Is STEM too Hard? Using Biglan to Understand Academic Rigor and Teaching Practices across Disciplines

Jessica Ostrow Michel
Corbin M Campbell
Karen Dilsizian
Teachers College, Columbia University

Abstract: *Absent from the higher education literature is the notion that disciplinary teaching and learning is situated in classrooms. Therefore, this study explored discipline categorized by classes (as per Biglan (1973) via quantitative observation in 459 courses across nine colleges and universities to understand whether academic rigor and cognitively responsive teaching vary by discipline. Findings revealed that soft discipline classes scored higher than hard discipline classes on all five teaching practices under investigation. Our findings support the use of theoretically derived heuristics rather than organizational structures for defining disciplines. Additionally, given that courses rooted in hard disciplines may lend less naturally to some of the teaching practices deemed important for student learning, we see STEM teaching and learning as an area in need of continued attention.*

Keywords: *stem teaching, classroom observation, quantitative observation, cognitive responsiveness, college teaching, academic rigor, discipline*

Jessica Ostrow Michel is an adjunct instructor in the Higher and Postsecondary Education program at Teachers College, Columbia University and a Postdoctoral Research Fellow at the Sustainability and Education Policy Network at the University of Saskatchewan. **Corbin M. Campbell** is Associate Professor of Higher Education at Teachers College, Columbia University. **Karen Dilsizian** oversees the College Readiness Program at a non-profit organization in Jersey City, NJ.

Introduction

The academic profession has been famously (or perhaps infamously) characterized as “small worlds, different worlds” due to the considerable differentiation in experiences in academe, particularly across disciplines (Becher, 1987, 1994; Braxton & Hargens, 1996; Clark, 1987). The vast differences in experiences in higher education institutions are not limited to faculty. Substantive scholarship on the experiences of administrators and students also highlight the importance of considering discipline in regard to important topics in higher education such as administrative leadership (Del Favero, 2005, 2006), student learning (Laird, Shoup, Kuh, & Schwarz, 2008; Pike & Killian, 2001), student and faculty research (Finnegan & Gamson, 1996; Craney, Mckay, Mazzeo, Morris, Prigodich, & de Groot, 2011), faculty turnover (Xu, 2008), and faculty beliefs and faculty teaching practices (Barnes, Bull, Campbell, & Perry, 2001; Laird et al., 2008).

The robust body of literature that includes discipline as context for understanding faculty roles illustrates differences between hard disciplines (or disciplines with single paradigms) versus soft disciplines (or disciplines with multiplistic paradigms). Hard disciplines are traditionally associated with Science, Technology, Engineering, or Mathematics (STEM). These disciplines have been a particular focus in higher education due to the demand for more students to graduate with STEM degrees in order for the United States to remain competitive internationally (Breiner, Harkness, Johnson, & Koehler, 2012; Mullis, Martin, Gonzalez, & Chrostowski, 2004). For example, as to soft versus hard discipline influencing faculty grading beliefs about gatekeeping, Barnes et al. (2001) found an influence of discipline on faculty conceptualization of teaching. Specifically, hard fields held higher gatekeeping perceptions of grading than soft fields. Furthermore, in terms of teaching, faculty in the soft fields devote most of their teaching time to lectures and seminars, while faculty members in hard fields spend most of their teaching time on laboratory instruction (Ballantyne, Bain, & Packer, 1999; Milem, Umbach, & Liang, 2004).

Overall, we found that the consensus from the literature (that includes discipline as a context for understanding teaching practices) showed that soft disciplinary courses hold promise for better teaching practices than their hard discipline counterparts. For instance, using

Biglan’s (1973) framework, Laird et al. (2008) found that faculty in soft, pure, and life disciplines scored higher in their self-reported practices that facilitate deep approaches to learning for students (although pure had a very small effect size). Furthermore, Braxton (1995) explored faculty members’ self-reported data and found that faculty in the soft disciplines integrate good principles of teaching to a greater degree than those in hard disciplines. Braxton, Olsen, and Simmons’ (1998) explanation for this difference pointed to a disproportionate level of pressure for faculty members across the two paradigms, as faculty in the hard sciences have more pressure to conduct research and are therefore less likely to adopt effective teaching practices. More recently, Kilgo, Culver, Young, and Paulsen (2017) replicated Braxton et al.’s (1998) study on student perceptions controlling for disciplinary differences using paradigmatic scores: soft fields versus hard fields. Kilgo et al.’s (2017) results replicated the findings from Braxton et al.’s (1998) using student-level, instead of faculty-level reports, in order to explore faculty use of good teaching practices. Further, they found that students taking a higher percentage of courses in the hard sciences perceived good practices of prompt feedback and high expectations/academic challenge, and an increase in cooperative learning.

Other research explores the intersection of discipline and course contexts, such as class size and mode (online versus on-site). For example, Johnson (2010) studied class size and student performance at a large public research institution using Biglan's (1973) classification of academic disciplines: hard versus soft, applied versus pure, and life versus non-life. Johnson (2010) found that class size did have a significant effect on students’ final grades across all disciplinary areas. However, while Johnson (2010) predicted that class size had a greater effect on final grades within the soft disciplines versus hard disciplines and the applied versus the pure, the results did not support this hypothesis. Rather, evidence showed that class size had a greater impact in the pure-soft-life disciplines than in applied-soft-life disciplines and in pure-hard-life disciplines than in applied-hard-life disciplines. The effect of class size in life versus non-life disciplines, however, did not produce any significant differences (Johnson, 2010). With regards to mode, the literature addressed student achievement and satisfaction as two ways to evaluate the quality of online education. While some scholarship on academic achievement have shown mixed reviews (Jung & Rha, 2000), some scholars have asserted that online education can be at least as effective as traditional classroom instruction (Allen & Seaman, 2004).

Although research has traditionally focused on policy stream at the institutional level, recent research has focused on situating college teaching within subject matter and disciplinary contexts (Lattuca & Stark, 2011; Neumann, 2014). Furthermore, countless studies in higher education that do not have disciplinary differences as the focus of the research control for discipline to isolate other aspects of the college educational experience, given the importance of discipline in the experiences of both students and faculty (Becher, 1987; Pascarella & Terenzini, 2005). To assist in discerning the vast literature on disciplinary experiences in higher education, we describe two important characteristics of disciplinary research that set the stage for the present study: the way discipline is conceived and the level of analysis.

Defining Disciplinarity

Defining *discipline* is a complex task. According to Finnegan and Gamson (1996), disciplines are “demarcated knowledge domains with distinctive epistemologies and methods. They are also cultures that are embodied in the social relations among members” (p. 152). Some scholars, such as Pike and Killian (2001), have suggested that prior research on the educational experiences of undergraduate students by discipline largely used organizational definitions of the concept rather than theoretically conceived discernment of disciplines. For example, Finnegan and Gamson (1996) studied four mathematics departments and four English departments to understand how faculty made meaning of an organizational shift from a focus on teaching to research. Craney et al. (2011) used student majors (organizationally structured) and then grouped the majors according to three broad categories (math/science, social science, and humanities) to study the outcomes of a summer student research program. Del Favero (2006) used unit affiliation to define 10 different disciplines for studying the cognitive complexity of university administrators. Through these organizational definitions of disciplines, researchers allow majors, departments, and divisions/schools within universities to discern and differentiate disciplines. However, these organizational ways of conceiving discipline were limited in that they relied on the assumption that disciplinary differences are structural and cultural rather than substantive, paradigmatic, or epistemological.

By contrast, other research on disciplinary differences in educational experiences in higher education considered theoretically driven heuristics for defining disciplinary categories. Perhaps the most widely used categorization, as mentioned earlier, is Biglan's (1973), which describes disciplines as hard versus soft (disciplines with single paradigms versus multiplistic paradigms); pure versus applied (knowledge for discovery versus applied knowledge); and life versus non-life (concerning life systems). Several studies of both students and faculty have used Biglan's (1973) work to categorize disciplines, finding, overall, that the hard/soft and pure/applied distinctions are more meaningful than life/non-life in understanding the differences in educational experiences in higher education (Barnes et al., 2001; Del Favero, 2005, 2006; Jung, 2012; Laird et al., 2008; Paulsen & Wells, 1998; Xu, 2008). For example, Laird et al. (2008) studied the disciplinary effect on deep approaches to learning, as well as on the relationships between deep approaches to learning and selected educational outcomes, and found that deep approaches to learning were more prevalent in Biglan's soft, pure, and life fields compared with their hard, applied, and non-life counterparts. Several scholars (e.g., Braxton, 1995; Smart & Ethington, 1995) found that student assessments in hard fields required memorization and application of course material, while assessment in soft disciplines tended toward exam questions requiring analysis and synthesis of course content. In terms of faculty preferences, instructors in soft disciplines tend to indicate a greater preference for teaching while those in hard areas show significant preference for research (Biglan, 1973; Clark, 1987). In contrast to organizational definitions of disciplines that are determined based on college structures, Biglan's (1973) theory distinguishes disciplinary categories through heuristics that are associated with individual or environmental understandings, worldviews, traits, or subject matter substance.

Level of Analysis

Beyond defining and conceptualizing the disciplines, researchers must also decide who makes up the disciplines and at what level disciplines should be measured. For example, disciplines could be determined by categorizing faculty into departments/fields (e.g., Barnes et al., 2001; Finnegan & Gamson, 1996; Hativa & Marincovich, 1995; Laird et al., 2008) or students into majors (Craney et al., 2011; Laird et al., 2008; Paulsen & Wells, 1998; Pike & Killian, 2001). Sometimes, discipline is categorized by asking faculty to self-affiliate with a specialization, and categorize the specializations into disciplines (Del Favero, 2005, 2006; Lindblom-Ylanne, Trigwell, Nevgi, & Ashwin, 2006; Xu, 2008). Upon further investigation, these discipline measures are determined at either the individual (student or faculty) or department level. Yet, disciplinary teaching and learning is also contextualized in courses and classrooms in addition to departments, fields, and majors.

All but absent from the literature on disciplinary differences in undergraduate education is the course level. In fact, we found no studies that demarcated the discipline at the course level (i.e., categorizing individual courses into disciplines), and very few studies asked faculty to think about courses or teaching responsibilities when categorizing faculty into disciplines. For example, Laird et al. (2008) reported faculty disciplines with regards to "the field in which they teach" (p. 476). Considering disciplinary differences at individual or departmental levels may be particularly problematic due to the variation in teaching and learning practices across course contexts. For instance, in a study of UK and Finnish faculty, Lindblom-Ylanne, Trigwell, Nevgi, and Ashwin (2006) found that the same faculty member teaching two different courses often has differing approaches to teaching depending on the course context. While course context matters, the methods of data collection on college educational experiences for students and faculty are mainly self-report survey (which may lend to a student or faculty level analysis) or institutional records (which may lend to a departmental or major analysis). In order to unpack disciplinary differences at the courselevel, studies will need to consider alternative methods that yield data on the teaching and learning process as it unfolds in the classroom.

It is at the crossroad of conceptualization and method that this paper makes its contribution to the literature on disciplinary differences in undergraduate education. Given the fact that prominent scholars in the field, like Clark (1983) and Becher (1994), refer to disciplines as the "life-blood" of higher education, since they provide the main organizing base, and the main social framework for higher education as an entity, we focus on a context for understanding disciplinary differences that may yield insights for the policy discussions of student learning and accountability in higher education. Further, although many studies of academic work examine differences in teaching based on different academic levels, institutional discipline, while often a component of such studies, is rarely the major focus (Neumann, 2001). As such, the purpose of this paper is to use Biglan's (1973) categorizations of discipline at the course-level (via quantitative observation) to understand whether academic rigor and teaching practices vary by discipline. Therefore, this study pursues the following research questions: (1.) To what extent does discipline (hard or soft, applied or pure) influence the level of academic rigor in course practices? And (2.) To what extent does discipline (hard or soft, applied or pure) influence cognitively responsive teaching practices?

Conceptual Framework

As aforementioned, we conceptualized our study exploring the extent to which discipline influences teaching practices through two frameworks, namely, academic rigor and cognitively responsive teaching.

Academic Rigor

Our first framework is *academic rigor*, which we defined as "teaching practices and coursework that challenge learners to sustain a deep connection to the subject matter and to think in increasingly complex ways about the course content and its applications" (Campbell & Dortch, 2018). As such, we study rigor in terms of the cognitive challenge of the course practices: the level of cognitive complexity posed by the course and the level of standards and expectations set for sustained attention in the course material. In particular, we explored two facets of rigor: cognitive complexity as well as standards and expectations.

Cognitive complexity, via Bloom et al.'s (1956) revised taxonomy, designates six levels of increasing cognitive complexity across a curriculum: remember, understand, apply, analyze, evaluate, and create (Anderson & Krathwohl, 2001). According to Bloom et al.'s (1956) revised taxonomy, some cognitive tasks require lower levels of complexity, like repeating course ideas (remembering) or summarizing main points (understanding). Other tasks require higher levels of complexity, like critiquing an argument based on a certain perspective (evaluating) or developing new hypotheses that build off earlier work (creating). We also explored standards and expectations laid out in the course (Astin, 1993), because prior research has found a strong association between standards and expectations and students' learning. For example, Kuh, Laird, and Umbach (2004) found that faculty's level of expectations for students was associated with the degree of students' engagement in the classroom. Taken together, cognitive complexity as well as standards and expectations, made up our academic rigor framework, one which allowed us to gauge how academic rigor differed by discipline. This is instrumental in the present study since prior literature highlights the notion that discipline often influences teachers' beliefs on academic rigor.

Cognitively Responsive Teaching

Teaching practices using Neumann's (2014) framework on cognitively responsive teaching stipulates three claims that together illuminate what good teaching in higher education comprises. First, Neumann (2014) suggests that good teachers provide students with the opportunity to confront and interact with a subject matter idea derived from a discipline. Here, the teacher identifies core ideas of the subject matter, determines the sequence in which they ought to be introduced, uses multiple examples and pedagogical strategies to share those ideas, and facilitates opportunities for students to situate ideas within a broader discipline or field.

Second, Neumann (2014) stipulates that good teachers connect students' learning with their prior knowledge and experiences, from both personal and cultural perspectives, because encounters with new subject matter ideas can then interact with students' prior knowledge in ways that challenge students emotionally and cognitively to accept new ideas (Bransford, Brown, & Cocking, 2000). As such, a good teacher surfaces students' prior knowledge and then probes the ways a student frames and works through the subject matter ideas. Third, Neumann (2014) suggests that good teachers support students cognitively and emotionally when the course leads them to question long-held beliefs, as they attempt to reconcile their prior knowledge with what they are learning. This claim allowed us to calibrate how implemented out by teachers in the classroom.

Neumann's (2014) cognitively responsive teaching contributed towards our exploration of the disciplinary effect on teaching practices because this framework emulates the complexity in today's higher education coursework. Cognitively responsive teaching situates the college teaching and learning process in disciplinary subject matter ideas; students' own cognitions of the subject matter; and the intersection of course content, student understanding, and faculty understanding.

Methods

The present study used data from the College Educational Quality (CEQ) study, a multi-institutional quantitiative observational study of college classrooms across nine institutions and 587 courses (Campbell, 2017). Quantitative observation uses a closed-ended and highly structured rubric with raters who have been trained to rate within the parameters of a specific study (Waxman, Tharp, & Hilberg, 2004). We conducted observational research in the setting of the social phenomenon as it naturally occurred (Alexander & Winne, 2006; Angrosino, 2012). More specifically, our observations took place in college classrooms in order to observe how educational practices unfold. Therefore, the CEQ study used rubrics that measured the coursework practices of academic rigor and cognitively responsive teaching, as prescribed by our conceptual framework. Given our interest in categorizing discipline at the course level, we coded each course according to Biglan's (1973) disciplinary categorizations (hard/soft; pure/applied; life/non-life). Prior scholarship that has explored quantitative observation suggests that this method has several benefits, such as "the ability to understand patterns in rich, theoretically driven constructs, which can be applied to understanding student experiences both within and outside of the classroom" (Campbell, 2017, p. 290).

Sample

The nine sites for our multi-institutional observational study of college classrooms, namely CEQ, were purposefully selected to include a range of types, selectivity, and resources. Of the institutions, three were public and six were private. Further, two were research universities, two were comprehensive institutions, and five were liberal arts colleges. Three were highly ranked selective, two were ranked moderately selective, and four were ranked >100 or unranked/ less or non-selective.

We stratified our sample by discipline, class size, class level, and faculty category because prior higher education literature has found that these characteristics are associated with academic rigor and teaching practices (Astin, 1993; Pascarella & Terenzini, 2005). We weighted our sample by course enrollment: a sample that considered the number of seats in each course better represented what a majority of students would experience at these institutions. We selected 350 courses from each institution, and the faculty members interested in participating had to consent to course observation; 34.3% of sampled faculty agreed to participate. We were not able to observe all of the courses due to scheduling and observer constraints, so we purposively selected courses that contributed to the representativeness of our sample.

We observed a large number of courses (50–100) per institution. We excluded data that was missing at random, leaving a final analytical sample of N=459. We present the representation of discipline, class size, and mode (whether the course was online or on-site) of observed courses compared to the population in Table 1. In running chi-square analyses, we found that our sample, in terms of the hard versus soft disciplinary paradigm, was representative of our population. However, in terms of the pure versus applied paradigm, our sample was overrepresented with pure courses. While this was not a representative sample of U.S. higher education institutions or courses, our purposeful sample offered insight into the extent to which, if at all, discipline influences teaching practices.

Table 1. Representativeness of Course Characteristics of Observed Courses (n=459) Compared with the Population of Courses

Characteristics	**Observed Courses**		**All Courses**	
	N	**%**	**N**	**%**
Class Size				
5-10	47	10.2	1239	19.8
11-25	209	45.5	3176	50.8
26-75	172	37.5	1635	26.2
76+	31	6.8	198	3.2
Mode				
On-site	439	95.6	5924	95.3
Online	20	4.4	295	4.7
Discipline				
Hard	154	33.6	1697	30.3
Soft	305	66.4	3911	69.7
Discipline				
Applied	137	29.8	2199	40.5
Pure	322	70.2	3234	59.5

Note. Classes in the sample were stratified by size and then weighted by seat to allow for larger courses to be selected at a greater rate than smaller classes.

Procedures

We followed established procedures for comprehensively training classroom observers for quantitative observation (Stallings & Mohlman, 1988). Observers were required to have an academic background in higher education teaching and learning; complete an extensive training; pass tests on their knowledge of the conceptual frameworks and the observer logistics; and pass an inter-rater reliability (IRR) certification, comparing observers' scores against a set of master ratings from the principal investigator.

After observers were comprehensively trained, they were dispersed among the nine participating sites. Each site team was led by experienced site leaders who had participated in a pilot study. Observations occurred during a one-week site visit during the mid-point of the semester. Two observers[1] rated the entirety of one class session for each course. Observers were matched to rate courses in their own disciplinary background whenever possible. The observers rated in real time—for example, observers would record the highest level of cognitive complexity in the lecture or class discussions, no matter when that took place. They entered classes without knowledge of a faculty member's rank or research and scholarly productivity.

Data Sources

Data were collected from classroom observations in accordance with quantitative observation procedures. Data were collected from classroom observations that used detailed, closed ended, specific rubrics guided by our theoretical frameworks on academic rigor and cognitively responsive teaching (Campbell, 2017; Stallings & Mohlman, 1988). Rubric design and development was assisted by both content experts (in college teaching) and methodological experts (in survey and rubric design), who tested the rubrics for content and response process validity.

Inte-Rrater Reliability

The inter-rater reliability of the observation data was calculated through a one-way, absolute, average-measure, mixed-effects, intra-class correlation (ICC) calculation, which was appropriate for the study

---

[1] In a few classes, due to scheduling conflicts, only one observer was present.

because of the ordinal nature of the observation categories. The classes were random but the coders were fixed (i.e., we are not generalizing to a larger pool of raters outside the study). We used more than one rater per class, and we had an interest in the consistency of the absolute value of the ratings (Hallgren 2012). The ICC across all items was 0.705 (which was considered good as per Cicchetti's (1994) cut-off values, which are fair: 0.4-.59; good: 0.60–0.74; excellent: 0.75). In order to ascertain that only reliable data were included in our analysis, we treated any values where the discrepancy between the two raters were greater than two response options as missing.

Scales

To create the scales, we averaged the ratings of the two observers for each item. Scale scores are means of the individual items that were scored within each construct (e.g., if a class had lecture and instructor questions but no class discussion or activities, the cognitive complexity scale score would have been calculated as the mean of the lecture and instructor scores). We provide items for each scale from the observation rubric in Table 2.[2]

The scales were validated using Confirmatory Factor Analyses (CFA). Model fit indices indicated excellent fit of a five factor inter-correlated model (RMSEA=.049, CI [.041, .057], CFI=.965, TLI=.956, SRMR=.047). The constructs were highly reliable (Coefficient-*H* ranged from .809 to .970; Hancock & Muller, 2006; Muller & Hancock, 2008). We created two academic rigor scales (cognitive complexity and standards and expectations), and three cognitively responsive teaching scales (subject matter, prior knowledge, and supporting changing views).

---

[2] For a detailed description of the observational protocol, including training, rubric tuning, and validation, contact primary author. We ran our analysis with the independent variables for life versus. non-life disciplinary classifications and did not find any significant differences in our results. Given that our sample decreased in size when including this control (because we were unable to code some classes as either life on non-life), we chose not to include it in our final analysis.

Table 2. Items from Observation Rubric Used in Scales

Conceptual Facet	Sub-Scale	Item	Response Option
Academic Rigor	*Cognitive Complexity*	*The instructor's lecture reflected what level of cognitive processing?* The questions asked by the instructor required students to…. The questions asked by students demonstrated the students' ability to… The class activities required students to… The class discussions demonstrated students' ability to…	Remember, understand, apply, analyze, evaluate, create
	Standards and Expectations	*The classroom's norms reflected what level of sustained attention in the course material?* Standards the instructor set for students being prepared for class… Standards set by other students (students reference class materials, students show they are knowledgeable about course content) for students being prepared for class… Standards the instructor set for students' class participation… Standards set by other students for students' class participation (student norms for participation and encouraging others to participate)…	Lowest, low, middle, high, highest

Cognitively Responsive Teaching	*Subject Matter*	The instructor introduced in depth one or more concepts that are central to the subject matter of the course… The instructor created multiple representations of the "core ideas"… The instructor introduced students to how ideas play out in the discipline/field…	Very ineffective: Instructor never attempted technique Ineffective: Instructor attempted the technique but the outcome was not achieved Somewhat effective: Instructor attempted the technique and the desired outcome was partially achieved Effective: Instructor attempted the technique and the desired outcome was achieved Very effective: Instructor demonstrated mastery of the technique and the desired outcome was achieved
	Prior Knowledge	The instructor surfaced students' prior knowledge about the subject matter "core ideas"… The instructor worked to understand students' prior knowledge about the subject matter "core ideas"…	
	Supporting Changing Views	The instructor helped students to realize the difference, similarities and sometimes conflict, between prior knowledge and new subject matter ideas…	
		The instructor provided a space for students to encounter dissonance between prior knowledge and new course material…. The instructor helped students to realize the difference, similarities and sometimes conflict, between prior knowledge and new subject matter ideas…	

Multivariate Analyses

We ran five Ordinary Least Squares (OLS) regressions with Stata 14 software. We ran the OLS regressions with robust standard errors to adjust the estimates due to departure of normality, which can also handle ordinal and continuous variables like the ones in this study. We also ran the OLS regressions with fixed effects because participating faculty members were sampled within institutions, and this study did not intend to examine institutional differences. Therefore, all variables of interest were situated at the individual course level (Raudenbush & Bryk, 2002). Our five dependent variables were both academic rigor scales (i.e., cognitive complexity, standards and expectations) and cognitively responsive teaching scales (i.e., subject matter, prior knowledge, supporting changing views). Our independent variables were soft versus hard discipline and applied versus pure discipline[3]. Our controls were class size and mode[4]. We determined the significance of predictors by $p < .05$.

Findings

The final data analytical sample included 459 cases. Descriptive statistics of variables employed in our study within this sample are reported in Table 3. In terms of academic rigor, the cognitive complexity high-scale mean was approaching analyze (mean=3.753, SD=.917) and the standards and expectation scale was approximately half (mean=3.211, SD=1.044). The cognitively responsive teaching scales, on average, ranged between somewhat effective to effective (subject matter scale mean=3.689, SD=.777; prior knowledge scale mean=2.904; SD=1.023; supporting changing views scale mean= 2.715, SD=1.130). Class size ranged from 5-10 students (10.2%) to 76+ students (6.8%). Most classes (95.6%) were on-site, while very few were online (4.4%). About two thirds of the courses in the sample were soft disciplines (66.4%; 33.6% were hard), and about two thirds of the courses in the sample pure disciplines (70.2% were pure; 29.8% were applied).

We tested assumptions of linearity, normality, and homoscedasticity using the following analyses: scatter plot of unstandardized residuals and predicted values, Q-Q plot, and the Kolmogorov-Smirnov test. Data were approximately linear and homoscedastic. Multicollinearity was not a concern, as evidenced by Eigen values > .01, condition indices < 50, and VIF < 2.5. Lastly, there were no outliers that had a marked influence on the regression estimates,

---

[3] We re-ran our analysis with the independent variable for life versus non-life disciplinary classifications and did not find any significant differences in our results. Given that our sample decreased in size when including this control (because we were unable to code some classes as either life or non-life), we chose not to include it in our final analysis

[4] We re-ran our analysis with a control for tenure or non-tenure track faculty status and did not find any significant differences in our results. Given that our sample decreased in size when including this control (because we were missing this data from one site), we chose not to include it in our final analysis. For more information on tenure versus non-tenure track faculty, please see Michrl et al.. (2018).

as evidenced by Studentized Deleted Residuals[5] <+-2.5 and Cook's D values less than .5. As a result, we retained all data in the analysis.

Table 3. Descriptive Statistics and Coding (N=459)

Variable	**Coding/ Frequency**	**Mean**	**SD**
Academic Rigor Cognitive Complexity Scale—High	Remember=1 Understand=2 Apply=3 Analyze=4 Evaluate=5 Create=6	3.753	.917
Academic Rigor Standards and Expectations Scale	None/almost none=1 Few=2 Half=3 Most=4 All/almost all=5	3.211	1.044
Cognitively Responsive Teaching Subject Matter Scale	Very Ineffective=1 Ineffective=2 Somewhat Effective=3 Effective=4 Very Effective=5	3.689	.777
Cognitively Responsive Teaching Prior Knowledge Scale	Very Ineffective=1 Ineffective=2 Somewhat Effective=3 Effective=4 Very Effective=5	2.904	1.023
Cognitively Responsive Supporting Changing Views Scale	Very Ineffective=1 Ineffective=2 Somewhat Effective=3 Effective=4 Very Effective=5	2.715	1.130
Class size	5-10 students: 10.2% 11-25 students: 45.5% 26-75 students 37.5% 76+ students: 6.8%		
Mode	On-Site: 95.6% Online: 4.4%		
Discipline: Soft versus Hard	Soft: 66.4% Hard: 33.6%		
Discipline: Applied versus Pure	Applied: 29.8%		

Our research questions explored the extent to which discipline (hard versus soft, applied versus pure) influenced academic rigor course practices, and the extent to which discipline (hard versus soft, applied versus pure) influenced cognitively responsive teaching practices. Table 4 provides standardized coefficients of predictor and discipline control variables on teaching practices.

Distinctive patterns were revealed in our findings. The regression models demonstrated that soft disciplines scored significantly higher in all five academic rigor and cognitively responsive teaching scales than hard disciplines (with β ranging from -.106 in the standards and expectations scale to -.376 in the cognitive complexity-high scale; $p \leq .05$). Further, the effect sizes of discipline on cognitive complexity and supporting changing views were particularly strong (β=-.376 and -.239, respectively). In a similarly robust pattern of findings, none of the five regression models demonstrated that applied versus pure categorizations of discipline explained a significant proportion of variance in either academic rigor or cognitively responsive teaching scales[6] ($p \geq .05$).

---

[6] It is worth noting here that we ran interaction effects between soft versus, hard, and applied versus pure, disciplinary categorizations. We had to run this analysis separately due to issues surrounding multicollinearity. We found that soft disciplines (soft and pure, and soft and applied) continued to be significant, not shedding any new light on our findings. As such, we do not include these models in this manuscript.

Table 4. Standardized Coefficients of Predictor and Discipline Control Variables on Teaching Practices (N=459)

	Cognitive Complexity —High Scale	**Standards and Expectations Scale**	**Subject Matter Scale**	**Prior Knowledge Scale**	**Supporting Changing Views Scale**
Class Size (5-10 students is reference group)					
11-25 students	-.072	-.096	-.122	-.085	-.100
26-75 students	-.236**	-.306***	-.154*	-.294***	-.287***
76+ students	-.134	-.187**	-.162**	-.271***	-.241***
Mode	.033	.029	.043	-.125*	-.107*
Discipline (Soft versus Hard)	-.376***	-.106*	-.187***	-.163***	-.239***
Discipline (Applied versus Pure)	.010	-.041	-.053	.025	.086
adj. R^2	.211	.138	.139	.199	.226

We controlled for class size and mode because prior research has shown these factors to influence teaching practices (Hativa, 1997; Pascarella & Terenzini, 2005; Toth & Montagna, 2002). All five scales showed a statistically significant difference in class size. Although there was no significant difference between classes with 5-10 students and 11-25 students ($p \geq .05$), classes with 26-75 students had weaker coursework practices than classes with 5-10 students in all five teaching practice scales (with β ranging from -.154 in the cognitively responsive teaching subect matter scale to -.306 in the cognitive complexity standards and expectations scale; $p \leq .05$). Additionally, four of the five scales (all but the cognitive complexity scale) had statistically significantly higher teaching practices in class sizes with 5-10 student when compared with class sizes of 76+ students (with β ranging from -.162 in the cognitively responsive teaching subect matter scale to -.271 in the cognitively responsive teaching prior knowledge scale; $p \leq .05$). Lastly, mode was only statistically significant in the cognitively responsive teaching scales of prior knowledge (β=-.125; $p \leq .05$) and supporting changing views (β=-.107; $p \leq .05$) meaning that traditional on-campus classes were better able to tap these two constructs in this study.

Discussion and Implications

In the present study, we found that soft fields, like humanities, provided students with stronger teaching practices than hard fields, like mathematics, across all five constructs. This robust finding resonates with prior literature, as, for example, faculty in the soft fields devote most of their teaching time to lectures and seminars, while faculty in hard fields spend most of theirs on laboratories and exercises (Ballantyne, Bain, & Packer, 1999; Milem, Umbach, & Liang, 2004). This finding is rooted in a longstanding discourse about disciplinary differences as notably characterized by Clark (1987) as "small worlds, different worlds." Here we considered three important broader contexts in higher education that intersect with our findings. First, the findings of this study suggested the importance of using theoretically derived heuristics to demarcate disciplines. Second, the ways that assumptions and norms in hard disciplines may lend less naturally to the specific teaching practices we examined in this study that have been deemed important for student learning (Bransford et al., 2000). Finally, important implications for the policy stream that focuses on increasing STEM undergraduate completion at-large and particularly for underrepresented students.

One finding of this study was the support for using theoretically derived heuristics rather than organizational structures for defining the disciplines. We echo the sentiments of Pike and Killian (2001) and Feldman, Smart, and Ethington (2008) who described the importance of using theory rather than organizational categories to differentiate disciplines for analysis. All too often researchers will add "department," "division," or "major" as a control for disciplinary experience in higher education quantitative research studying students or facult,. These organizationally derived proxies for discipline more accurately reflect organizational reporting than epistemological, ontological, axiological, paradigmatic, or methodological convictions. This study found that Biglan's (1973) categorization of soft disciplines produced moderate effects on the undergraduate education experience in courses. The strong influence of disciplines on academics' beliefs, on teaching and on students' learning, would suggest that disciplines could benefit from greater systematic study, especially regarding their effect on the quality of teaching and learning in higher education. The findings of this study are aligned with the broader literature base; previous research using student surveys and institutional records found that soft disciplinary courses hold promise for better teaching practices than their hard discipline counterparts (Kilgo et al. 2017; Laird et al., 2008).

Different disciplinary norms, traditions, assumptions and practices may influence faculty teaching, and, therefore, the particular teaching practices we examined in this study varied according to whether the course utilized a single paradigm (hard courses) or multiplistic paradigm (soft courses). When considering the specific constructs we examined, the paradigmatic differences seem to resonate well with the findings. For example, the level of cognitive complexity that we examined runs on a continuum from remembering (memorizing and repeating back facts) through more complex skills (analyzing, evaluating, creating). Perhaps disciplines with single paradigms focus more on "facts" than on the interpretation that may be necessary for more cognitively complex processing. Similarly, another construct we examine that had a large effect size was "supporting changing views," wherein the faculty member supports students (both cognitively and emotionally) in working through the dissonance they may experience when encountering new subject matter material. This form of teaching practice may resonate more strongly with the assumptions that undergird multiplistic paradigms because it requires the ability to see subject matter as a subjectively understood concept (i.e., one student may understand it differently than another based on their own lived experiences).

Given that courses rooted in hard disciplines may lend themselves less naturally to some of the teaching practices deemed important for student learning (Bransford et al., 2000), we consider that this is embedded in a larger conversation about STEM undergraduate education. There is mounting fear that the United States is not producing a sufficient number of hard discipline graduates, such as those in science, technology, engineering, and mathematics (STEM) fields (Breiner, Harkness, Johnson, & Koehler, 2012; Thomasian, 2011). This fear is ignited by evidence that the United States is falling behind, as reported by international large scale assessments (such as Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA), which measure K-12 students' math and science performance levels (Mullis, Martin, Gonzalez, & Chrostowski, 2004). Further, policy reports at the higher education level warn that if the United States is to maintain global competitiveness, it must produce approximately one million more STEM workers over the next decade (President's Council of Advisors on Science and Technology, 2012).

Good college teaching in STEM disciplines is an important factor towards achieving the goal of increasing STEM-literate citizens. Understanding the disciplinary contexts that shape undergraduate course experiences, particularly those in the hard fields like STEM, can sharpen the conversation about academic rigor and college teaching practices. For example this study found that hard discipline courses, such as STEM, focus on lower levels of cognitive complexity and access and use, instructively, students' prior knowledge when compared to courses in soft disciplines. This leads to additional questions: why do hard courses use lower levels of complexity? Are there STEM faculty who teach similar content in ways that reflect higher level complexity? If so, how? What does "prior knowledge" look like with specific content in STEM disciplines—perhaps if STEM faculty were taught how to look for this lived and cultural knowledge, they would be better equipped to use that knowledge, instructively, in their courses. We suggest that better teaching practices could be carried out by faculty in order to make hard discipline classes more engaging, and culturally relevant, in a way that would inspire a community of STEM learners. While there is a burgeoning scholarship that examines how to do this better (for example, Hora & Ferarre, 2014; Weiman, 2015), our study findings suggested that teaching practices in hard disciplines can still be improved.

The lack of marginalized racial, ethnic, and socioeconomic minorities in STEM has been widely discussed among policymakers and researchers (Burke & Mattis, 2007). For instance, Bianchini's (2013) report published by the National Academies, *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads* (2013), recommended ways to increase minority students in STEM fields. This report, like many others, detailed the problem of underrepresentation of minorities in STEM by placing the onus on systemic barriers such as lack of preparation, access and motivation, affordability, and academic and social support. The report notes that the United States is at a crossroads and that in order to move beyond it, we must "increase the participation and success of underrepresented minorities in STEM education" (2013, p. 7). In exploring possible best practices toward achieving increased minority

STEM participation in higher education, the report details many important areas, such as resources, research experiences, mentoring, and tutoring. This report, though, overlooks the teaching and learning practices that occur in the classroom. However, the teaching practices examined in this study (e.g., prior knowledge and supporting changing views) have particular relevance for students of color and other underrepresented groups (Castillo-Montoya, 2017; Neumann, 2014). Given that there is a policy push to educate more STEM students, and particularly more racial minorities and women, and that this study found that strong teaching practices (including culturally relevant teaching practices) were less likely to occur in hard disciplines, we see STEM teaching and learning, particularly for minorities, as an area in need of continued attention.

Limitations

This study was exploratory in nature with a sample of only nine institutions, and is therefore not representative of all higher education institutions. However, as this was a multi-institutional quantitative observational study of disciplinary effect on coursework practices, we see this as an important step toward examining in-class practices. In addition, while this study examined differences in some our scales for academic rigor and cognitively responsive teaching across disciplinary categorization, it did not explore the role of institutional culture on how faculty approach teaching. Because organizational, political, economic and social factors affect teaching and faculty careers (Biglan, 1973; O’Meara & Campbell, 2011), we suggest future research incorporate organizational frameworks to explore how pedagogical choices vary across faculty cultures and disciplines, and that these studies observe select courses across the semester.

References

Alexander, P. A., & Winne, P. H. (2006). *Handbook of educational psychology*. Mahwah, NJ: Erlbaum.

Allen, I. E., & Seaman, J. (2004). *Entering the mainstream: The quality and extent of online education in the United States, 2003 and 2004*. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.

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

Angrosino, M. (2012). Observation-based research. In J. Arthur, M. Waring, R. Coe & L. Hedges (Eds.), *Research methods & methodologies in Education*, 165-169, Los Angeles, CA: Sage Publications.

Astin, A. W. (1993). *What matters in college? Four critical years revisited.* San Francisco, CA: Josey-Bass..

Ballantyne, R., Bain, J. D., & Packer, J. (1999). Researching university teaching in Australia: Themes and issues in academics' reflections. *Studies in Higher Education*, *24*(2), 237-257.

Barnes, L. L., Bull, K. S., Campbell, N. J., & Perry, K. M. (2001). Effects of academic discipline and teaching goals in predicting grading beliefs among undergraduate teaching faculty. *Research in Higher Education*, *42*(4), 455-467.

Becher, T. (1987). The disciplinary shaping of the profession. In B. R. Clark (Ed.), *The academic profession: National, disciplinary, and institutional setting*. Berkeley and Los Angeles: University of California Press.Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education*, *19*(2), 151-161.

Bianchini, J. A. (2013). Expanding underrepresented minority participation: America's science and technology talent at the crossroads. *Science Education*, *97*(1), 163-166.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of applied psychology*, *57*(3), 195-203.

Bloom, B. S., Englehart, M. B., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, the classification of educational goals – Handbook I: Cognitive domain*. New York: McKay.

Bransford, J. D., Brown, A., & Cocking, R. (2000). How people learn: Mind, brain, experience,and school. *Washington, DC: National Research Council.*

Braxton, J. M. (1995). Disciplines with an affinity for the improvement of undergraduate education. *New Directions for Teaching and Learning*, (64), 59-64.

Braxton, J. M., & Hargens, L. L. (1996). Variations among academic disciplines: Analytical frameworks and research. *Higher Education: Handbook of Theory and Research, 11*, 1–45.

Braxton, J. M., Olsen, D., & Simmons, A. (1998). Affinity disciplines and the use of principles of good practice for undergraduate education. *Research in Higher Education*, *39*(3), 299-318.

Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics*, *112*(1), 3-11.

Burke, R. J., & Mattis, M. C. (Eds.). (2007). *Women and minorities in science, technology, engineering, and mathematics: Upping the numbers*. Edward Elgar Publishing.

Campbell, C. M. (2017). An inside view: The utility of quantitative observation in understanding college educational experiences. *Journal of College Student Development, 58*(2), 290-299.

Campbell, C. M., & Dortch, D. (2018). Reconsidering academic rigor: Posing and supporting rigorous course practices at two research institutions. *Teachers College Record*.

Castillo-Montoya, M. (2017). Deepening understanding of prior knowledge: What diverse first-generation college students in the U.S. can teach us. *Teaching in Higher Education*, *22*(5), 587-603.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284-290.

Clark, B. R. (1987). *The academic life. Small worlds, different worlds. A Carnegie Foundation special report*. Princeton University Press, 3175 Princeton Pike, Lawrenceville, NJ 08648.

Craney, C., Mckay, T., Mazzeo, A., Morris, J., Prigodich, C., & de Groot, R. (2011). Cross discipline perceptions of the undergraduate research experience. *The Journal of Higher Education*, *82*(1), 92-113.

Del Favero, M. (2005). The social dimension of academic discipline as a discriminator ofacademic deans' administrative behaviors. *The Review of Higher Education*, *29*(1), 69-96.

Del Favero, M. (2006). An examination of the relationship between academic discipline and cognitive complexity in academic deans' administrative behavior. *Research in Higher Education*, *47*(3), 281-315.

Feldman, K. A., Smart, J. C., & Ethington, C. A. (2008). Using Holland's theory to study patterns of college student success: The impact of major fields on students. *Higher Education: Handbook of Theory and Research, 23,* 329-380.

Finnegan, D. E., & Gamson, Z. F. (1996). Disciplinary adaptations to research culture in comprehensive institutions. *Review of Higher Education*, *19*, 141-178.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23-24.

Hancock, G. R., & Mueller, R. O. (Eds.). (2006). *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing, Inc.

Hativa, N. & Marincovich, M. (1995). *Disciplinary differences in teaching and learning: Implications for practice.* San Francisco, CA: Jossey-Bass.

Hativa, N. (1997) Teaching in a research university: Professors' conceptions, practices, and disciplinary differences. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, Il.

Hora, M.T. & Ferrare, J. J. (2014). Remeasuring postsecondary teaching: How singular categories of instruction obscure the multiple dimensions of classroom practice. *Journal of College Science Teaching, 43*(3), 36-41.

Johnson, I. Y. (2010). Class size and student performance at a public research university: A cross-classified model. *Research in Higher Education*, *51*(8), 701-723.

Jung, I., & Rha, I. (2000). Effectiveness and cost-effectiveness of online education: a review of the literature. *Educational Technology*, *40*(4), 57-60.

Jung, J. (2012). Faculty Research Productivity in Hong Kong across Academic Discipline. *Higher Education Studies, 2*(4), 1-13.

Kilgo, C. A., Culver, K. C., Young, R. L., & Paulsen, M. B. (2017). The relationship between students' perceptions of "Good Practices for Undergraduate Education" and the paradigmatic development of disciplines in course-taking behavior. *Research in Higher Education*, *58*(4), 430-448.

Kuh, G. D., Laird, T. F. N., & Umbach, P. D. (2004). Aligning faculty activities & student behavior: Realizing the promise of greater expectations. *Liberal Education*, *90*(4), 24-31.

Laird, T. F. N., Shoup, R., Kuh, G. D., & Schwarz, M. J. (2008). The effects of discipline on deep approaches to student learning and college outcomes. *Research in Higher Education*, *49*(6), 469-494.

Lattuca, L. R., & Stark, J. S. (2011). *Shaping the college curriculum: Academic plans in context*. San Francisco, CA: John Wiley & Sons.

Lindblom-Ylänne, S., Trigwell, K., Nevgi, A., & Ashwin, P. (2006). How approaches to teaching are affected by discipline and teaching context. *Studies in Higher Education, 31*(3), 285-298.

Michel, J. O., Chadi, D., Jimenez, M., & Campbell, C. M. (2018). Ignis fatuus effect of faculty category: Is the tenure versus non-tenure debate meaningful to students' course experiences?. *Innovative Higher Education*, *43*(3), 201-216.

Milem, J. F., Umbach, P. D., & Liang, C. T. (2004). Exploring the perpetuation hypothesis: The role of colleges and universities in desegregating society. *Journal of College Student Development*, *45*(6), 688-700.

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. *Best practices in quantitative methods*, *488-508*.

Mullis, I. V., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.

Neumann, A. (2014). Staking a claim on learning: What we should know about learning in higher education and why. *The Review of Higher Education*, *37*(2), 249-267.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, *26*(2), 135-146.

O'Meara, K., & Campbell, C. M. (2011). Faculty sense of agency in decisions about work and family. *The Review of Higher Education*, *34*(3), 447-476.

Pascarella, E.T. & Terenzini, P.T. (2005). *How college affects students: Findings and insights from twenty years of research.* Jossey-Bass: San Francisco, CA.

Paulsen, M. B., & Wells, C. T. (1998). Domain differences in the epistemological beliefs of college students. *Research in Higher Education, 39*(4), 365-384.

Pike, G. R., & Killian, T. S. (2001). Reported gains in student learning: Do academic disciplines make a difference?. *Research in Higher Education*, *42*(4), 429-454.

President's Council of Advisors on Science and Technology (2012). Engage to Excel: Producing One Million Additional College Graduates With Degrees in Science, Technology, Engineering, and Mathematics. US Government Printing Office. Retrieved from: www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excelfinal_2-25-12.pdf.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage Publications.

Smart, J. C., & Ethington, C. A. (1995). Disciplinary and institutional differences in undergraduate education goals. *New Directions for Teaching and Learning*, *64*, 49-57.

Stallings, J.A., & Mohlman, G.G. (1988). Classroom observation techniques. In J.P. Keves' (Ed.), *Educational research, methodology, and measurement: An international handbook.* Oxford, England: Pergamon, 469-474.

Thomasian, J. (2011). Building a science, technology, engineering, and math education agenda—An update of state actions. Washington, DC: National Governors Association. Retrieved from http://www .nga.org/files/live/sites/NGA/home /nga-center-for-best-practices/col2 -content/list-center-bottom/top-down loads/building-a-science-technology.

Toth, L.S. & Montagna, L.G. (2002). Class size and achievement in higher education: A summary of current research. *College Student Journal, 36*(2), 253-260.

Waxman, H. C., Tharp, R. G., & Hilberg, R. S. (Eds.). (2004). *Observational research in US classrooms: New approaches for understanding cultural and linguistic diversity*. Cambridge, MA: University Press.

Weiman, C. (2015). Comparative Cognitive Task Analyses of Experimental Science and Instructional Laboratory Courses. *The Physics Teacher, 53*(6), 349-351.

Xu, Y. (2008). Faculty turnover: Discipline-specific attention is warranted. *Research in Higher Education*, *49*(1), 40–61.