Reconsidering Academic Rigor: Posing and Supporting Rigorous Course Practices at Two Research Institutions

Authors:

Corbin M. Campbell
campbell2@tc.columbia.edu
212-531-5182
Teachers College, Columbia University
525 W. 120th St., Box 101
New York City, NY 10027

Corbin M. Campbell, Ph.D., is Assistant Professor of Higher Education at Teachers College, Columbia University. Her work, broadly situated, examines the organizational contexts that support learning and growth for students and faculty in higher education. Her recent work, awarded a 2015 National Academy of Education/Spencer Foundation postdoctoral fellowship, studies new ways to conceptualize and measure the educational quality of colleges and universities that more closely reflect the teaching and learning process. By understanding college quality from a teaching and learning perspective, her work questions the current institutional prestige and reward structures in higher education. Recent publications on these topics have appeared in: *Journal of Higher Education*, *Research in Higher Education*, the *Review of Higher Education*, and *Higher Education: Handbook of Theory and Research.*

Deniece Dortch, Ph.D.
deniece.dortch@utah.edu
801.581.6064
The University of Utah
College of Education
1721 Campus Center Drive, SAEC #2251
Salt Lake City, UT 84101

Deniece Dortch, Ph.D., is a Postdoctoral Research Fellow in the College of Education at The University of Utah. Through her program of research, she studies the experiences of Black graduate students, and the institutional policies and practices that influence students' academic and workforce trajectories. Dr. Dortch earned her Doctorate in Educational Leadership & Policy (with a concentration in Higher Education) from the University of Wisconsin-Madison, a Master's in Education (with a concentration in Higher & Postsecondary Education) from Teachers College, Columbia University, a Master of Arts in Intercultural Service, Leadership & Management (with a concentration in Diversity Leadership & Social Justice) and a Bachelor's of Arts in Language & International Trade from Eastern Michigan University. Dr. Dortch is the inaugural Director of the African American Doctoral Scholars Initiative at the University of Utah. Prior to beginning doctoral work, Dr. Dortch served as the Program Director of Leadership Programs at the Memorial Student Center at Texas AM University. She is originally from Holland, Michigan.

In press at *Teachers College Record*

Description: This study reconsiders academic rigor by using a new conceptual framework that focuses on rigorous course practices and by using quantitative observational methods at two selective research institutions. Findings suggest that rigor can be cultivated by classroom environments and educational practices, such as active learning and cognitively responsive teaching.

Structured Abstract

Background/Context: U.S. institutions of higher education have been criticized for limited learning gains and lacking rigor (Arum & Roksa, 2011). Most understandings of academic rigor in higher education focus on how rigor manifests in students in terms of amount of work or approach to learning (Entwistle & Ramsden, 1983; Kuh et al., 2005; Reason, Cox, McIntosh, & Terenzini, 2010).

Purpose/Objective/Research Question/Focus of Study: This study examines rigor as posed by course practices. We define rigorous course practices 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. This study sought to further the discourse on college academic rigor by describing rigor in coursework at two selective research institutions and examining which course contexts and teaching practices were associated with higher levels of rigor.

Setting: We studied two highly ranked, highly residential, selective, very high research institutions on the East Coast of the United States: a mid-sized (< 5,000 undergraduates), private, urban institution and a large (~15,000 undergraduates), public institution.

Population/Participants: We sampled 400 courses at each institution. Of the faculty who taught these courses, 31.4% agreed to participate. We conducted 150 class observations: 99 at site one and 51 at site two.

Research Design: This study used a quantitative observational protocol.

Data Collection and Analysis: Data were collected during a week-long site visit with observers using a structured rubric. Data were analyzed using descriptive statistics and OLS regression in blocks, partitioning the variance in academic rigor that can be explained by course characteristics (e.g., class size, discipline) and teaching practices (active learning, cognitively responsive teaching).

Findings/Results: Most courses in our sample focused on applying, and 85% of courses achieved a higher order level of cognitive complexity (analyzing, evaluating, or creating) at some point during the class session. Active learning and cognitively responsive teaching practices were associated with higher cognitive complexity and greater standards and expectations in the course.

Conclusions/Recommendations: The discourse on academic rigor in higher education warrants further scrutiny and, could be balanced by studies that provide greater depth in the educational practices in classrooms. This study suggests that institutions and faculty may have a significant role in scaffolding rigor. Academic rigor is not simply about having bright, dedicated, and hard-working students, but, is also determined by classroom environments and processes that can be cultivated.

Acknowledgement: This research was graciously supported by a National Academy of Education/Spencer Foundation postdoctoral fellowship.

Reconsidering Academic Rigor: Posing and Supporting Rigorous Course Practices at Two Research Institutions

Learning, growth, and development in college require students to experience some form of challenge, dissonance, vigilance, questioning of assumptions, or crisis of understanding (Astin, 1984; Kuh, Kinzie, Schuh, Whitt & Associates, 2005; Neumann, 2014; Pace, 1980; Sanford, 1967). This model of learning is rooted in decades of research on cognitive and psychosocial development (e.g., Erikson, 1968; Perry, 1968) indicating that students need to experience a combination of challenge and support in order for learning and development to occur. Yet, critics of U.S. higher education claim that colleges and universities offer an education that is lacking in academic rigor (Arum & Roksa, 2011; Bok, 2006; Campbell & Cabrera, 2014; Jaschik, 2011; Pascarella, Blaich, Martin, & Hanson, 2011). Arum and Roksa's book, *Academically Adrift*, states that students today spend less time on academic requirements, yet earn higher grades and graduate at higher rates. The public media has also questioned the rigor of a college degree, as seen in the *New York Times*: "In just 10 days, academically deficient players could earn three credits and an easy 'A' from Western Oklahoma State University for courses like 'Microcomputer Applications' (opening folders in Windows) or 'Nutrition' (stating whether or not the students used vitamins)" (Carey, 2012, para. 2). These dubious accounts of college rigor raise questions about how rigor manifests (or not) in college classrooms and which classroom practices support higher levels of rigor.

Most current understandings of academic rigor in higher education focus on how rigor manifests in students: the amount of student work (e.g., number of hours spent studying, number of large papers; Kuh et al., 2005; Pace, 1980) or the student's approach to learning (e.g., deep approaches to learning; Entwistle & Ramsden, 1983, Nelson Laird, Seifert, Pascarella, Mayhew & Blaich, 2014; Reason, Cox, McIntosh, & Terenzini, 2010). Research that understands the number of hours that students worked, the number of papers they write, or their approach to learning yields important information about how students experience rigor. Yet, this research does not provide a window into the practices that underlie this commitment by students. This study considers an alternative viewpoint on rigor by focusing on the level of rigor as *posed by the course practices.* We define rigorous course practices 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. 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. This study sought to further the discourse on college academic rigor by exploring rigor in coursework at two selective research institutions (one private and one public) and examining which course contexts and teaching practices are associated with higher levels of rigor.

Is College Rigorous?

One of the most controversial examples of academic rigor research is derived from Arum and Roksa's (2011) book, *Academically Adrift,* which describes that although a modest proportion of college students do make incremental gains in critical thinking and reasoning skills, the majority make little to no substantial gains in these areas. In this book, Arum and Roksa contend that there is a connection between student's lack of "hard work" in college (i.e. less time spent on coursework) and the limited learning gains:

Students' lack of academic focus at today's colleges…has had little impact on their

GPA… as they have developed… 'the art of college management' in which success is achieved…not through hard work but through 'controlling college' by shaping schedules, taming professors, and limiting workload. (p. 4)

Additional efforts to assess academic rigor have employed self-report surveys to give descriptive accounts of students' time use, with a particular focus on the hours that students spend studying. Babcock and Marks (2011) studied the amount of time students spent studying over time from 1961 to 2003, and found that full-time students reduced the time they studied, on average from 40 to 27 hours per week. This study was convincing in that it accounted for framing effects across different surveys, disciplinary effects, and the evolving composition of college students over time. In a summary they prepared for the American Enterprise Institute, Babcock and Marks (2010) largely attributed this decline to falling standards in higher education and increasing emphasis on student "leisure." McCormick (2011) discussed other possible ways to account for the declining time spent, including the changing nature of higher education institutions during this time period, reward structures that focus less on teaching, and the changing nature of pedagogy, which focuses on practices such as collaborative learning and service learning that may be rigorous, but not counted in traditional "studying" time.

The discourse on low academic challenge in college (Arum & Roksa, 2011; Braxton, 1993; Bok, 2006; Campbell & Cabrera, 2014; Pascarella et al., 2011) has been combined with the pressure to improve graduation rates (Frost & Teodorescu, 2001). Despite students spending less time on academic requirements, graduation rates and grades have increased (Arum & Roksa, 2011; Frost & Teodorescu, 2001; Hu, 2005; Rojstaczer & Healy, 2012). Rojstaczer and Healy (2012) collected historical and contemporary grading data from over 200 four-year colleges and universities and found that, "on average across a wide range of schools, A's represent 43% of all letter grades, an increase of 28 percentage points since 1960" (p. 1). Rojstaczer and Healy attribute the grade inflation to an erosion of academic standards over time, largely driven by the consumerism in higher education (i.e. the desire to keep students satisfied rather than accountable) and the lack of regulation in grading by faculty. Likewise, Babcock and Marks (2011) attribute the decline in time spent studying to how institutions cater to student preferences for leisure (via, for example, student teaching evaluations) and how there are few institutional structures that reward faculty for rigorous teaching. McCormick (2011) argues that these declines are more complex than simply responding to student desire for leisure, taking into account that the entire system of higher education has changed via the massification of higher education and rewards structures that emphasize research over teaching. Indeed, there are several forces in higher education today that place importance on increasing enrollments and improving graduation rates, and few forces that focus on the level of student learning (Babcock & Marks, 2011; Campbell & Cabrera, 2014; McCormick, 2011).

As such, one strand of research on college rigor posits that students study fewer hours than in past decades and get better grades. Yet, the amount of time spent studying may be a blunt tool to understand a robust construct like college rigor. As such, student self-report surveys have offered descriptive accounts of more robust understandings of rigor in the college academic experience. For example the National Survey of Student Engagement (NSSE) created four engagement indicators that comprise "academic challenge" (2013), including, for example "quantitative reasoning" (e.g., students' reports of how often during the current school year they "used numerical information to examine a real-world problem") and "learning strategies" (e.g., students' reports of how often during the school year they "reviewed [their] notes after class." They also created scales to measure Deep Approaches to Learning (DAL), which include integrative learning, high order learning, and reflective approaches to learning (Nelson Laird et al., 2014). Although several studies have connected the DAL scales to broad-based college learning outcomes (Nelson Laird et al., 2014; Reason et al., 2010; Renaud & Murray, 2007; Wang et al., 2014), Campbell and Cabrera (2014) found that Deep Approaches to Learning were not a strong predictor of GPA, adding to the concern that academic rigor, such as deep approaches to learning, may not be rewarded in the higher education grading structure.

In sum, the descriptive accounts of rigor have progressed from understanding rigor as the number of hours worked to a broader view that examines students' description of the depth of their learning experiences. While this change may represent progress in the conceptualization of rigor, many of the studies that have fueled the concern about college academic rigor have stemmed from a single standardized test or self-report survey. Certain scholars have questioned whether self-report surveys accurately measure educational experiences and gains (Bowman, 2010; Campbell & Cabrera, 2011; Porter, 2011, 2013), while others question the ability of a single standardized test to understand the breadth of academic rigor at the college level (Liu, Bridgeman, & Adler, 2012; Pascarella et al., 2011). Additionally, the studies that rely on students' self-reports ask them to aggregate their assessment of rigorous practices across all their coursework over an academic year. This removes rigor from the course experience, leaving aside subject matter and disciplinary contexts for rigor and the specific teaching practices that bring about a rigorous experience for students. As such, questions remain about the claim that a college education is lacking in rigor and how rigor manifests within courses.

Rigor as Embedded within College Courses

The perspective on rigor that students' can offer is important—how students feel about the level of challenge in coursework is certainly relevant to understanding college rigor. Yet, understanding how rigor manifests in courses may provide alternate explanations for the current rhetoric around rigor. To date, few studies have examined rigor at the course level. Braxton and Nordvall (1985) investigated the level of cognitive complexity in 158 examinations across liberal arts institutions and found that institutions with higher selectivity tended to require examinations with higher cognitive complexity. Braxton (1996) did a follow up study with 115 examinations across 40 research institutions, with different findings: selective and non-selective institutions largely had similar levels of cognitive complexity. There are also a few more recent examples of studies of cognitive complexity at the course level. Weiman (2015) compared the cognitive tasks required of expert physicists as they design and execute research to the cognitive tasks required of students in introductory laboratory courses. In another example, as a facet of a study on college teaching, Hora and Ferrare (2014) examined the level of cognitive engagement in STEM courses using an observational protocol. While Weiman's and Hora and Ferrare's studies shed light on cognitive engagement at the course level, these studies do not connect the findings to the narrative on college rigor.

Connecting the recent theories of college teaching and learning with the narrative on college rigor may uncover different understandings than prior research focusing on student's reports of their experiences across course contexts. For example, by looking descriptively at the number of hours that students work (across all their coursework) and the proportion of students that receive A grades, Babcock and Marks (2011) offered that students' emphasis on leisure was driving grade inflation. An alternative explanation for the increase in grades over time may be the proliferation of mastery learning in the past few decades. In essence, this perspective would hold that that the goal is not to "weed out" learners in courses, but instead to ensure that all learners achieve mastery of a topic. Several "mastery learning programs" were developed for college coursework during the late 1960s, 70s, and 80s—for example Bloom's Learning for Mastery (LFM) and Keller's Personalized System of Instruction (PSI) (Kulik, Kulik, & Bangert-Drown, 1990). In these programs, all students were required to master individual units of course content before moving forward to new content. Following this perspective, an increase in grades over time would signify that more students are developing mastery of the course content rather than evidence a lowering of standards.

In another example of how teaching and learning theory and the narrative on college rigor may intersect, there may be a difference between the frequency of engaging in rigorous behaviors and the depth of engagement in those behaviors. Students may self-report how often they were engaged in analyzing problem sets or writing long papers (NSSE, 2013). However, theories from the learning sciences demonstrate the importance of learning in-depth fewer core concepts of subject matter in a course (Bransford, et al, 2000). It may, in essence, be more rigorous to examine, very deeply, one core concept from specific subject matter, than to work a greater number of hours on a breadth of topics. By examining rigor at the course level, it may be possible to understand the depth of cognitive engagement in specific course material.

As such, examining rigorous course practices may complement the studies of student's perceptions of rigor across their undergraduate experience. Understanding the extent to which college courses challenge learners to sustain a deep connection to the subject matter and to think in increasingly complex ways, perhaps, can offer different insight into the current narrative on college rigor.

Educational Practices that May Support Rigorous Coursework

In addition to examining, descriptively, how rigorous college coursework is, there is also evidence that certain contexts and teaching practices may be associated with greater rigor.

Course Contexts

Course characteristics, such as class size, class length, class level, and discipline, can serve as an organizing backdrop for student educational experiences in their coursework. Pascarella and Terenzini's (2005) review of the past two decades of research on the effects of class size in higher education acknowledged that larger classes have a negative effect on student's ability to acquire subject matter knowledge. In 1983, Fischer and Grant conducted an observational study of approximately 150 college classrooms and examined the relationship between the level of cognitive complexity using Bloom's framework and course contexts (class size), institutional context (public/private) and teaching practices (active learning). They found that smaller classes had higher cognitive complexity than larger classes. The length of the class session could also be an important course characteristic for rigor. It is possible that a longer class session could achieve a higher level of complexity than a shorter class session. Institutions have manipulated course duration and number of class sessions (e.g., intensive, time-shortened summer classes) to enhance student convenience and persistence (Daniel, 2000). The class level (introductory or advanced) could also play a role in rigor. Introductory gateway courses to popular majors may contain intentionally high levels of rigor to weed out less talented or devoted students (Johnstone & Maloney, 1998) or advanced courses may require more rigor to correlate with students' academic progress throughout college. Similarly, there may be disciplinary differences in the level of rigor in coursework given the important role of discipline in the lives of faculty (Becher and Trowler, 2001) and in faculty teaching (Neumann, 2014). These course contexts may work together in interesting ways with regard to rigor. For example, while introductory courses in STEM may have high standards, they also are often high enrollment classes, which may contrast with the level of cognitive complexity that can be achieved.

Teaching Practices

Although there are many forms of college teaching practices that have been studied over time and are shown to influence classroom learning, we focus on two that may have particular applications to facilitating rigorous coursework: the landmark theory of active learning and the developing theory of cognitively responsive teaching. Decades of research has demonstrated the critical role of certain pedagogies that require active involvement from students in their learning process (Astin, 1993; Carini, Kuh, & Klein, 2006; Dillon, 1990; McKeachie & Kulik, 1975; Pascarella & Terenzini, 2005). This body of literature runs opposed to classroom practices that see faculty as purveyors of knowledge and students as passive receivers. Active learning practices, such as collaborative assignments and projects, have been adopted as a high-impact practice by the American Association of Colleges and Universities (AAC&U, 2007). Similar practices, such as student faculty interaction, student collaboration, and class discussion, also play a central role in the new NSSE engagement measures (NSSE, 2013). Given the importance placed on active learning practices, courses that include an interactive component, like class discussions or class activities, may influence the level of rigor in courses.

The past three decades of research on college teaching has largely focused on broad categories of pedagogical techniques, such as active or passive teaching practices. Following this perspective, a meta-analysis of findings about lecturing versus active learning techniques calls for abolishing lecturing altogether in STEM education (Freeman et al., 2014). More recent college teaching theorists argue that broad sweeping categories (e.g., lecture versus active learning) can obscure subtler teaching practices, imbedded within disciplinary contexts (Hora & Ferrare, 2014; Neumann, 2014). In essence, these categories may not facilitate an understanding of the content of learning, who the learners are, or what the instructors do in their teaching practices (Neumann, 2014). By contrast, theories that are brought from the K-12 fields and the learning sciences attempt to bring instructors, learners, and subject matter to the center of studies of college teaching and learning. For example, Neumann (2014) contends that good college teaching requires instructors to be responsive to student's cognition of the subject matter. Neumann describes that instructors do this by teaching subject matter core ideas in-depth, by understanding student's prior knowledge of the subject matter, and by bridging student's prior understandings to the new course ideas. Given that this framework focuses on framing course content toward student's cognitions, it is possible that these practices could support the level of cognitive complexity in the classroom.

Viewing rigor as the cognitive challenge required by coursework, this study explores new ways of understanding academic rigor in higher education courses and adds to the growing narrative on college rigor by asking the following:

1. Given the emerging view of rigor in college coursework, how academically rigorous are college courses, overall, at two selective, research universities (one private and one public)?
2. Which course characteristics (class size, length of class, class level, and discipline) are associated with a greater level of rigor in courses?
3. Holding course characteristics constant, are certain teaching practices (active learning and cognitively responsive teaching) associated with the level of academic rigor in courses?

Conceptual Framework

This study examined rigorous course practices (i.e. challenge) as well as the educational practices that provide scaffolding for a higher level of rigor (i.e. support). As such, the framework for this study is grounded in three conceptual tenets that support the specific variables we use in this study and the relationships we posit among such variables: (a) learning and development requires both challenge and support (Pace, 1984; Sanford, 1967); (b) certain teaching practices and course characteristics can support the level of rigor in college courses (Braxton, Milem, & Sullivan, 2000; Fischer & Grant, 1983); and (c) academic challenge can be seen uniquely be examining individual course contexts and practices (Bensimon, 2007; Braxton & Nordvall, 1985; Dowd, Sawatzky, & Korn, 2007; Nordvall & Braxton, 1996). By focusing on the *rigorous practices* as they manifest in the classroom, this study provides a new vantage point for understanding academic rigor at two selective research institutions.

Perhaps a metaphor for the framework employed in this study is to consider the typical "sink or swim" idiom for rigor, which focuses on students' determination to stay afloat. Instead, the framework we used in this study would suggest that instructors have a role in finding a deep pool and throwing students in (i.e. they pose the challenge). Simultaneously, instructors can also find an appropriate life vest for each student and provide swim lessons before they throw the students in the water (i.e. educational practices that support rigorous coursework).

We present a visual for the conceptual framework that guided this study in Figure 1. In this figure, the hypothesis that there is a relationship between supportive classroom practices (course characteristics and teaching practices) and rigorous coursework is supported by Sanford's 1967 landmark theory of challenge and support. Under this theory, Sanford posits that supports must be in place in order for students to meet greater challenges in the academic environment, and that the pairing of challenge and support is what ultimately produces learning and development. Sanford built his theory off of decades of psychological research that demonstrates that "a person develops through being challenged: for change to occur, there must be internal or external stimuli which upset his existing equilibrium" (p. 51). At the same time, Sanford's theory contends that the challenge can be so great that an individual retreats and no learning or development takes place. As such, Sanford's theory offers that individuals must be "readied" for the challenge that they will encounter. Sanford suggests that educational processes (such as teaching and institutional and class contexts) can provide scaffolding for productive challenge to take place in students, resulting in learning and development. This landmark theory has been cited and reinforced in countless more recent studies of college student learning and development (e.g., Lechuga, Clerc, & Howell, 2009; Reason, Cox, Quaye, & Terenzini, 2010; Terenzini, 1999).

We clarify that this model posits a pairing between challenge and support rather than a directionality. For example, it does not matter whether a teacher creates supports in order to facilitate a higher level of challenge or whether an instructor seeks a high level of rigor and then provides the associated scaffolding to support students through the challenge. Perhaps faculty who plan courses with a high level of cognitive complexity therefore integrate supportive practices, such as active learning. On the other hand, it may be that using active learning in the classroom spurs greater levels of complexity among students. As such, there is likely reciprocity in the challenge and support hypothesis. Regardless of the directionality, Sanford posits that the pairing of challenge and support, together, facilitate learning. Support alone, does not produce the necessary conditions for learning and development. Simultaneously, challenge can be overwhelming without the scaffolds and supports that ready a student for such challenge.

As such, the association between supportive practices and rigorous coursework is of interest. Specifically, we focus on the association between the level of rigor in coursework and certain classroom contexts and teaching practices practices given the initial evidence that such practices facilitate higher order thinking skills (Braxton, Milem, & Sullivan, 2000; Fischer & Grant, 1983). Examining the teaching practices and course contexts that are paired with a high level of rigor could be instructive for college teachers and other stakeholders who are interested in understanding how rigor manifests in higher education. Our operationalization of rigor is supported by Bensimon's (2007) assertion that ways of defining success in higher education should be focused on college practices and practitioners rather than on student characteristics and effort. As such, we defined rigor using theories of cognitive complexity and the standards in college courses. We expand upon these definitions, below.

[INSERT FIGURE 1 ABOUT HERE]

Rigorous Courses: The Deep End

Given that we study rigor in terms of the cognitive challenge of the course, we chose a two faceted framework that would allow us to consider both the depth and the breadth of cognitive challenge as required by the course. To understand the depth of cognitive challenge we borrowed from Braxton and Nordvall (1985), who used Bloom's 1956 taxonomy of levels of cognitive complexity to operationalize rigor (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). In this study we used the revision of Bloom's taxonomy conducted by Anderson and Krathwohl in 2001. The revised Bloom's taxonomy contends that academic work can require students to engage cognitively in six increasingly complex levels: remember, understand, apply, analyze, evaluate, and create.

We defined the breadth of cognitive challenge as the level of standards and expectations for the work in the course (Astin, 1993; Hu, 2005; Rojstaczer & Healy, 2012; Stallings, 1980). This aspect of the framework follows Astin's (1993) conception of quality of effort and the long-established relationship between learning and time on task (e.g., Stallings, 1980). While several previous studies have examined how many hours students spend in coursework (e.g., Babcock & Marks, 2011), we focus on the manifestation of expectations for sustained attention in the classroom. For example, we examined how the instructor sets an expectation that students are prepared for the course and the expectations for students to remain engaged in course content during the class. The quantity of work and sustained attention required by coursework is tied to the standards and expectations set by instructors (Hu, 2005; Kuh, Laird, & Umbach, 2004; Rojstaczer & Healy, 2012).

Practices that Support Rigorous Courses: A Life Vest

In addition to examining challenging academic practices, this study investigated whether certain course characteristics and classroom practices support a higher level of rigor in courses. We examined course characteristics (class size, class length, class level, and discipline) as possible contexts for rigorous practices. We used Biglan's (1973) description of disciplines as hard versus soft (disciplines with single paradigms versus multiplistic paradigms) and pure versus applied (knowledge for discovery versus applied knowledge). Demarcating disciplines using theoretical heuristics produces more effects than using organizational definitions of disciplines (Pike & Killian, 2001).

We further studied two forms of teaching practices as possible facilitators of rigorous coursework: active learning and cognitively responsive teaching. We consider "active learning" practices as any pedagogical techniques that involve students in their learning experience, such as group work, case studies, class discussions, structured activities (Braxton, Milem, & Sullivan, 2000). In this way, we are assuming the behavioral lens on active learning, which is prominent in the higher education literature base (Chi & Wylie, 2014). According to Chi and Wylie, "Behavioral engagement refers generally to the notion of participating" as compared to the motivation to participate or the level of cognitive engagement (p. 219). In addition to active learning, we observed cognitively responsive teaching according to Neumann's (2014) claims on teaching and learning. According to Neumann, good teaching requires: (I) deliberately orchestrating an encounter of subject matter ideas culled from the discipline; (II) connecting student's learning of subject matter to prior knowledge and understandings based on their life and cultural experiences; and (III) creating space for students to experience the dissonance between what they already know and the new subject matter ideas and supporting students in working through the cognitive and emotional features of the dissonance.

Methods

Data collected for this study were derived from a larger study of two selective, highly residential, very high research universities (one private and one public) aimed at exploring alternative ways of understanding and assessing college educational quality (including academic rigor) at the institutional level (see *website—obscured for blind review*). We collected data using quantitative observation for a representative sample of courses across the two institutional sites. Quantitative observation allowed for a deeper investigation of rigorous coursework and the practices that support high levels of rigor as they unfolded between faculty and students, between course ideas and students, and among students in a course setting.

By "quantitative observation" we refer to an observation protocol that uses a closed-ended, highly structured rubric and coding scheme with raters who have been specifically trained to rate according to the conceptual framework used in this study (Angrosino, 2007). This kind of quantitative observational protocol can be used to measure the pervasiveness of a pre-defined phenomenon (such as rigorous course practices at two institutions), in contrast to a more open-ended, in-depth qualitative observational protocol. This method has been widely used in K-12 education as a means for evaluating teaching effectiveness, teaching quality, and effective educational practices (Croninger, Buese, & Larson, 2012; Hill, Charalambous, & Kraft, 2012). Structured observation has also been used in the past decade to study teaching and learning in higher education, particularly (but not exclusively) in the STEM disciplines (see, for example author, in press; Hora & Ferrare, 2014; Weiman, 2015). We used our conceptual understanding of academic rigor to guide development of the detailed rubrics and while formulating training protocols for our raters. We discuss the training and other protocols for the observations in the Procedures section, below.

Sites

Given that we studied educational practices that support rigorous courses, we selected two institutions that purport a high level of rigor based on their levels of prestige and selectivity. We provide broad descriptions about the sites to assist with contextualizing findings and transferability to similar settings while maintaining confidentiality for the sites. Both sites are highly ranked, highly residential, selective, very high research institutions on the East Coast of the United States. One is a mid-sized (< 5,000 undergraduates), private, urban institution. The other is a large (~15,000 undergraduates), public institution in a typical college town. In terms of student characteristics, both sites have high 4 year graduation rates, low transfer rates (both in and out), approximately equal proportions of men and women students (transgender not reported), and more than 2/3 of undergraduates are white, non-Hispanic students.

Sample

There were two different sampling considerations for this study: (a) ability to generalize to the *population of courses* across the two institutions and (b) representation of the "course experience" across the two institutions as defined by *seats in courses*. A sample that considered the number of seats in each course better represented what a majority of students would experience at these institutions: although most courses (69%) had small enrollments (5-25 students), small classes only accounted for 32% of the seats in classes overall. Similarly, although only about 11% of courses had more than 50 students enrolled, these larger classes accounted for almost half of the enrollment in terms of seats. Given our conceptual framework, it was particularly important not to over-emphasize smaller classes—both active learning and cognitively responsive teaching may be easier to enact in courses with smaller enrollments.

To meet our first sampling goal, generalizing to the *population of courses* at the two institutions, we employed a stratified random sampling design to determine which courses would be selected to participate (Lohr, 2009). To get the most representative sample possible for aggregating courses in the class observations to the institution level, we stratified undergraduate courses by class size, level of class (intro or advanced), and discipline (college/school). Prior higher education literature has linked these factors to academic rigor, teaching quality or learning outcomes (Astin, 1993; Daniel, 2000; Johnstone & Maloney, 1998; Pascarella & Terenzini, 2005; Stallings, 1980; Toth & Montagna, 2002). To allow for meeting our second sampling consideration, generalizing to the “course experience” and ensuring representation by seats in courses, we also weighted our sample of courses by seats within class size (again, ensuring that very small classes did not have an undue influence on our findings of the average educational experience for students).

We sampled 400 courses at each institution, and of the faculty who taught these courses, 31.4% agreed to participate. Whereas this response rate may not be entirely ideal, it compares well to national surveys of research universities. While non-governmental national faculty survey response rates across all institutional types are approximately 40-50% (Milem, 1999;

FSSE, 2015), response rates for faculty surveys at research institutions and particularly public research institutions are considerably lower (Eagan et al., 2014). We removed classes with less than 5 students, independent studies, and duplicate courses with the same instructor. At each institution, we had more faculty consent to observation than we had time slots for observers. We selected among the pool of consented courses to achieve our stratified sampling scheme (i.e. if too many consented courses were at noon, we would observe the one that had course characteristics needed to allow for better representation across class size, discipline, and class level). We completed 150 class observations: 99 at site one and 51 at site two. The second site was smaller and had a more condensed schedule (more courses midday). While our sample was smaller at the second institution, our representation across class size, division, and class year was similar in institution 1 and 2. Additionally, we do not compare institutions for the purposes of this study. Table 1 shows the course characteristics (class size, class level, discipline, faculty category) of observed courses and the population of courses. The observed courses were statistically representative of the population of courses in terms of discipline and class level. Courses taught by tenure stream faculty were over-represented in our sample compared to the population of courses at site 2 (site 1 did not provide data on faculty category).

[INSERT TABLE 1 ABOUT HERE]

Procedures

For the class observations, we followed the procedures for quantitative observation (Johnson & Christensen, 2010), which specify that the rubrics and instructions be closed ended, detailed, and specific. The purpose of the class observations was to get a snapshot in time of the average level of academic rigor in the courses across each university. Most classes were observed by two observers during a one-week site visit in mid-spring semester. Due to scheduling conflicts, some courses were rated by one observer (n=24). The observers rated the entire duration of one class period using the observation rubric. All class interactions and activities were included as a part of the rating (e.g., lecture, group discussions, presentations, assignments).

We comprehensively trained 14 observers on the conceptual frameworks, the observation protocol, and the rubrics. Observers were working toward a graduate degree in higher education with a focus on teaching and learning. They were required to take a four-hour training and prove strong reliability with the rest of the research team on rubric ratings during the practice observation. We used a one-way, absolute, average-measure, mixed-effects intra-class correlation (ICC) calculation of the inter-rater reliability, which is appropriate for the current study due to the ordinal nature of the observation categories, more than one rater per class, and our interest in the consistency of the absolute value of the ratings (Hallgren, 2012). The ICC across all items included in this study was .646, with ICCs of sub-scales ranging from .575 to .697. Any ratings that had a discrepancy between raters greater than two response options was removed, leaving 94% of observation ratings included in the analyses.

Given our interest in subject-matter content, observers were matched with courses in their disciplinary expertise whenever possible. We also did post-hoc analyses to determine the extent to which disciplinary expertise matters in the ratings. We investigated whether the inter-rater reliability would be greater in courses where both observers had subject matter expertise compared with observers where only one had substantive subject matter expertise. We found that although the pairings of observers who both had expertise had slightly higher inter-rater reliability compared to pairing one expert observer with one non-expert observer (ICC=.748, .715, respectively), both pairings had good inter-rater reliability. The similarity in the two inter-rater reliabilities seems to indicate that disciplinary background did not have a large effect on ratings. Yet, we do see observation of specific disciplines as an important area of future research.

Data Sources

We created rubrics for the class observations to assess educational quality as defined by academic rigor, teaching quality, and learning goals. The present study focuses on the items that measured academic rigor and teaching quality. The development of the rubrics was assisted by both content experts (in academic rigor, teaching quality, and learning goals) as well as methodological experts (in survey and rubric design), who tested the rubrics for content and response process validity.

Following our conceptual framework on academic rigor, we created three scales: two scales on cognitive complexity and one scale on standards and expectations. Based on our conceptual framework on teaching practices we created one scale to measure cognitively responsive teaching, and a composite measure for active learning. Each of the four scales (cognitive complexity—high, cognitive complexity—average, standards and expectations, and cognitively responsive teaching) were created using Principal Components Analysis. Loadings across all constructs were greater than .4, with only two items less than .7. The Cronbach's Alpha across all scales ranged from .822 to .970. Scale scores are means of the individual items that were scored within each construct (e.g., if a class had lecture and handout, but no questions, discussion, or activities, the cognitive complexity scale scores would have been calculated as the mean of the lecture and handout scores). We provide sample items from each scale in Table 2[1]. We also provide examples of how each of these scales might have manifested in a course in Table 3 (academic rigor) and Table 4 (cognitively responsive teaching). We further describe the operationalization of cognitive complexity, standards and expectations, cognitively responsive teaching, and active learning, below.

[INSERT TABLES 2, 3, & 4 ABOUT HERE]

The two sub-scales for the cognitive complexity of the course were the highest level of complexity and the average level in order to score the variation in complexity that could take place during one class period. The highest cognitive complexity was the point within the class period that reached the highest level of complexity, compared to the average level of complexity across the class period. For example a class may have focused mainly at understand (average level of complexity) but dabbled in analyze (highest level of complexity). The sub-scale on standards and expectations focused on the standards set for preparation for and participation in class. For example, in observation, these standards could have been seen in the instructor requiring students to talk about readings from the previous week, or in student norms for being prepared for class, such as bringing questions to class about the week's course content.

To answer research question three, we created one cognitively responsive teaching scale and one active learning composite item. The cognitively responsive teaching scale was created as a composite mean of three sub-scales, following Neumann's framework on teaching quality: subject matter teaching (for example, the instructor using multiple representations of core ideas), prior knowledge (for example, the instructor working to understand a student's prior views of the course content), and supporting changing views (for example, helping students to bridge prior understandings with new course material). To measure active learning during observation, we used two rubric items that assessed the presence of active learning techniques. The first measured whether the class included class activities, defined as a structured activity where students engaged with the course content (e.g., case studies, course material reviews using clickers, group work). The second item assessed the presence of class discussion, defined as back and forth conversation between instructor and students or among students about the course content. Given that both discussion among students and structured activities have been consistently connected to student learning in prior literature (Pascarella & Terenzini, 2005), and given that the literature has largely viewed active learning as a broad category, compared against didactic approaches (Hora & Ferrare, 2014), we coded active learning dichotomously: the course used “active learning” practices if it contained class activities and/or class discussion.

Analysis

In order to answer research question 1, we weighted our observation data by class size to better generalize to the population of courses and then ran descriptive statistics. To answer the final two research questions, we ran two Ordinary Least Squares (OLS) regression models. Variables were entered into the regression model in two blocks to partition the variance explained by course characteristics from the two teaching practices of interest (i.e. to determine whether active learning and cognitively responsive teaching influence rigor above and beyond the variance explained by course characteristics). Block 1 included course characteristics (site, class size, length of class, class level, discipline). Block 2 included active learning techniques (whether the class included class discussions or activities) and the scale for cognitively responsive teaching (i.e. Neumann’s (2014) framework). The dependent variables were two academic rigor scales: (a) the average level of cognitive complexity; and (b) the standards and expectations of work. We determined the significance of predictors by $p < .05$. We used the standardized Beta coefficients (β) to interpret effect sizes, and used Keith’s (2015) cut-off scores of > .05 as small; > .1 as moderate, and >.25 as large. 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 Eigenvalues >.01, condition indices <50, and VIF < 2. Last, there were no outliers that had a marked influence on the regression estimates, as evidenced by Studentized Deleted Residuals <+-2 and Cook's D values less than .5. As a result, we retained all data in the analysis. Table 5 displays the descriptives and frequencies for included variables and scales.

[INSERT TABLE 5 ABOUT HERE]

Limitations

One limitation of this work is the scope of the study: two selective research institutions (one private, one public). While results may be applicable to other similar research institutions, care should be used when generalizing to other institutions and especially other institutional types. Additionally, the class observations represent a snapshot in time during mid-semester. Results might have been different if observations had occurred earlier or later in the semester. Results do not capture how instruction progresses across the semester. Additionally, while many observers/raters had some disciplinary background in the course they rated, their level of expertise was not that of a faculty member in the discipline. Moreover, while the training process was designed to ensure consistency across raters, individuals with different understandings of the conceptual frameworks may have rated differently. Likewise, by using a rubric we were explicitly narrowing the view of rigor that we could understand. Future research should use open-ended qualitative observation to further understand how rigor manifests in college classrooms. Additionally, it is possible that the faculty who agreed to participate were more likely than non-responders to use active learning and cognitively responsive practices. This potential non-response bias may mean that the results presented here demonstrate the rigor and the practices that support rigor among faculty who place higher importance on teaching. Finally, the intent of this study was to investigate rigorous coursework—future studies should examine how rigorous coursework relates to students' rigorous performance (e.g., exams, papers).

Results

Research Question 1

To answer research question 1 about the level of rigor in coursework across a public and a private institution, we present detailed, item-by-item and average scale score descriptive results in Table 6.

Cognitive complexity. Raters scored the cognitive complexity required by several facets of the class, including: lecture, handouts or visual material, questions asked by students, questions asked by the instructor, class discussions, and class activities. For the average class, the average level of complexity demonstrated across the class period focused on applying (M=2.89, SD=0.90), and the highest level achieved during class focused on analyzing (M=3.79, SD=1.02). Outside of central tendency, another way to view the findings is to consider the proportion of courses that achieved high and low levels of cognitive complexity. Most courses (85%) achieved a higher order level of cognitive complexity (analyzing, evaluating, or creating) at some point during the class (e.g., during lecture, in the visuals, or during an activity). Additionally, 31% of courses achieved the highest level (create). Yet, these frequencies indicate courses that may have just touched on these higher order levels—for example, a course that focused mainly on remembering or understanding information, but briefly asked students to analyze the course material. When considering the level of cognitive complexity that was central in the class (i.e. the average level of complexity demonstrated during the course), only 21% of courses focused the course content on higher-order levels (analyzing, evaluating, or creating) and none of the courses focused at the create level. This may be explained given that Bloom's framework suggest that these levels are cumulative, and higher order levels build off of earlier lower-order learning. As such, courses that achieve the "create" level will do so by building up from tasks and processes that required remembering and understanding course content.

Focusing in on each of the class components, the two that required the highest levels of complexity were class discussions and class activities (Table 6). Instructor lecture and visuals (e.g., handouts and power point presentations) required lower levels of cognitive complexity. 47% of class activities and 31% of class discussions reached the Create level, by contrast with only 18% of instructor's lectures. While, less than one half of courses included class discussions and about one third of the courses included activities, 71% of courses included at least one of these two active learning techniques.

Standards and expectations. During class observations, on average, about half of students in class were expected to be prepared (e.g., using readings in discussion) or to participate in class (e.g., class activity). For example, an instructor might have asked questions about the readings in class, but only certain students had the opportunity to respond and others were not called upon (by contrast with teaching techniques where all students would have been required to demonstrate their knowledge of the reading or participate). The instructor used teaching techniques that required all students to be prepared for the day's lesson in only 27% of classes. Observers found that most enrolled students attended class (82%), but not all students were engaged in course material during class. In only 27% of the classes, instructors required *all* students to participate. In the average class that was observed, some students were engaged while others were unengaged. There were also 5% of courses where most students "appeared unengaged, bored, or were doing something unrelated to the course during class."

[INSERT TABLE 6 ABOUT HERE]

Research Questions 2 and 3

To answer the final two research questions, we present our results from the two regression models. Table7 provide standardized coefficients and p-values for predictors as well as model ΔR^2values.

Course characteristics. Overall, course characteristics as a group (site, class size, course length, class level, and discipline) were significant contributors to both of the models predicting cognitive complexity and standards and expectations (Table 7). While course characteristics, taken together, had a significant effect on the model, only two individual course characteristics were significant contributors. Smaller classes were associated with higher standards and expectations with moderate to large effect sizes. “Soft” disciplinary courses were associated with higher cognitive complexity compared to hard disciplines, again with a moderate effect.

Teaching Practices. The two teaching practices investigated in this study (active learning and cognitively responsive teaching) exerted significant and positive effects on both cognitive complexity and standards and expectations. This finding is robust when controlling for course characteristics. For example, active learning and cognitively responsive teaching, together, explained 12.5% of the variance in the cognitive complexity scale and 32.4% of the variance in the standards and expectations scale, above and beyond site, class size, class length, class level, and discipline. Active learning techniques contributed significantly to both models (average cognitive complexity, standards and expectations; Table 7). Courses that included class discussions and/or activities were associated with higher levels of academic rigor, with greater effects on standards and expectations (large effect) than on cognitive complexity (moderate effect). Cognitively responsive teaching from Neumann's (2014) framework had a significant and positive association with large effects on both models of rigor: cognitive complexity and standards and expectations (Table 7).

[INSERT TABLE 7 ABOUT HERE]

Practices that Support Rigor in Larger Courses. Enrollment is concentrated in courses that are larger (67.4% of seats are in courses with >25 enrolled). One critique of active learning and cognitively responsive teaching may be that they are easier to facilitate in smaller classes. Given this, we ran two exploratory regression models (unweighted) to determine the practices that support rigorous coursework in medium and large sized courses (enrollments >25; *n*=47; Table 8). Given the small sample size, we use conservative p-values (p<.01). In the models using larger courses, course characteristics did not exert significant influence on either cognitive complexity or standards and expectations. In terms of teaching practices, active learning and cognitively responsive teaching both had a significant association with standards and expectations. Yet, only cognitively responsive teaching had a significant association with the level of cognitive complexity. The effect size of the association between cognitively responsive teaching and cognitive complexity and standards and expectations within these larger courses were rather large (β =.642, .681, respectively). While these findings may be practically important considering the different teaching contexts in larger and smaller courses, the results should be interpreted with caution given the small sample size.

[INSERT TABLE 8 ABOUT HERE]

In Table 9, below, we summarize and compare the variables that showed significant differences across the academic rigor scales. The table shows that the associations for both active learning and cognitively responsive teaching are considerable. Courses with an active learning component, on average, scored 16.4% higher on standards and expectations than the courses that did not include active learning. Cognitively responsive teaching showed a substantive association for larger courses: with a one unit increase in effectiveness in cognitively responsive teaching, the cognitive complexity score and the standards and expectations scores increased by 23.9% and 25.2%, respectively. Again, given the small sample size of larger courses, while promising, this result should be interpreted with caution.

[INSERT TABLE 9 ABOUT HERE]

Discussion

The current rhetoric about college rigor painted by the media (e.g., Carey, 2012) and higher education critics (Arum & Roksa, 2011; Bok, 2008) would suggest that students are not working hard and still achieving high marks. Our study considers academic rigor differently by investigating rigorous course practices: the way that rigor is posed by coursework and the educational contexts and practices that support a rigorous college education. We situate our findings in our specific institutional sites: two selective, highly residential research institutions that tout rigor as a facet of their education. Our findings advance the conversation on college academic rigor in three main ways. First, our findings contribute to understanding rigor by describing rigorous course-based practices (level of cognitive complexity and standards and expectations) across these two selective institutions. Second, we found that rigorous coursework may be scaffolded by certain course contexts, active learning techniques, and cognitively responsive teaching. Finally, cognitively responsive pedagogies may be particularly salient in their association with rigor in larger courses. This third finding, while notable, should be treated as exploratory, warranting replication, given the small sample size of larger courses in this study.

In terms of cognitive complexity, we found that the average college classroom at these two institutions focused mainly on understanding and applying information, and occasionally analyzing the course material. To make meaning of this finding, we can map the revised Bloom's taxonomy categories onto other frameworks that have been used to understand college rigor. The deep approaches to learning framework has recently been a focus of higher education learning research (Campbell & Cabrera, 2014; Nelson Laird et al., 2014; Wang et al. 2014). This framework is based on research in the 1970s by Marton and Saaljo (1976) that demonstrated that students had two distinct ways of making meaning of the information they learned. One approach to learning, termed "deep learning," focused on the interconnections between facts, seeking a broader conceptual understanding of the course material. The deep approach to learning maps onto the upper categories in the hierarchy of the revised Bloom's taxonomy—applying, analyzing, evaluating, and creating. By contrast, a second approach to learning described by Marton and Saaljo, termed "surface learning" or "rote learning" focused on learning individual facts to meet course requirements. Surface learning can most clearly be seen in the first two categories of the Bloom's revised taxonomy—remember and understand (Anderson & Krathwohl, 2001). Recent research on student's approaches to learning in a higher education setting suggests that students taking a deep approach to learning influences broad-based learning outcomes, such as critical thinking, moral reasoning, need for cognition, and literacy attitudes (Nelson Laird et al., 2014; Reason et al., 2010; Wang et al., 2014). It appears that our study would support that most classes (mid-semester) are focusing on fostering both surface (understand) and some deep approaches to learning (apply), but only dabble in other categories of Bloom's taxonomy that might map onto a deeper approach, such as analyzing and evaluating.

Scholars of teaching and learning in higher education (Hora & Ferrare, 2014; Nelson Laird et al., 2014; Renaud and Murray, 2007) as well as practitioners (AAC&U, 2007) have supported the need for higher-level cognitive demand in the college classroom. For example, the American Association of Colleges and Universities describes that cognitive processes and skills, such as integrating, applying, and analyzing information, are key to success in today's workforce and for developing leaders and citizens. Hora and Ferrare described that the teaching and learning process in higher education is complex, and the level of cognitive engagement in the exchange between students and faculty is one important facet of college teaching. In our study of courses via observations, we found that although most courses occasionally used the higher order cognitive processes (e.g., evaluate and create), few courses focused at these levels. Given that we studied only one class session, mid-semester for each course, additional studies should examine whether the level of rigor in courses would be different at other points in the semester and whether the cognitive complexity builds over time and with specific course content.

Additionally, standards and expectations for participation and preparedness could be increased as evidence of greater rigor in coursework. Our raters found that while most students attended class, instructors only required part of the class to participate or be prepared. For example, an instructor might have asked questions about the readings, but only some students answered. In this area, we connect our findings to the studies that describe declines in hours worked by students over time (e.g., Arum & Roksa, 2011; Hu, 2005). Perhaps the lack of expectations around preparation in the classroom (as seen in the present study) may be fostering lower student motivation outside of class.

Turning to practices that support rigorous coursework, like previous studies suggest, we found that course characteristics and certain pedagogies matter in the educational experience of students (e.g., Carini et al., 2006; Daniel, 2000; Johnstone & Maloney, 1998; Pascarella & Terenzini, 2005). In this study, course characteristics (e.g., class size) provided a backdrop for rigor to take place. It may be that instructors find it more manageable to keep high expectations and pose higher levels of cognitive complexity in courses with smaller enrollments—but this effect was smaller in comparison to the effects of the teaching practices. Additionally, courses from soft disciplines demonstrated higher levels of cognitive complexity than courses from hard disciplines. This, perhaps, could be explained by the different manifestations of Bloom's framework in different disciplines. Perhaps the work that students must accomplish in "understanding" course material in math, science, engineering, and other hard disciplines, spans a broader proportion of the course-taking than in disciplines, such as sociology, history, or art, where analysis and evaluation may be a greater focus in the discipline. To further understand this finding, we recommend future studies situating rigor within disciplines.

Active learning pedagogies, such as class discussions and activities, was associated with the level of academic rigor in college coursework. Active learning techniques had a particularly strong association with the standards and expectations in the classroom, with moderate to large effect sizes. This is not surprising given that active learning techniques typically require participation from students and, therefore, require students to be prepared for class. The findings of the association between active learning techniques and academic rigor are supported by previous literature that asserts the important role active pedagogies in student learning. For instance, instructional practices that stress interaction, feedback, and collaborative learning have been found to be associated with gains in problem-solving skills, occupational awareness, and professional confidence among engineering students (Cabrera, Colbeck & Terenzini, 2001; Colbeck, Cabrera & Terenzini, 2001; Volkwein, Lattuca, Harper & Domingo, 2007).

Even when controlling for course characteristics and active learning techniques, cognitively responsive teaching had a significant association with both academic rigor scales, and was particularly related to rigor in larger courses (large effect size). Here we consider the *what*, *how*, *and who* of teaching. Neumann's (2014) framework is centered on the understanding that teaching occurs in specific disciplinary, instructor, and learner contexts. Good college teaching requires attention to the interaction between the subject-matter core ideas embedded in a discipline, students' prior understandings and backgrounds with regard to the subject matter, and instructors own reflexivity of how they understand and translate the course material based on their students' ways of thinking. We found that teaching that is cognitively responsive supports both cognitive complexity and standards in the classroom. This finding resonates with the recent critiques of broad-based categories of pedagogical techniques (e.g., lecture v. active learning; Hora & Ferrare, 2014). Perhaps, it is not only that a teacher makes room for discussion or facilitates an activity that matters for increasing rigor: the content of that discussion or activity appears to be a critical ingredient. For example, the rigor of a course may be supported by a teacher's ability to use multiple perspectives when teaching an in-depth subject matter concept or by a teacher's ability to connect new course material with students' prior knowledge and understandings. Culturally relevant pedagogy (CRP; Ladson-Billings, 1995) may provide an example of a teaching practice that capitalizes on the prior knowledge component of Neumann's framework. CRP connects the culture of the students, prior knowledge, and relevant outside experiences with new knowledge to support students' learning processes. The results of this study support that cognitively responsive teaching, such as CRP, may facilitate a higher level of rigor in coursework.

Cognitively responsive teaching, defined using Neumann's (2014) framework, appears to be particularly important in larger courses. In our exploratory models on academic rigor in larger courses, active learning techniques were not significantly related to cognitive complexity, whereas cognitively responsive teaching had a rather large effect. Perhaps faculty teaching larger courses have a harder time managing broad class discussions and activities, but they are still able to find ways to connect course content with student's backgrounds and the knowledge (often cultural) they bring to class that is reflective of those backgrounds (Gonzalez, Moll, & Amanti, 2005). The robust finding regarding the importance of cognitively responsive teaching warrants further study and replication due to the exploratory nature of these findings with a small sample. Additionally, qualitative studies could explore how instructors deepen subject matter knowledge embedded in the discipline and access and use student's prior knowledge to support students in learning new course material.

Overall, most courses did include an active learning component, but did not capitalize on the cognitively responsive teaching practices that influence rigor. Almost three-quarters of courses included class discussions and/or activities—which this study found supports rigor, particularly in smaller classes. However, most courses did not effectively embrace cognitively responsive teaching. Although most instructors were at least moderately effective in teaching in-depth subject matter ideas (e.g., using multiple representations of the core course ideas), most did not attempt (or were ineffective in) the two other facets of Neumann's (2014) framework (i.e. accessing students' prior knowledge and supporting students through the dissonance between old understandings and new course ideas). Future research could investigate whether this finding was limited to mid-semester given that this study only examined these practices at one time point. This may, perhaps, have implications for faculty development: although many faculty are socialized in their subject-matter expertise and disciplinary conventions, perhaps fewer are socialized towards understanding and using, instructively, students' prior knowledge.

Finally, the framing of this study contributes to a different understanding of academic rigor than the current narrative. This study focused on the academic rigor of colleges and universities as it manifests in college coursework and course practices rather than in student performance and effort. Conceptually, this view advances the narrative by considering that academic rigor in college is a partnership among subject matter, educators, and learners. Perhaps, students may not work more hours unless the coursework (as it manifests among subject matter, faculty, and students) requires students to think in complex ways, draws students to the subject matter, and holds high expectations for students to remain engaged with the course content. This conceptual understanding of rigor resonates with Bensimon (2007), who described the importance of considering educational practices and practitioners as evidence of the quality of education.

Further, this new conceptual view required methods that would allow us to witness course-based practices. We found that using a quantitative observational protocol enriched our ability to answer some questions in greater depth. For example, we were able to watch the educational processes unfold among coursework, students, and faculty—revealing a window into the "who," "what," and "how" of teaching practices. We could witness an instructor react to a student who was struggling with understanding a new course concept or another student making a connection between course material and personal history. Whereas active learning techniques might be captured with a student or faculty survey (e.g., class discussions yes/no/how often), the newer frameworks that focus on a more contextualized understanding of college teaching, such as Neumann's (2014) framework or Hora and Ferrare's (2014) teaching dimensions, may require alternative methods of data collection. This study suggests that utilizing course-based conceptualizations and more in-depth methods such as quantitative observation may yield another viewpoint on academic rigor in higher education that can complement the research on student's attitudes, behaviors, and effort.

Limitations of this Study and Directions for Future Research

This emerging study examines rigor as embedded within courses and teaching practices. The results point to the association between the level of cognitive complexity reached and active learning and cognitively responsive teaching. Yet, this study was framed in particular institutional and course contexts, namely two highly selected research institutions with observations of single class sessions during mid-semester. Additional study is warranted to understand the interaction of teaching practices and the level of cognitive complexity as the course unfolds across the semester. Further, discipline-specific study could examine whether certain course practices have a stronger association with rigor in the context of specific subject-matter. Finally, this study speaks to the rigorous course practices at two highly selective institutions. Given the diversity of higher education institutions in the U.S., future studies should seek to understand how rigor manifests in courses at different kinds of institutions (for example, high diversity, online, or for-profit institutions).

In conclusion, the discourse on academic rigor in higher education warrants further scrutiny. The scathing accounts of American higher education based on standardized test scores, such as Arum and Roksa (2011), could be balanced by studies that provide greater depth and access to the educational practices in institutions and in classrooms. This study only begins to consider rigorous coursework at two selective research institutions, mid-semester. When taking the pulse of rigorous coursework at two research institutions, we wonder, can and should universities be doing more? If more could be done, this study suggests that institutions and faculty may have a significant role in scaffolding rigor. The level of rigor is not simply about having bright, dedicated, and hard-working students, but, perhaps, is also determined by classroom environments and processes that can be cultivated.

References

Association of American Colleges and Universities (AAC&U; 2007). *College learning for the new global century*. Washington, DC: Association of American Colleges and Universities.

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. V. (2007). *Naturalistic Observation*. Walnut Creek, CA: Left Coast Press.

Arum, R., &Roksa, J. (2011).*Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297-307.

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

Babcock, P. & Marks, M. (2011). The Falling Time Cost of College: Evidence from half a century of time use data. *The Review of Economics and Statistics, 93*(2), 468–478.

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: intellectual enquiry and the cultures of disciplines*. Philadelphia, PA: Open University Press.

Bensimon, E. M. (2007). The underestimated significance of practitioner knowledge in the scholarship of student success. *Review of Higher Education, 30*(4), 441–469. doi:10.1353/rhe.2007.0032

Biggs, J. and Tang, C. (2011). *Teaching for quality learning at university: What the student does.* New York, NY: McGraw-Hill

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57*(3), 195.

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, NY: McKay.

Bok, D. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton, NJ: Princeton University Press.

Bowman, N. A. (2010). Can 1st-year students accurately report their learning and development? *American Educational Research Journal, 47*, 466–496. doi:10.3102/0002831209353595

Bransford

Braxton, J.M. (1993). Selectivity and rigor in research universities. *Journal of Higher Education,64*(6),657-675.

Braxton, J. M., & Nordvall, R. C. (1985). Selective liberal arts colleges: Higher quality as well as higher prestige? *The Journal of Higher Education*, 538-554.

Cabrera, A. F., Colbeck C. L. & Terenzini P. T. (2001). Developing performance indicators for assessing classroom teaching practices and student learning. *Research in Higher Education, 42*(3), 327-352.

Campbell, C.M. & Cabrera, A.F. (2011). How sound is NSSE: Investigating the psychometric properties of NSSE at a public research-extensive institution. *Review of Higher Education, 35*(1), 77-103. doi:10.1353/rhe.2011.0035

Campbell, C. M. & Cabrera, A. (2014). Making the mark: Are deep learning and GPA related? *Research in Higher Education*, *55*, 494-507. doi: 10.1007/s11162-013-9323-6

Carey, K. (2012. December 10). Who will hold colleges accountable? *New York Times,* Retrieved from http://www.nytimes.com/2012/12/10/opinion/who-will-hold-colleges-accountable.html?_r=0 .

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education*, *47*(1), 1–32. doi:10.1007/s11162-005-8150-9

Chi, M. T. H. & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243.

Colbeck C. L., Cabrera, A. F. & Terenzini, P. T. (2001). Learning professional confidence: Linking teaching practices, students' self-perceptions, and gender. *Review of Higher Education, 24*(2), pp.173-191.

Croninger, R.G., Buese, D., & Larson, J. (2012). A mixed-methods look at teaching quality: Challenges and possibilities from one study. *Teachers College Record*, *114*(4), 1-36. Retrieved from http://www.tcrecord.org/library/content.asp?contentid=16649

Daniel, E.L. (2000). A review of time-shortened courses across disciplines. *College Student Journal, 34*(2), 298-308.

Dillon, J.T. (1990). *The practice of questioning.* New York, NY: Routledge.

Dowd, A. C., Sawatzky, M. & Korn, R. (2011). Theoretical foundations and a research agenda to validate measures of intercultural effort. *The Review of Higher Education, 35*, 17-44.

Eagan, M. K., Stolzenberg, E. B., Berdan Lozano, J., Aragon, M. C., Suchard, M. R. & Hurtado, S. (2014). *Undergraduate teaching faculty: The 2013–2014 HERI Faculty Survey*. Los Angeles: Higher Education Research Institute, UCLA.

Entwistle, N. J. & Ramsden, P. (1983). *Understanding student learning*. London, UK: Croom Helm.

Erikson, E.H. (1968). *Identity: Youth and crisis*. NewYork, NY: Norton.

Fischer, C. G. & Grant, G. E. (1983). Intellectual Levels in College Classrooms. In C. L. Ellner & C. P. Barners [Editors]. *Studies of College Teaching*. Lexington Books: Lexington, MA.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). *Active learning increases student performance in science, engineering, and mathematics*. Proceedings of the National Academy of Sciences, 111, 8410–8415. doi:10.1073/pnas.1319030111.

Frost, S. H. & Teodorescu, D.(2001). Teaching excellence: How faculty guided change at a research university. *The Review of Higher Education, 24*(4), 397-415. doi:10.1353/rhe.2001.0007

Gonzalez, N., Moll, L. C., & Amanti, C. (2005). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quantitative Methods Psychology, 8*(1), 23-34.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study, *Educational Researcher*, *41*(2), 56-64. doi:10.3102/0013189X12437203

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. Retrieved from http://www.nsta.org/college/

Hu, S. (2005). Beyond grade inflation: Grading problems in higher education. *ASHE Higher Education Report, 30*(6), 1-99. San Francisco, CA: Jossey-Bass.

Jaschik, S. (2011, January 18). 'Academically Adrift'. *Inside Higher Education.* Retrieved from http://www.insidehighered.com/news/2011/01/18/study_finds_large_numbers_of_college_students_don_t_learn_much

Johnstone, D.B. & Maloney, P.A. (1998).Enhancing the productivity of learning: Curricular implications. *New Directions for Higher Education, 103,* 23-34. doi:10.1002/he.10302

Keith, T. Z. (2015). *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling (2nd ed.).* New York, New York: Routledge.

Kuh, G. D., Kinzie, J., Schuh, J., Whitt, E., & Associates. (2005). *Student success in college: Creating conditions that matter*. San Francisco, CA: Jossey-Bass.

Kuh, G. D., Nelson Laird, T. F., & Umbach, P. D. (2004). Aligning Faculty Activities & Student Behavior: Realizing the Promise of Greater Expectations. *Liberal Education, 90*(4), 24-31.

Kulik, C. L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, *60*(2), 265-299.

Ladson-Billings, G. (1995). But that's just good teaching: The case for culturally relevant pedagogy. *Theory into Practice, 34*(3), 159-165. doi:10.1080/00405849509543675

Lattuca, L. R. & Stark, J. S. (2009). *Shaping the college curriculum: Academic plans in context* (2nd edition). Jossey-Bass.

Lechuga, V. M., Clerc, L. N., & Howell, A. K. (2009). Power, privilege, and learning: Facilitating encountered situations to promote social justice. *Journal of College Student Development, 50*(2), 229-244.

Lohr, S. L. (2009). *Sampling: design and analysis.* Boston, MA: Cengage Learning.

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher, 41*(9), 352-362. doi:10.3102/0013189X12459679

Marton, F., & Säaljö, R. (1976). On qualitative differences in learning—ii Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology, 46*(2), 115-127.

McCormick, A.C. (2011) It's about time: What to make of reported declines in how much college students study. *Liberal Education, 97*(1), 30-39.

McKeachie, W. J. & Kulik, J. (1975).Effective college teaching. *Review of Research in Education, 3,* 165-209.

National Survey of Student Engagement. (2013). *A fresh look at student engagement—Annual results 2013*. Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved from http://nsse.iub.edu/NSSE_2013_Results/pdf/NSSE_2013_Annual_Results.pdf

Nelson Laird, T. F., Seifert, T. A., Pascarella, E. T., Mayhew, M. J., & Blaich, C. F. (2014). Deeply Affecting First-Year Students' Thinking: Deep Approaches to Learning and Three Dimensions of Cognitive Development. *The Journal of Higher Education*, *85*(3), 402-432. doi:10.1353/jhe.2014.0017

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. doi:10.1353/rhe.2014.0003

Neumann, A. & Pallas, A. (forthcoming). *College Teaching Reconsidered: Repairing the heart of a college education* [working title]. Johns Hopkins University Press.

Nordvall, R. C., & Braxton, J. M. (1996). An alternative definition of quality of undergraduate college education: Toward usable knowledge for improvement. *The Journal of Higher Education*, 483-497.

Pascarella, E.T., Blaich, C., Martin, G. L., & Hanson, J. M. (2011). *How robust are the findings of Academically Adrift? Evidence from the Wabash National Study*. *Change: The Magazine of Higher Learning*, *43*(3), 20-24. doi: 10.1080/00091383.2011.568898

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

Perry, W. G., Jr. (1968). *Forms of intellectual and ethical development in the college years: A scheme*. New York, NY: Holt, Rinehart & Winston.

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

Porter, S. R. (2011). Do college student surveys have any validity. *Review of Higher Education, 35*, 45-76. doi:10.1353/rhe.2011.0034

Reason, R. D., Cox, B. E., McIntosh, K., & Terenzini, P. T.(2010, May). *Deep learning as an individual, conditional, and contextual influence on first-year student outcomes*. Paper presented at the Annual Forum of the Association for Institutional Research, Chicago, IL.

Reason, R. D., Cox, B. E., Quaye, B. R. L., & Terenzini, P. T. (2010). Faculty and institutional factors that promote student encounters with difference in first-year courses. *The Review of Higher Education, 33*(3), 391-414.

Renaud, R. D., & Murray, H. G. (2007). The validity of higher-order questions as a process indicator of educational quality. *Research in Higher Education, 48*(3), 319-351. doi:10.1007/s11162-006-9028-1

Rojstaczer, S. & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record, 114*(7), 1-23.

Sanford, N. (1967). *Where colleges fail: A study of the student as a person*. San Francisco, CA: Jossey-Bass.

Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. B. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research, 4*(3), 248-267.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher, 9*(11), 11-16.

Terenzini, P. T. (1999). Research and Practice in Undergraduate Education: And Never the Twain Shall Meet? Higher Education, 38(1), 33-48.

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.

Volkwein, J. F., Lattuca, L.S., Harper, B. J. & Domingo, R. J. (2007). Measuring the impact of professional accreditation on student experiences and learning outcomes. *Research in Higher Education, 48*(2), 251-282

Wang, J.-S., Pascarella, E. T., Nelson Laird, T. F., & Ribera, A. K. (2014). How clear and organized classroom instruction and deep approaches to learning affect growth in critical thinking and need for cognition. *Studies in Higher Education*, doi: 10.1080/03075079.2014.914911

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

Wolf-Wendel, L., Ward, K., & Kinzie, J. (2009). A tangled web of terms: The overlap and unique contribution of involvement, engagement, and integration to understanding college student success. *Journal of College Student Development, 50*(4), 407–428. doi:10.1353/csd.0.0077

Endnote:

[1]The rubrics for the study, additional information about validity testing, and information about observer training and tuning the rubrics are available from the PI, author: author email.

Tables

Table 1

Observed and population course characteristics.

Characteristics	Observed Courses		All Courses	
	N	%	N	%
Class Size[1]				
5-25	46	30.7	1785	68.6
26-50	42	28.0	522	20.1
51-100	28	18.7	166	6.4
>100	34	22.7	128	4.9
Class Level				
Introductory	79	52.7	1268	48.8
Advanced	71	47.3	1333	51.2
Discipline[2]				
Soft	100	69.4	1606	71.3
Hard	44	30.6	646	28.7
Discipline2				
Applied	42	28.2	810	36.8
Pure	102	70.8	1393	63.2
Tenure Line[3]				
Tenure Track	30	58.8	245	38.0
Non-Tenure Track	21	41.2	399	62.0

[1] Smaller classes were purposefully under-sampled to better represent the student experience of courses by taking into account the number of seats in courses.
[2] Courses that were a blend of two disciplinary categories were marked as missing (e.g., applied & pure)
[3] The public institution did not provide data on whether faculty were tenure stream.

Table 2

Sample items from Observation rubrics.

Conceptual Facet		Item	Response Option
Academic Rigor	Cognitive Complexity	The instructor's lecture reflected what level of cognitive processing?	Remember, understand, apply, analyze, evaluate, create
	Standards and Expectations	Standards (the expectations set through words, actions, or norms) for students being prepared for class	**Lowest** (no preparedness was expected), **low, middle, high, highest** (all of the class was expected to be prepared)
Active Learning		NA = Class did not contain class discussions (conversations among students about course material)	**NA**
Cognitively Responsive Teaching	Subject Matter Learning	The instructor created multiple representations of the "core ideas,"	**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"	
	Supportive Learning	The instructor helped students to realize the difference, similarities and sometimes conflict, between prior knowledge and new subject matter ideas.	

Table 3

Illustration of Academic Rigor frameworks in a fictitious course in art history.

Cognitive Complexity	Remember	Understand	Apply	Analyze	Evaluate	Create
	The instructor asks students to match the names of artistic periods with their associated dates.	The instructor asks students to summarize the characteristics of the Romantic period in the arts in their own words.	The instructor teaches students about 3 artistic techniques from the Romantic period and then asks students whether and how these techniques are used in paintings from a local art gallery.	The instructor compares and contrasts two paintings, and then asks students to determine the similarities and differences in the techniques used and the societal influences of the time period.	Students compare and contrast two paintings in terms of how effective they were at representing the social movement of the time period. They are asked to use evidence based on their knowledge of techniques and artistic periods to support their response.	Students present a comprehensive critique of an analysis of the romantic period using evidence, and then suggest, based on prior literature, a new framework to consider in this analysis that would provide different insight into the paintings during that era.
Standards & Expectations	Lowest			Highest		
	• The instructor did not ask students to demonstrate that they understood material from the readings or the prior week's class. • A student could have sat through the class without having read the week's readings or thought about the topic of the class without being noticed. • On the rare occasion, when the instructor paused for "Questions?" No students raised their hands. • There was no norm for participation during class.			• Upon entry to the class, students turned in a short paper about the prior week's readings. • The students immediately sat down and brought out their notes from the readings and questions they had about the prior week's materials. • In the power point presentation, there were activities built in that required all students to participate (such as pair and share). • The instructor facilitated a discussion about the readings, and took note of who participated.		

Table 4. Illustration of Neumann's (2014) framework in a fictitious course in quantitative research methods in psychology.

	Very Ineffective Example	Very Effective Example
Subject Matter Knowledge	• The instructor introduces the definition of research, but does not situate it more broadly in the discipline of psychology. • The instructor does not demonstrate to the class how the core ideas of today's class are inter-related with other classes in the semester • The instructor includes only one perspective on the core ideas • The instructor does not let students explore the limits and constraints of the core ideas for the day's class • For example, the instructor defines the concept of objectivity without letting students explore the assumptions in the field that undergird this concept	• The instructor defines, in depth, core ideas in quantitative research and includes ideas about the assumptions to quantitative methods from the discipline of psychology. • The instructor explains these core ideas in multiple ways and using multiple pedagogies. • The instructor maps out how the course and core ideas fit into psychology as a whole and the interconnections among sub-fields. • The instructor discusses the progression of the course and how the ideas of this class are interconnected and build off other classes during the semester.
Prior Knowledge	• The instructor assumes that the core ideas can be learned similarly by all students. • The instructor uses an example of quantitative research that is not relevant to student's lives. • The instructor presents the core ideas, such as "objectivity," without seeking to understand what students know about this core idea and related concepts from their own lives. • The instructor does not give students an opportunity to consider how their own experiences shape their understanding of research	• Instructor surfaces and explores student's prior knowledge. • For example, the instructor brings forth examples that explore the assumptions of what quantitative research is in Psychology. • Instructor asks students to relate their own epistemological understandings to those of the researchers in the examples. • The instructor might ask students how they came to have these understandings of the meaning of "research" based in their own lives.
Supporting Changing Views	• The instructor does not explore whether student's prior understandings of research conflict with the new core ideas of quantitative methods. • The instructor does not give students space to explore the dissonance they may feel about the conflicting views in their own lives and the course . • The instructor does not support the emotional or intrapersonal features of this dissonance, for example in their math identity—"I am not good at math so I can't do research"). • The instructor does not draw connections (cognitively) between old conceptions from students' lives and new course ideas.	• The instructor notes that some students are grappling with fears of math that are blocking them from exploring a deeper understanding of quantitative methods. • In response to this dissonance, the instructor pulls out part of a student's personal history that shaped their understanding of "math" and connects it with the new assumptions of quantitative methods, thereby supporting the student in moving towards a new understanding.

Table 5

Descriptives for model variables and factors weighted to course population

	Frequency	Mean/SD	Coding	Alpha	Inter-rater ICC
Course Characteristics					
Site	Site 1: 66% Site 2: 34%				
Class Size	5-25: 68.9% 26-50: 19.9% >50: 11.2%				
Length of Class	<= 60 minutes: 16.5% 61-120 minutes: 76.0% >120 minutes: 7.5%				
Class Level	Introductory: 48.9% Advanced: 51.1%				
Discipline	Hard: 24.5% Soft: 75.5% Applied: 29.5% Pure: 70.5%				
Teaching Practices					
Active Learning	Yes: 71.1%				
Cognitively Responsive Teaching		3.11/0.88	1: low; 5: high	.877	.583
Academic Rigor Scales					
Average Cognitive Complexity		2.89/0.90	1: Remember (low) 6: Create (high)	.904	.691
Standards and Expectations		3.62/0.87	1: low; 5: high	.888	.729

Table 6

Item by item descriptives – weighted to course population

Construct/Item	N~	Mean	Std. Dev.
Class Observations			
Highest Cognitive Complexity *	**148**	**3.89**	**0.98**
Instructor's lecture	130	4.07	1.13
Level of handouts or other visual material	111	3.35	1.13
Class activities	53	4.55	1.34
Questions asked by instructor	126	4.18	1.12
Class discussions	65	4.60	1.01
Questions asked by students	120	3.98	1.16
Average Cognitive Complexity *	**149**	**2.89**	**0.90**
Instructor's lecture	128	2.88	1.01
Level of handouts or other visual material	114	2.55	0.93
Class activities	54	3.59	1.23
Questions asked by instructor	126	3.04	1.11
Class discussions	61	3.51	1.20
Questions asked by students	130	2.89	1.03
Standards and Expectations	**150**	**3.62**	**0.87**
Standards set by instructor for students being prepared for class@	144	3.31	1.28
Standards set by instructor for students' class participation (amount)@	147	3,38	1.16
Standards set by instructor for students' class participation (quality—cognitive complexity)*	145	3.21	0.87
Amount of students actively listening, participating, or taking notes	148	4.16	0.95
Amount of students unengaged, bored, or doing something unrelated to the course during class#^	147	1.94	0.97

Notes:
* Response options: 1= remember; 2=understand; 3=apply; 4=analyze; 5=evaluate; 6=create
# Note items were not mutually exclusive: students could have been both actively listening and doing something unrelated to the course; Response options: 1=None or almost none; 2=Few; 3=Half; 4=Most; 5=All or almost all
@Response options: 1=No preparation/participation expected, 2=few students were expected to be prepared/participate; 3=Half were expected to be prepared/participate; 4=Most were expected to be prepared/participate; 5=All were expected to be prepared/participate
^ When scale was created, this variable was reverse coded
~ Lower Ns indicate courses that did not exhibit that specific form of pedagogy (e.g., did not have class discussions, only lecture).

Table 7

Standardized Regression Results

	Variable/Factor	Reference Group	*β*	S.E.	Sig.
Average Cognitive Complexity Regressed on Course Characteristics and Teaching Practices R^2=.236					
Course Characteristics (ΔR^2=.111, p<.05)					
Site	Site 2	Site 1	.035	.180	
Class Size	Medium (26-50)	Small (5-25)	-.021	.187	
	Large (>50)		-.067	.267	
Length of Class	61-120 minutes	Short (<= 60 minutes)	-.057	.251	
	>120 minutes		-.005	.349	
Class Level	Advanced	Introductory	.042	.147	
Discipline	Hard	Soft	-.213	.175	*
	Pure	Applied	.020	.172	
Teaching Practices (ΔR^2=.125, p<.001)					
	Active Learning	No active learning	.175	.171	*
	Cognitively Responsive Teaching Scale	NA	.299	.099	*
Standards and Expectations Regressed on Course Characteristics and Teaching Practices R^2=.588					
Course Characteristics (ΔR^2=.265, p<.001)					
Site	Site 2	Site 1	.014	.124	
Class Size	Medium (26-50)	Small (5-25)	-.288	.128	*
	Large (>51)		-.222	.183	*
Length of Class	61-120 minutes	Short (<= 60 minutes)	-.008	.172	
	>120 minutes		-.041	.240	
Class Level	Advanced	Introductory	-.036	.101	
Discipline	Hard	Soft	.045	.120	
	Pure	Applied	-.085	.118	
Teaching Practices (ΔR^2=.324, p<.001)					
	Active Learning	No active learning	.426	.117	*
	Cognitively Responsive Teaching Scale	NA	.319	.068	*

Table 8

Standardized Regression Results for Larger Courses

	Variable/Factor	Reference Group	*β*	S.E.	Sig.
Cognitive Complexity Regressed on Course Characteristics and Teaching Practices in Larger Courses: R^2=.429					
Course Characteristics (ΔR^2=.035, p>.05)					
Site	Site 2	Site 1	.069	.196	
Length of Class	61-120 minutes >120 minutes	Short (<= 60 minutes)	-.160 .072	.168 .387	
Class Level	Advanced	Introductory	.086	.146	
Discipline	Hard Pure	Soft Applied	-.038 -.106	.148 .167	
Teaching Practices (ΔR^2=.394, p<.001)					
	Active Learning	No active learning	.033	.149	
	Cognitively Responsive Teaching Scale	NA	.642	.088	*
Standards and Expectations Regressed on Course Characteristics and Teaching Practices in Larger Courses: R2=.619					
Course Characteristics (ΔR^2=.080, p>.05)					
Site	Site 2	Site 1	-.131	.150	
Length of Class	61-120 minutes >120 minutes	Short (<= 60 minutes)	.066 -.064	.129 .296	
Class Level	Advanced	Introductory	.168	.112	
Discipline	Hard Pure	Soft Applied	.072 .068	.113 .128	
Teaching Practices (ΔR^2=.539, p<.001)					
	Active Learning	No active learning	.215	.113	*
	Cognitively Responsive Teaching Scale	NA	.681	.067	*

Table 9

Percentile points above comparison group or above the mean (per unit increase) according to standardized betas.

	All Courses		Larger Courses	
	Cognitive Complexity	Standards and Expectations	Cognitive Complexity	Standards and Expectations
Discipline (Hard v. Soft)	-8.4	Not Sig.	Not Sig.	Not Sig.
Class Size (v. small):				
Medium	Not Sig.	-11.3	NA	NA
Large	Not Sig.	-8.7	NA	NA
Active Learning	6.9	16.4	Not Sig.	8.5
Cognitively Responsive Teaching	11.7	12.5	23.9	25.2

Figure

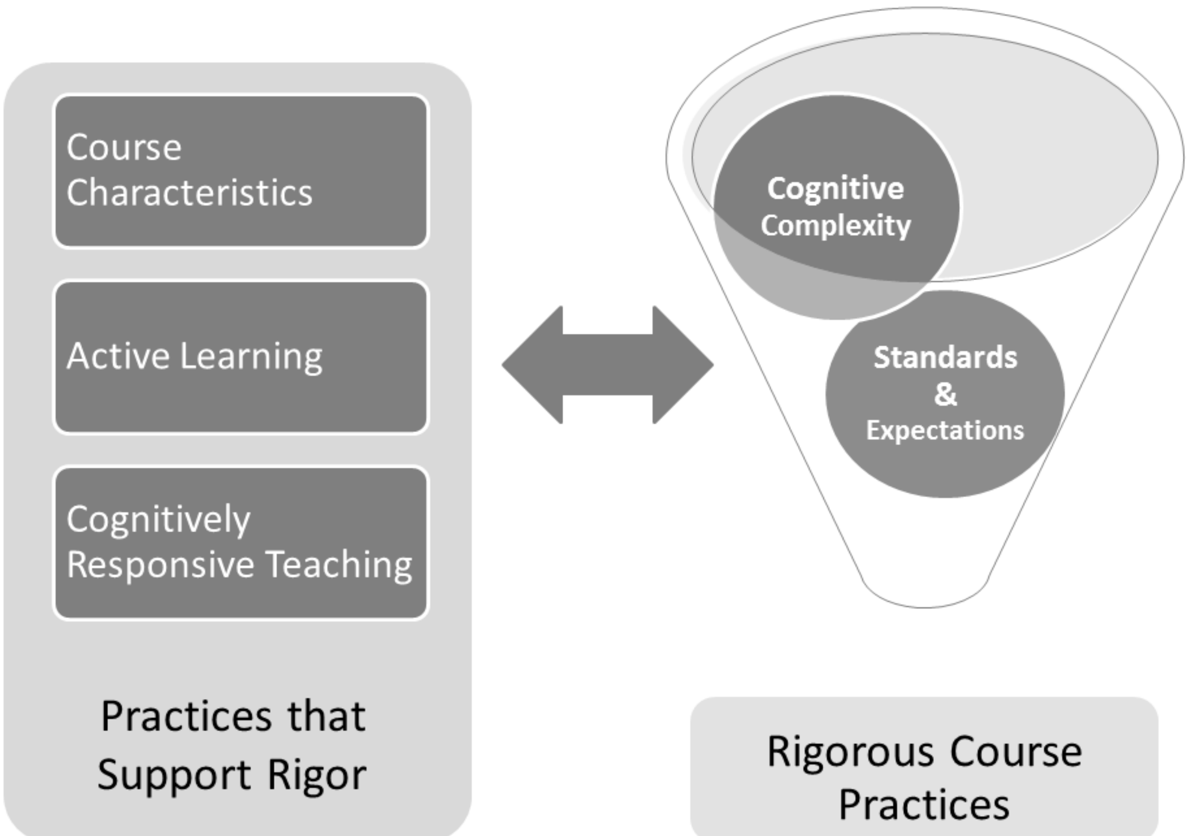

Figure 1. Conceptual Model