higher education: A comparison of eight change strategies," *J. Eng. Educ.*, vol. 103, no. 2, pp. 220–252, Apr. 2014, doi: 10.1002/jee.20040.

[3] C. Reidsema, R. Hadgraft, I. Cameron, and R. King, "Change Strategies for Educational Transformation," *Australasian Journal of Engineering Education*, vol. 19, no. 2, pp. 101–108, Jan. 2013, doi: 10.7158/22054952.2013.11464083.

[4] A. C. Phillips *et al.*, "Development and validation of the guideline for reporting evidence-based practice educational interventions and teaching (GREET)," *BMC Med. Educ.*, vol. 16, no. 1, p. 237, Sep. 2016, doi: 10.1186/s12909-016-0759-1.

[5] R. A. Streveler, S. Brown, G. L. Herman, and D. Montfort, "Conceptual change and misconceptions in engineering education: Curriculum, measurement, and theory-focused approaches," in *Cambridge handbook of engineering education research*, Cambridge University Press, 2015, pp. 83–102.

[6] R. A. Streveler, M. Borrego, and K. A. Smith, "9: Moving from the scholarship of teaching and learning to educational research: An example from engineering," *To Improve Acad.*, vol. 25, no. 1, pp. 139–149, Jun. 2007, doi: 10.1002/j.2334-4822.2007.tb00479.x.

[7] M. Borrego, S. Cutler, M. Prince, C. Henderson, and J. E. Froyd, "Fidelity of implementation of research-based instructional strategies (RBIS) in engineering science courses," *J. Eng. Educ.*, vol. 102, no. 3, pp. 394–425, Jul. 2013, doi: 10.1002/jee.20020.

[8] J. L. Bernstein, "Unifying SoTL methodology: Internal and external validity," *TLI*, vol. 6, no. 2, pp. 115–126, Sep. 2018, doi: 10.20343/teachlearninqu.6.2.9.

[9] N. Haigh and A. J. Withell, "The Place of Research Paradigms in SoTL Practice: An Inquiry," *TLI*, vol. 8, no. 2, pp. 17–31, Oct. 2020, doi: 10.20343/teachlearninqu.8.2.3.

[10] J. Brennan, T. Fernandez, and J. Tranquillo, "Professionalization and the forgotten system: Observed practices and perceptions at the intersection of informal and formal faculty development," *To Improve Acad.*, vol. 41, no. 2, Nov. 2022, doi: 10.3998/tia.440.

[11] L. S. Robins, C. B. White, and J. C. Fantone, "The Difficulty of Sustaining Curricular Reforms: A Study of 'Drift' at One School," *Acad. Med.*, vol. 75, no. 8, p. 801, Aug. 2000,

[12] J. R. Century and A. J. Levy, "Sustaining change: A study of nine school districts with enduring programs," in *A paper presented at the Annual Meeting of the American Educational Research Association*, sustainability2003.terc.edu, 2002.

[13] M. Priestley, "Making the Most of the Curriculum Review: Some Reflections on Supporting and Sustaining Change in Schools," *Scott. Edu. Rev.*, vol. 37, no. 1, pp. 29–38, Mar. 2005, doi: 10.1163/27730840-03701004.

[14] A. Malorni, A. Diaz, M. S. Spencer, and T. Jones, "Collaborative autoethnography as a Tool for Research–Practice partnerships: Facilitating Self and School Transformation," *Qualitative Social Work*, vol. 22, no. 4, pp. 643–662, Jul. 2023, doi: 10.1177/14733250221088211.

[15] N. Bissett, S. Saunders, and C. Bouten Pinto, "Collaborative Autoethnography: Enhancing Reflexive Communication Processes," in *Ethnographic Research and Analysis: Anxiety, Identity and Self*, T. Vine, J. Clark, S. Richards, and D. Weir, Eds., London: Palgrave Macmillan UK, 2018, pp. 253–272. doi: 10.1057/978-1-137-58555-4_14.

[16] D. J. McGill and W. W. King, *Engineering Mechanics : An Introduction to Dynamics*, 4th ed. Tichenor Publishing, 2003. Available: https://www.amazon.com/Engineering-Mechanics-Introduction-David-McGill/dp/0742134938

[17] G. T. Me, "Course Syllabus - ME2202 Dynamics of rigid bodies," 2022.

[18] L. B. Nilson, *Specifications Grading*, 1st ed. Routledge, 2014.

[19] T. M. Fernandez, K. M. Martin, and R. T. Mangum, "Whose grade is it anyway?: Transitioning engineering courses to an evidence-based specifications grading system," *2020 ASEE Virtual*, 2020,

[20] J. Bransford, A. L. Brown, and R. Cocking, "How people learn: Brain, mind, experience, and school," 1999,

[21] R. A. Streveler, R. L. Miller, A. I. Santiago-Román, M. A. Nelson, M. R. Geist, and B. M. Olds, "Rigorous methodology for concept inventory development: Using the' Assessment Triangle' to develop and test the Thermal and Transport Science Concept Inventory (TTCI)," *The International journal of engineering education*, vol. 27, no. 5, pp. 968–984, 2011,

[22] T. E. Adams, S. L. H. Jones, and C. Ellis, *Autoethnography*. Oxford University Press, 2015.
[23] L. B. Nilson, “Yes, Virginia, There’s a Better Way to Grade,” Jan. 18, 2016. Available: https://www.insidehighered.com/views/2016/01/19/new-ways-grade-more-effectively-essay. [Accessed: Feb. 08, 2024]
[24] D. Montfort, S. Brown, and D. Shinew, “The personal epistemologies of civil engineering faculty,” *J. Eng. Educ.*, vol. 103, no. 3, pp. 388–416, Jul. 2014, doi: 10.1002/jee.20050.
[25] N.-H. Kang, “Learning to teach science: Personal epistemologies, teaching goals, and practices of teaching,” *Teaching and Teacher Education*, vol. 24, no. 2, pp. 478–498, Feb. 2008, doi: 10.1016/j.tate.2007.01.002.
[26] M. Borrego, “Conceptual difficulties experienced by trained engineers learning educational research methods,” *J. Eng. Educ.*, vol. 96, no. 2, pp. 91–102, Apr. 2007, doi: 10.1002/j.2168-9830.2007.tb00920.x.
[27] C. M. Gray and T. M. Fernandez, “When world (view) s collide: Contested epistemologies and ontologies in transdisciplinary education,” *The International journal of engineering education*, vol. 34, no. 2, pp. 574–589, 2018,
[28] C. M. Gray, A. L. Toombs, and C. McKay, “Meaning Reconstruction as an Approach to Analyze Critical Dimensions of HCI Research,” in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, in CHI EA ’16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 328–340. doi: 10.1145/2851581.2892571.
[29] C. A. E. Goodhart, “Problems of Monetary Management: The UK Experience,” in *Monetary Theory and Practice: The UK Experience*, C. A. E. Goodhart, Ed., London: Macmillan Education UK, 1984, pp. 91–121. doi: 10.1007/978-1-349-17295-5_4.