Form for submitting extended proposals
for consideration for the
2018 ROBERT J. MENGES AWARD FOR OUTSTANDING RESEARCH IN EDUCATIONAL DEVELOPMENT

Instructions:

- Boxes will expand to accommodate text.
- Total **word count must not exceed 2000 words** for the body of the proposal, excluding references and appendices.
- Supporting information (e.g. tables, figures, images, references, instruments, details of experimental design) may be placed in appendices. Though not limited, the strongest proposals are typically supported by no more than 10 pages of appendices. To conserve space, for example, you can place multiple figures on one page, single-space survey instruments, etc. Keep in mind, the selection committee is not required to read beyond this general limit.
- Be sure to include the word counts in each section, as well as the total for all sections (see below). Proposals without the word counts noted will not be read.
- Incomplete proposals will not be read.
- As a final step, **"blind" your proposal** by removing any direct references to you, your co-authors, institution, and supporting publications. Be sure to blind all parts of your proposal, including appendices.
- Send your proposal to the committee chair in **MS Word format**. For consistency, it is helpful if you use calibri, 11 pt font.

RESEARCHERS NAMES (please indicate primary contact with a *):	INSTITUTION(s):	EMAIL ADDRESS(ES):
Lindsay B. Wheeler*	University of Virginia	lsb4u@virginia.edu
Dorothe Bach	University of Virginia	djb4d@virginia.edu
Add lines as necessary.		

Please leave the rest of this page blank.

SESSION TITLE:	Making Assessment Matter: Linking Interventions, Instructional Practices, and Academic Achievement

1. RESEARCH QUESTION(S) & WHY THEY ARE IMPORTANT TO THE FIELD:

Tracing the impact of educational development from intervention to student learning has historically been difficult (Beach, Sorcinelli, Austin, & Rivard, 2016; Chism, Holley, & Harris, 2012; Hines, 2009, 2011). More recently, however, a growing number of studies are being designed that attempt to link participation in Center for Teaching and Learning (CTL) interventions to changes in student outcomes (e.g., Condon, Iverson, Manduca, Rutz, & Willett, 2016; Meizlish, Wright, Howard, & Kaplan, 2017).

This study contributes to these efforts by examining the impact of educational development interventions at one university along two dimensions of Kreber and Brook's six-step impact model (2011): Participants' teaching performance and student learning. Like Meizlish, et al. (2017), we employed a quasi-experimental study design through which we compared groups of Science, Technology, Engineering and Math (STEM) instructors that differed in their level of participation in the local CTL's instructional development initiatives. Whereas most educational development studies make inferences from self-reported data (e.g. surveys of instructors' beliefs and self-efficacy, student evaluations of teaching), our study is unique in its combination of several direct measures. Specifically, we assessed instructors' teaching practices through an analysis of course syllabi and classroom observations, and leveraged institutional grade data to link teaching practices to student success (Figure 1). In addition, the study contributes to the research on a pressing higher education issue: the success of underrepresented minority (URM) students in STEM fields.

The research questions driving the study include:

1. What differences, if any, exist in the instructional practices and student success for courses taught by instructors who have and have not participated in CTE interventions?
2. What differences, if any, exist between courses taught by new faculty who have and have not participated in a New Faculty Program (NFP)?
3. What differences, if any, exist between courses taught in a department where a cohort of instructors have and have not engaged in a STEM Program (STEMP)?

WORD COUNT
314

2. DESCRIPTION OF RESEARCH DESIGN:

This quantitative study was conducted at a mid-Atlantic, research-intensive university during the 2016-2017 academic year. Additional details about the study design and referenced tables are in the Appendix.

Context

Instructors who participated in the study were grouped into these four categories depending on their level of engagement with CTL activities:

- *No engagement* – instructors who have not engaged in any CTL programming.
- *CDI* – instructors who completed a week-long Course Design Institute (35 contact hours) and designed a learning-focused course.

- *NFP* – instructors who completed a program for new faculty in their first three years at the institution that included CDI and a semester- or year-long Faculty Learning Community (FLC). The NFP FLC included a half-day retreat, seven 90 minute meetings facilitated by CTL staff, and an individual teaching consultation (22 additional contact hours; 57 contact hours total). Although NFP is open to faculty from all disciplines, the present study includes only STEM faculty who participated in NFP in 2015-2017.
- *STEMP*–instructors who completed a program for STEM faculty teaching large-enrolment courses that included CDI and a year-long FLC. The STEMP FLC included seven 60 minute meetings facilitated by CTL staff, peer-observations of other STEMP participants, and an individual consultation with CTE faculty (14 additional contact hours; 45 contact hours total). In the present study, the assessment of STEMP focuses on a group of six faculty from a department of 16 (37.5%) who participated as a cohort in 2015-2016 (Department A).

Participants

A total of 150 STEM instructors participated in the study. They ranged in their instructor status (e.g., graduate student, full professor) and participation in CTE interventions (i.e., none, CDI, STEMP, NFP) (Table 1). A total of 239 STEM undergraduate courses across 19 departments comprised the data set. They were representative of the 1005 total undergraduate STEM courses offered during the 2016-2017 academic year (Table 2). We compared Department A (STEMP cohort) to Department B, a comparable 'control' department (Table 3).

Data collection

To measure instructional practices, we collected:

- *Classroom observation data* - Observational data was collected using the Classroom Observation Protocol for Undergraduate STEM (COPUS) (Smith, Jones, Gilbert, & Wieman, 2011). COPUS captures the presence or absence of 26 different student and instructor behaviors in 2-minute time increments. Trained undergraduate students observed each of the 239 courses twice during the semester. From the individual COPUS behaviors, a COPUS profile value was calculated (copusprofiles.org; Stains et al., 2018), ranging from 1 (didactic instruction) to 7 (student-centered instruction). These profiles were then collapsed into didactic (1), interactive (2), and student-centered (3) clusters to show a hierarchical range of active learning present (Table 4).
- *Course syllabi* - We obtained 196 syllabi and used a validated rubric (Authors, 2014) to assess the degree to which a syllabus achieves a learning orientation. Two trained graduate students scored each syllabus placing them on a spectrum from learning-focused (a score of 46) or content-focused (a score of 0), which was then collapsed into content-focused (0-16), transitional (17-30), and learning-focused (31-46) clusters.

To measure student success, we collected:

- *Student grade data* – We obtained counts of students who received As, Bs, Cs, D/F/withdraw (DFW) for each observed course, disaggregated by gender and race. Black and Hispanic students were grouped into the category 'Underrepresented Minority' (URM). A % DFW rate and % A rate were calculated for each course (e.g., number DFWs/total number of students x 100). We also obtained grade data for all courses in Departments A, which contained a cohort of six faculty who participated in STEMP, pre- and post-STEMP program.

To account for differences across courses, we collected:

- *Instructor/course information and CTL participation* - We gathered information that included which CTL intervention instructors participated in, instructor status, course level, and course size.

Data analysis

To better understand course level outcomes, we selected participation in the four different CTL interventions as our independent variable and COPUS profile score, syllabus scores, % grade rates for students and subgroups of students as our dependent variables. We explored differences in instructor and course information for each independent variable group (Table 5).

To answer *research question 1,* we ran an ANCOVA to explore differences in COPUS profiles and URM % DFW rates for CTL intervention groups while controlling for number of enrolled students and level of course (all assumptions met). We ran ANOVAs to explore differences in syllabus scores and t-tests to explore differences between URM and White % DFW for CTL intervention groups.

To answer *research question 2,* we ran non-parametric Mann-Whitney U tests to compare differences in outcomes for courses taught by NFP participants (n=35) and courses taught by new faculty with no CTL engagement (n=29). We also ensured courses in each group were not significantly different in size and level.

To answer *research question 3,* we ran non-parametric Mann-Whitney U tests to compare differences in observations of courses in Department A (n=17) and Department B (n=29). Grade data for all courses in Department A were compared pre- (2014-2015) and post- (2016-2017) STEMP participation.

WORD COUNT
823

3. LITERATURE REVIEW & THE RELATIONSHIP OF THE LITERATURE TO YOUR RESEARCH QUESTION(S):

This study draws on and makes a contribution to two strands of research: 1. It contributes to current efforts to trace the impact of educational development initiatives from intervention to student outcomes. 2. It adds to a body of research examining the achievement gap for underrepresented minority students in STEM.

1. Educational developers are increasingly asked to demonstrate the efficacy of their work and to provide evidence that the investment in CTLs is worthwhile. This study contributes to research focused on establishing the link between participation in education development, improved teaching and student learning (Condon et al., 2016). Specifically, drawing on research on intensive course design workshops (Authors, 2016) and faculty learning communities (Gehrke & Kezar, 2017), it examines the impact of different CTL intervention types and durations. Unique in its design, this study uses multi-indicator strategy and direct measures such as syllabi and classroom observations to assess changes in teaching performance as well as institutional grade data to approximate student success. It builds on and moves beyond educational development research that relies on

indirect measures such as student evaluations of teaching and reported learning gains (e.g., Condon et al., 2016; Meizlish et al., 2017).

2. Achievement gaps between URM students and non-URM students in STEM courses are well documented in the literature (e.g.; National Science Foundation, 2017), and attrition from STEM courses have been attributed to limited student engagement (e.g., Gasiewski, Eagan, Garcia, Hurtado, & Chang, 2012; Daempfle, 2003). Incorporation of active and collaborative learning can positively impact cognitive and affective outcomes, success for students in STEM courses (e.g., Cooper, Cox, Nammouz, Case, & Stevens, 2008; Freeman et al., 2014), and reduce the achievement gap (Haak, HilleRisLambers, Pitre, & Freeman, 2011). However the majority of these studies also utilize self-report data to identify whether a course engages students in active learning. The present study uses observational data to further understand the ways in which instructional practices may impact student success, particularly for URM students.

WORD COUNT
326

4. FINDINGS, INCLUDING THEIR SIGNIFICANCE & LIMITATIONS:

Below we present the main findings for each research question, and outline the significance of this work. We also note where results are descriptive and cannot be generalized to the population. Further details on results, implications, and limitations can be found in the Appendix.

RQ1: Overall impact of CTL interventions

- Courses taught by instructors who completed either STEMP or NFP exhibited significantly more student-centered instruction compared to courses taught by instructors with no engagement in CTL programs, even when controlling for class size and level (Table 7).
- Syllabi of courses taught by instructors who engaged with one of our three CTL interventions were significantly more learning-focused compared to syllabi for courses taught by non-CTL engaged instructors.
- Courses taught by NFP and STEMP participants appear to be closing a commonly reported failure gap, particularly when looking at smaller courses (n<60) (Figure 2).
- Descriptively, URM student % DFW rates are consistent among courses classified as didactic (>80% lecture), regardless of instructor engagement in CTL programs. In courses classified as student-centered, URM and White student % DFW gap closes for STEMP (M=~3%) and NFP faculty (M=~7%) while URM % DFW rates for non-engaged participants were nearly quadruple White % DFW rates (Figure 3).

RQ2: NFP program

- Courses taught by new faculty who participated in NFP exhibited significantly more student-centered instruction and more learning-focused syllabi compared to courses taught by non-CTL engaged instructors (Table 8, Figure 4, 5).
- Descriptively, small courses (n<60) taught by NFP participants are more interactive and have much smaller % DFW rates when compared to courses taught by new faculty who have not participated in any CTL interventions (Figure 6).

RQ3: STEMP program

- Observed courses in Department A (STEMP cohort) had significantly more student-centered instruction and more learning-focused syllabi compared to observed courses in Department B (non-CTL engaged) (Table 9).
- Not only are the % DFW rates significantly lower for Department A (Table 9), but the difference between URM and White student % DFW rates are virtually non-existent for Department A (Figure 7).
- Descriptively, the overall department % DFW rate for URM students nearly halves after a cohort of faculty in Department A participated in STEMP (Figure 8).

Implications:

1. Prolonged support for faculty, particularly for new faculty and departmental cohorts, appear to have positive impacts on instructional practices. In some cases these changes in instructional practices appear to be linked to reduced failure rates and a closing of failure gaps between URM and White students.
2. Instructors without support in implementing active learning strategies may be doing more harm than good, particularly for URM students.
3. Even with a sample of 239 courses, only descriptive statistics could be determined as sub-group samples were small. But the results are promising. They clearly show the potential for educational developers to demonstrate CTL impact through a research design that leverages teaching observation and institutional student success data. Our study calls for a similarly systematic and rigorous, larger-scale, cross-institutional research project that will allow us to further substantiate the link between intervention, instructor practice and success for all students.

WORD COUNT
506

TOTAL WORD COUNT FOR ALL 4 SECTIONS (MAX 2000 WORDS):	**1969**

REFERENCES

Authors (2014).
Authors (2015).
Authors (2016).
Authors (in preparation).
Authors (in preparation).
Bain, K. (2004). What makes great teachers great. *Chronicle of Higher Education*, *50*(31), B7-B9.
Beach, A. L., Sorcinelli, M. D., Austin, A. E., & Rivard, J. K. (2016). *Faculty development in the age of evidence: Current practices, future imperatives*. Sterling, VA: Stylus.
Chang, M.J., Cerna, O. Han, J., & Sàenz, V. (2008). The Contradictory Roles of Institutional Status in Retaining Underrepresented Minorities in Biomedical and Behavioral Science Majors. *The Review of Higher Education*, *31*(4), 433–464. https://doi.org/10.1353/rhe.0.0011
Chism, N. V. N., Holley, M., & Harris, C. J. (2012). Researching the impact of educational development: Basis for informed practice. *To Improve the Academy, 31*, 129–145.
Condon, W., Iverson, E. R., Manduca, C. A., Rutz, C., & Willett, G. (2016). *Faculty development and student learning: Assessing the connections*. Bloomington: Indiana University Press.

Cooper, K. M., Downing, V. R., & Brownell, S. E. (2018). The influence of active learning practices on student anxiety in large-enrollment college science classrooms. *International Journal of STEM Education*, *5*(1), 23. https://doi.org/10.1186/s40594-018-0123-6

Cooper M. M., Cox Jr, C. T., Nammouz M., Case E. and Stevens R., (2008), An assessment of the effect of collaborative groups on students' problem-solving strategies and abilities, *Journal of Chemistry Education, 85*(6), 866.

Daempfle, P. A. (2003). An analysis of the high attrition rates among first year college science, math, and engineering majors. *Journal of College Student Retention: Research, Theory & Practice*, *5*(1), 37-52.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *BioScience*, *61*(7), 550–558. https://doi.org/10.1525/bio.2011.61.7.9

Fink, L. D. (2013). *Creating Significant Learning Experiences, Revised and Updated. An Integrated Approach to Designing College Courses*. San Francisco, CA: Jossey-Bass.

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(23), 8410-8415.

Gasiewski, J. a., Eagan, M. K., Garcia, G. a., Hurtado, S., & Chang, M. J. (2012). From Gatekeeping to Engagement: A Multicontextual, Mixed Method Study of Student Academic Engagement in Introductory STEM Courses. *Research in Higher Education*, *53*(2), 229–261. https://doi.org/10.1007/s11162-011-9247-y

Gehrke, S., & Kezar, A. (2017). The roles of STEM faculty communities of practice in institutional and departmental reform in higher education. *American Educational Research Journal*, *54*(5), 803-833.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*(6034), 1213-1216.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984. https://doi.org/10.1002/tea.20439

Hines, S. R. (2009). Investigating faculty development program assessment practices: What's being done and how can it be improved? *Journal of Faculty Development, 23*, 5–19.

Hines, S. R. (2011). How mature teaching and learning centers evaluate their services. *To Improve the Academy, 30*, 277–289.

Kreber, C., & Brook, P. (2001). Impact evaluation of educational development programmes. *International Journal for Academic Development*, 6(2), 96-108

Meizlish, D.S., Wright, M.C., Howard, J., & Kaplan, M.L. (2017): Measuring the impact of a new faculty program using institutional data, International Journal for Academic Development

National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees systemic change to support diverse student pathways*. Washington, DC: The National Academies Press.

National Science Foundation, National Center for Science and Engineering Statistics, (2017), Women, minorities, and persons with disabilities in science and engineering: 2017, *Special Report NSF 17-310*, Arlington, VA, available at www.nsf.gov/statistics/wmpd/.

President's Council of Advisors on Science and Technology (PCAST) (2012). *Engage to Excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*.

Schunk, D. H., Pintrich, P. R. & Meece, J. R. (2007). *Motivation in Education: Theory, Research, and Applications* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Seymour, E., & Hewitt, N. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview Press.
Smith, M. K., Jones, F. H. M., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate stem (COPUS): A new instrument to characterize university STEM classroom practices. *CBE Life Sciences Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154
Snyder, J. J., Sloane, J. D., Dunk, R. D. P., & Wiles, J. R. (2016). Peer-Led Team Learning Helps Minority Students Succeed. *PLoS Biology*, *14*(3), 1–7. https://doi.org/10.1371/journal.pbio.1002398
Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., … Young, A. M. (2018). Anatomy of STEM teaching in North American universities. *Science*, *359*(6383), 1468–1470. https://doi.org/10.1126/science.aap8892
Svinicki, M. D. (2004). *Learning and Motivation in the Postsecondary Classroom*. Bolton, MA: Anker.
Wiggins, G. & McTighe, J. (2005). *Understanding by Design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
Winkelmes, M. A., Bernacki, M., Butler, J., Zochowski, M., Golanics, J., & Weavil, K. H. (2016). A teaching intervention that increases underserved college students' success. Peer Review, 18(1/2), 31.

Appendices: Supporting information (e.g. tables, figures, images, references, instruments, details of experimental design) may be placed in appendices. Though not limited, the strongest proposals are typically supported by no more than 10 pages of appendices. To conserve space, for example, you can place multiple figures on one page, single-space survey instruments, etc. Keep in mind, the selection committee is not required to read beyond this general limit.

MENGES PROPOSAL APPENDIX

Section 1: Research questions

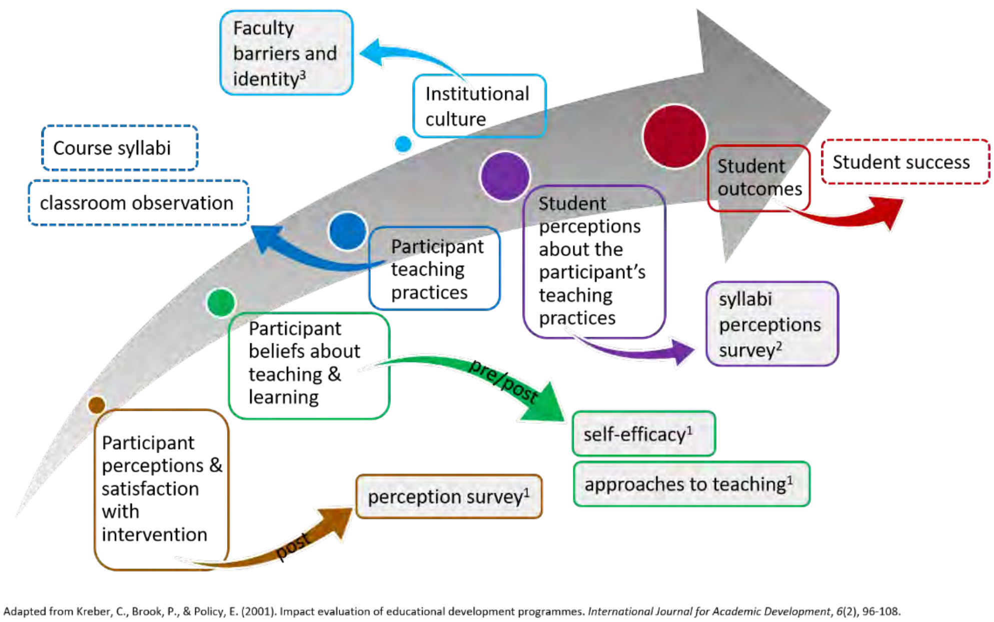

Figure 1. Modified Impact Model to Measure CTL Interventions. Dotted boxes indicate the components of the present study, while greyed boxes indicate prior research conducted in our CTL. [1] Authors, 2014; Authors, in preparation; [2] Authors, 2015; [3] Authors, in preparation)

Section 2: Description of Research Design

Context

Course Design Institute (CDI). Course design interventions are a widely adopted in educational development. Our university's CDI is a week-long intensive workshop (35 contact hours) that, like many others, draws on ideas from backward and integrated course design (Fink, 2013; Wiggins & McTighe, 2005), educative assessment (Huba & Freed, 2000; Wiggins, 1998), active learning (Bonwell & Eison, 1991), student motivation (Schunk, Pintrich & Meece, 2007; Svinicki, 2004), and transparent assignments (Winkelmes et.al. 2016) to support faculty in designing learning-focused course. At the time of the present study, 508 individuals have participated in CDI, 144 of which were STEM instructors.

During CDI, faculty produce promising course syllabi (Bain, 2004) that compellingly communicate to students why they should care about material (relevance), what they will learn (goals and objectives), what students will do to achieve the objectives (transparent assessments and assignments aligned with objectives), and how they will be supported to be successful in the course (environment, inclusivity). Throughout the week, faculty engage in small groups and are supported by a trained facilitator. Faculty also have opportunities to meet one-on-one with other faculty, facilitators, and trained undergraduate students to receive feedback on their course design and syllabus. Prior research on our CDI demonstrates its positive impact on instructors' self-reported understandings and confidence and demonstrated shifts in course syllabi to more learning focused following CDI (Authors, 2016). The present study aims to extend this research and compare CDI participants to those who have not engaged in CTL interventions.

New Faculty Program (NFP). Programs focused on new faculty are among services most frequently offered by CTL (Beach, Sorcinelli, Austin, & Rivard, 2016) and research suggests that efforts focused on future and new faculty have the potential of being more impactful than those aiming to change the practices of more established colleagues (e.g., Ebert-May et al., 2011). Like at many other institutions, the goal of our NFP program is to help a new generation of faculty adopt evidence-based teaching practices and become reflective practitioners. Faculty enrolled in the NFP program first complete the CDI described above (35 hours). They then participate in a semester- or year-long FLC consisting of a half-day retreat, seven 90-minute meetings facilitated by CTE staff, and an individual teaching consultation (22 hours).

Unlike other early-career faculty programs that focus on instructional development though introductory workshops on a range of different and often disconnected teaching topics, the NFP FLC is tightly focuses on supporting faculty in implementing the learning-focused courses they designed during the intensive CDI. The NFP FLC engages participants in a cycle of deliberate classroom experimentation, analysis of and reflection about their experience, and gives them the opportunity to provide and receive peer- and expert feedback before implementing a new teaching strategy. Meizlish, Wright, Howard, & Kaplan (2017) speculate that programs with more extended emphasis on course design, might be anticipated to have more impact. The present study focused on a subset of NFP faculty from the 2015-2017 cohorts who teach in STEM disciplines to provide evidence to support, or refute, the speculations of Meizlish et al. (2017).

STEM Program (STEMP). Calls for reform in undergraduate STEM education (e.g., National Academies of Sciences, Engineering, and Medicine, 2016; President's Council of Advisors on Science and Technology [PCAST], 2012;) demonstrate the need for educational development programs focused on STEM faculty, particularly for instructors teaching large-enrollment courses (e.g., Seymour & Hewitt, 1997). The goal of our STEMP program is to support STEM faculty teaching large-enrolment courses implement evidence-based instruction in their classes. Engagement in STEMP included CDI (35 hours) and a year-long FLC (14 hours). The STEMP FLC involved seven 60-minute meetings facilitated by CTL staff, peer-observations of other STEMP participants, and an individual consultation with CTE faculty. During the FLC, members could learn from and encourage one another in the redesign and administration of their new "learner-centered" courses. Topics and literature concerning learner-centered pedagogy were often discussed, and members of the cohort were able to discuss their own questions and experiences as part of the process.

While other professional development programs have engaged faculty from a single department (e.g., Ebert-May et al., 2011), we embedded a cohort of six faculty from the same department within the 2015-2016 STEMP FLC community (comprised of 10 total faculty across four STEM departments). This allowed for development of their department-level community of practice but also provided opportunities to engage with and develop cross-disciplinary relationships, which research suggests is one effective change strategy (Henderson, Beach & Finkelstein, 2011). For the STEMP program, these six non-tenure track faculty from Department A were the focus for assessing STEMP in the present study.

Participants and recruitment

All instructors-of-record who taught the 1006 undergraduate STEM lecture or seminar course during the fall 2016 and spring 2017 semesters were emailed to voluntarily participate in the IRB-approved study. The goal for recruitment was to obtain a representative sample of instructors and courses across all STEM disciplines in order to be able to generalize the claims made from the study to the entire undergraduate STEM lecture/seminar course population at our university. For example, if 10% of all undergraduate STEM lecture and seminar courses were chemistry courses, we wanted our sample to include approximately 10% chemistry courses. After obtaining an initial set of volunteers, we calculated the percent of courses in each department and level (e.g., introductory, upper-level) to determine where our sample was over- or under-represented based on the total population. Our samples were well-matched on course level each semester, and we contacted chairs for departments that were under-represented to request they send out our recruitment email to instructors in their department.

Participants included 150 of the 447[1] instructors (33.6%) who ranged in their instructor status (e.g., graduate student, full professor) and in their engagement in the various Center interventions (i.e., none, CDI, NFP, STEMP). These participants were fairly representative of the total population of instructors for these variables (Table 1). Courses included 123 of the 522 total courses taught in the fall 2016 semester (23.6%) and 116 of the 483 total courses taught in the spring 2017 semester (24.0%). A total of 239 of the 1005 courses (23.8%) across 19 STEM departments were included in the data set and were representative of the percent of lecture/seminar courses offered in each department and at each level (Table 2).

Table 1. *Instructor Demographics for Sample and Population of STEM Undergraduate Courses*

		Sample, n (%)	Population, n (%)
Intervention	none	140 (58.6)	668 (66.5)
	CDI	28 (11.7)	192 (19.1)
	STEMP	36 (15.1)	71 (7.1)
	NFP	35 (14.6)	74 (7.4)
Instructor Type	Graduate student	12 (5.0)	68 (6.8)
	Non-tenure track	78 (32.6)	383 (38.1)
	Tenure track Assistant Professor	39 (16.3)	94 (9.4)
	Tenured Associate Professor	54 (22.6)	125 (12.44)
	Full Professor	56 (23.4)	273 (27.2)
	Unknown	---	62 (6.2)

Note: Because the unit of analysis in the study is the course, the description of the instructors is based on the total number of courses. Therefore, if an instructor taught multiple courses, the instructor is counted as many times as they taught.

Table 2. *Course Details for Sample and Population of STEM Undergraduate Courses*

Science & Math Departments	Sample, n (%)	Population, n (%)	Engineering Departments	Sample, n (%)	Population, n (%)
Astronomy	4 (1.7)	27 (2.7)	Applied Math	17 (7.1)	92 (9.2)
Biology	21 (8.8)	51 (5.1)	Biomedical	9 (3.8)	32 (3.2)
Chemistry	14 (5.9)	58 (5.8)	Civil	9 (3.8)	22 (2.2)
Environmental Sci	3 (1.3)	64 (6.4)	Chemical	4 (1.7)	20 (2.0)
Mathematics	31 (13.0)	123 (12.3)	Computer Science	34 (14.2)	77 (7.7)
Physics	19 (7.9)	39 (3.9)	Electrochemical	8 (3.3)	27 (2.7)
Psychology	13 (5.4)	63 (6.3)	Engr & Society	20 (8.4)	113 (11.2)
Sociology	9 (3.8)	67 (6.7)	Mechanical	5 (2.1)	35 (3.5)
Statistics	5 (2.1)	38 (3.8)	Materials Science	8 (3.3)	24 (2.4)
			Systems	6 (2.5)	33 (3.3)
STEM class level	Sample, n (%)	Population, n (%)	STEM class level	Sample, n (%)	Population, n (%)
100	59 (24.7)	221 (22.0)	300	53 (22.2)	260 (25.9)
200	65 (27.2)	229 (22.8)	400	62 (25.9)	295 (29.4)

To assess the STEMP program, we focused on Department A, where a subset of six instructors had participated in STEMP. These instructors comprise 37.5% of the faculty (n=16) who taught in the

[1] The sum of fall and spring semester instructors listed above is larger than the total number of instructors as instructors taught multiple courses in the same semester or taught both semesters.

department during the 2016-2017 academic year when observations were conducted. We also identified a comparable 'control' department, Department B, where none of the observed participants (n=16 of 51 total Department B instructors, 31.4% participation) engaged in any CTL interventions. Department A and B were similar in disciplinary content, % URM, and % White students (Table 3); however, they differed in the number of courses at various levels (e.g., lower, upper) as well as the dispersion of instructors of different status. The majority of instructors in Department A were non-tenure track faculty teaching across all course levels, while the majority of instructors in Department B were graduate students teaching in lower-level courses and tenure track/tenured teaching in upper-level courses[2].

Table 3. *Comparison of Department A (STEMP cohort) and Department B (non-engaged department)*

Student make-up, M (SD)	Department A, n=70	Department B, n=120
Total students, Mean (SD)	40.71 (11.41)	35.96 (9.24)
% URM, Mean (SD)	9.63 (6.84)	12.70 (8.33)
% White, Mean (SD)	55.93 (11.24)	53.35 (11.00)

Data collection and coding

Data were collected from courses taught by participating instructors, all of whom consented to IRB, during the fall 2016 and spring 2017 semesters. Data sources included classroom observations, course syllabi, student grade data, and instructor information. From the data collected, we coded/scored each component individually to create a data point for each outcome measure for each individual course. Both the data collection and coding procedures are described below.

Observations. We collected observational data using the Classroom Observation Protocol for Undergraduate STEM (COPUS) (Smith, Jones, Gilbert, & Wieman, 2013), which captures the presence or absence of 26 different student and instructor behaviors in 2-minute time increments. Undergraduate students conducted the classroom observations using COPUS. They received 7 hours of training on COPUS, which included coding and discussing 6 video-taped STEM classrooms, a live observation, and a final video to obtain reliability. All observers were reliable on the final training video prior to collecting observational data (all κs >.8). Each course was scheduled to be observed twice, each during a 4-week time block (weeks 3-7 and weeks 8-12 of the semester) at a time that each instructor chose to represent typical instruction. Days that included exams or non-typical activities, as reported by the instructor, were avoided. A total of 451 observations were conducted by a single or set of paired observers, averaging 1.89 observations for each course[3]. Because of scheduling conflicts, some classes were observed only once.

The COPUS behaviors for individual observations were first converted into a percent of total time (number of times coded/total 2-minute increments in the observation x 100). All observations for a course received an average percent of total time for each of the 26 COPUS behaviors[4], and from these individual behaviors, a COPUS profile value was obtained (copusprofiles.org) (Stains et al., 2018). These COPUS profiles ranged from 1 (didactic instruction) to 7 (student-centered instruction), which were then

[2] Differences in course level and instructor status across the compared departments may be a confounding factor in the data. We address these differences in the analysis section below.
[3] Stains et al. (2018) suggests four observations are needed to characterize classroom instruction. We made the decision to sample more classes for fewer observation data points and understand this is a limitation of the study.
[4] We random checked inter-rater reliability on a 10% subset of paired observations to ensure we could then collapse these data. All sampled pairs were reliable (all κs >.7), with the exception of two observations. After contacting the observers, the discrepancies were resolved.

collapsed into didactic (1), interactive (2), and student-centered (3) to demonstrate a hierarchical range of active learning present (Table 4).

Table 4. *Description of COPUS Profiles Used to Characterize Classroom Instruction*

Grouping	Profile #	Description
Didactic	1	Greater than 80% lecture with very few student questions
	2	Greater than 80% lecture with few clicker questions and group work
Interactive	3	Lecture that includes group activities
	4	Lecture that includes group work associated with clicker questions
Student-centered	5	Group work activities consistently used throughout class
	6	Process Oriented Guided Inquiry Learning (POGIL)-like instruction with group worksheets and one-on-one instructor-student interactions
	7	Varied group work activities used throughout class

Course syllabi. We also requested course syllabi for the 239 courses observed, and a total of 196 number of syllabi were collected (82%). A small number of instructors (n=17, 11.3%) taught multiple sections of the same course, which were included in our sample, and the syllabus for the course was associated with these multiple sections. A previously developed and validated syllabus rubric was used to score each syllabus (Authors, 2014). The rubric includes four main criteria to assess the learning-focus of the syllabus: 1) learning goals and objectives, 2) assessment activities, 3) schedule, and 4) overall learning environment. The thirteen components within these criteria are scored as strong, moderate, and weak and are used to calculate a score ranging from learning-focused (a score of 46) or content-focused (a score of 0).

Two graduate student researchers trained on the syllabus rubric scored all syllabi. Training included a 2-hour meeting, followed by trainees' scoring two rounds of 10 syllabi each, discussing coding results with the first author, and reaching consensus on all training syllabus coding prior to scoring the syllabi data for the study. All 196 syllabi were scored by each researcher individually, and any discrepancies in the individual coding were resolved upon discussion. The first author coded a 10% subset of these syllabi to compare her scores with the final component scores of the graduate student researchers. Any major discrepancies were discussed and resolved. From the syllabus scores, each course received a score indicating whether the syllabus was content-focused (score less than 17=1), transitional (scores between 17-30=2), and learning-focused (scores greater than 30=3).

Student grade data. We obtained student grade data from our institutional assessment center for all observed courses[5]. These grade data included the total number of students enrolled in the course, which were used to determine class size for each course. We also obtained counts of students who received As, Bs, Cs, D/F/withdraw (DFW)[6] for each course. These counts were also obtained disaggregated by gender and race. The number of African American and Hispanic students who received a particular grade were combined into a category labeled 'Underrepresented Minority' (URM)[7]. A % DFW rate and % A rate were calculate for each course overall (e.g., number DFWs/total number of students x 100). These rates were also calculated for students identified as male, female, White, and URM (e.g., number URM As/total number of URM x 100). Of the 239 courses observed, 23 (9.6%) of the courses did not have URM students, which reduced the sample size when URM grade rates were used in the analysis. To understand the impact of a cohort of STEMP faculty on a department, student grade

[5] These data were anonymized at the student level and were obtained by the researchers as grade counts.
[6] We use DFW and failure synonymously but are aware that withdrawals are not necessarily considered as such.
[7] Research typically includes Native American in the URM category (Chang et al., 2008); however, in our data set, Native American students were ~0.01% of the sample. Therefore we chose to exclude them from this group.

data were obtained for all courses in Department A (STEMP cohort department) for the year before (2014-2015) and after (2016-2017) participants engaged in STEMP.

Instructor information. We gathered information on each instructor that included which CTL intervention(s) they had participated in as well as instructor status. Given the hierarchical nature of the intensiveness of our interventions as described above, CTL intervention status was coded as none (0), CDI (1), STEMP (2), and NFP (3). We obtained instructor status and hiring date from our institutional assessment center and triangulated these information from university websites to identify new faculty (hired since 2013). Instructor status included graduate students (0), non-tenure track instructors and teaching faculty (1), tenure track assistant professors (2), tenured associate professors (3) and full professors (4).

Data analysis

From the data collection and coding, each course was associated with a set of data from the observation, syllabus, grade data, and instructor information (Table 5). Then these data were analyzed to answer the research questions. The independent variable was Center intervention status, and the dependent variables examined were COPUS profile score, syllabus scores, % grade rates for students and subgroups of students. Of the 239 courses observed, 24 (10%) did not have any URM students. When % DFW rates were included in the analyses, the courses with no URM students were excluded from the data set so as to not skew the data.

Table 5. *Example Set of Data for an Individual Course*

	Course	COPUS Group	Syllabus score	Syllabus code	Class size	Size code	%DFW	%DFW white	%DFW URM	Center intervention	Instructor status	New instructor
Descriptive	Intro Chem	Didactic	12	Content	400	extra large	10	5	13	None	Instructor	Yes
Coded	Chem 100	1	12	1	400	4	10	5	13	0	1	1

RQ1 Analyses (Overall Impact). We first explored the data to determine whether particular course or student variables differed across each group. No differences existed for class or student characteristics (Table 6); however, prior research suggest class size and level (e.g., lower- vs. upper-level course) may be factors to consider in characterizing instructional practice (Seymour & Hewitt, 1997). Therefore we included class size and level as covariates when assumptions of homogeneity of variance and homogeneity of slopes were met for the outcome variable (COPUS profile scores and URM % DFW rates). We then ran an ANCOVA to explore the impact of CTL interventions on these two outcome measures while controlling for number of enrolled students and level of course. Assumptions were not met for syllabus scores, so an ANOVA was run instead. Paired t-tests were used to explore differences in grade data within each group (e.g., difference in %DFW for URM and White students).

Table 6. *Comparison of Mean Information for Each Intervention Group*

	Class information		Instructor information		Student information		
	# students	Class level	Instructor type	% new faculty	% URM	% White	% female
None	82.02 (88.96)	1.49 (.52)	2.44 (1.38)**	20.71 (.41)**	11.28 (9.13)	53.24 (14.58)	43.41 (19.65)
CDI	59.75 (41.31)	1.39 (.50)	2.54 (1.04)**	28.60 (.46)**	10.70 (5.82)	58.11 (8.23)	46.48 (18.90)
STEMP	118.08 (113.83)	1.44 (.50)	2.17 (1.23)	11.11 (.32)**	10.87 (6.11)	57.19 (10.67)	48.53 (17.42)*
NFP	86.77 (85.38)	1.43 (.50)	1.46 (.56)	100.00 (.00)	9.51 (5.77)	50.86 (10.48)	36.50 (15.93)

Note. Data reported are means and standard deviations. Class level coded as freshmen (1), sophomore (2), junior (3), and senior (4). Instructor type coded as graduate student (0), instructor (1), assistant professor (2), associate professor (3), and full professor (4). ** significantly different than NFP, p<.05.

We also examined differences in outcome measures for smaller courses (<60 students) separately to understand differences in CTL interventions specific to these class sizes. Given the small size of each intervention group, we explored the data to ensure normality for each outcome measure. In cases where the homogeneity of variance assumption was violated (i.e. Levene's test <.05), a non-parametrics Mann-Whitney U test was run.

RQ2 Analyses (NFP). There were significant differences between NFP and other intervention groups based on instructor status and seniority (Table 6), so we focused solely on new faculty for this analysis. Because the number of courses taught by new faculty were so small for the CDI (n=8) and STEMP (n=4) intervention groups, we only compared differences in outcome measures for courses taught by new faculty that participated in NFP (n=35) and those that have not engaged in any center interventions (n=29). Outcome measures for the NFP and non-engaged groups were not normal, so we conducted non-parametric Mann-Whitney U tests to identify differences in observed instructional practices, course syllabi, and student % DFW rates. We also calculated frequencies to descriptively characterize differences between the two groups.

RQ 3 Analyses (STEMP cohort). Similar to the NFP data, the small number of courses observed in Department A (n=17) and Department B (n=29) dictated the use of non-parametrics Mann-Whitney U tests to identify differences in observed instructional practices, course syllabi, and student failure rates between the two departments. Comparing two departments where instructor status and course level differ, as mentioned above, may limit our ability to attribute differences to the STEMP program. However, analyses of the overall data reveal that instructional practices and student success did not differ by instructor status and class level. To triangulate our data set, we also calculated descriptives for overall departmental grade data for Department A in the year prior to (2014-2015) and following (2016-2017) their participation in STEMP.

Section 4: Findings, Significance, & Implications

Below we present the tables and figures referenced in the proposal and any additional details about these findings. Our sample includes purposefully sampled 239 courses, so we feel confident that many of our significant findings for the overall sample can be generalized to the STEM undergraduate population at our university. However, our sample sizes became very small for some subgroups, and we acknowledge that these smaller groups may not represent those populations. We differentiate these findings from the inferential findings by stating they are descriptive in nature and call for further research to examine these preliminary findings.

RQ1: Overall impact of CTL interventions

Overall there were significant differences between CTL intervention groups in measures of instructional practices but not in student success measures (Table 7). We found there were significant differences between CTL intervention groups' observed instructional practice, even when controlling for class size and type, $F(3, 233)=5.27$, $p<.01$. Courses taught by both STEMP and NFP participants had significantly more observed active learning than courses taught by participants who had not engaged in CTL interventions. We also found syllabi scores were significantly different between the groups, $F(3, 190)=15.80$, $p<.001$. Courses taught by participants who had engaged in any one of our three CTL interventions had significantly more learning-focused syllabi than courses taught by participants who have not engaged with our CTL.

Exploring student grades, there were no significant differences in student % grade rates between the CTL interventions. In other words, the % DFW for URM students in courses taught by CDI participants ($M=7.33$, $SD=10.65$) was not significantly different than %DFW for URM students in courses taught by STEMP participants ($M=8.45$, $SD=10.91$), $F(213, 3)=.47$, $p=.71$. However, there were significant differences within each group when comparing grades for URM and White students. The %DFW rates for URM students was significantly higher than %DFW rates for White students for all groups *except* NFP. There were no significant differences between URM and White student %DFW rates for courses taught by NFP faculty, t(30)=1.90, p=.067. Alarmingly, White students' %A rates were significantly higher than URM %A rates across the board and do not appear to be impacted by any of the CTL interventions.

Table 7. *Differences in Instructional practice and student success between CTL intervention groups.*

	Instructional practices, Mean (SD)		Student success, Mean (SD)			
	COPUS Group	Syllabus Score	URM %DFW	White % DFW	URM % A	White % A
None	1.44 (.68)	7.74 (7.42)	10.06 (16.04)	4.93 (6.75)	33.44 (31.30)	51.05 (24.30)
CDI	1.68 (.77)	16.41 (10.11)**	7.33 (10.65)	2.41 (3.84)	39.80 (34.10)	54.58 (23.84)
STEMP	1.75 (.87)*	17.65 (9.95)**	8.45 (10.91)	3.98 (4.76)	29.81 (29.81)	42.32 (21.53)
NFP	1.97 (.75)**	19.89 (8.92)**	7.62 (11.33)	3.99 (4.29)	38.34 (28.07)	57.14 (23.81)

Note. * significant from none, p<.05, ** significant from none, p<.01. COPUS group ranges from didactic (1), interactive lecture (2) and student centered (3). Syllabus score ranges from content-focused (0) to learning-focused (46).

With the challenges in engaging and retaining students in large, introductory STEM courses (e.g., Seymour & Hewitt, 1997), we focused on %DFW rates for smaller courses (n<60)[8] as a subset of our data (Figure 2). Descriptively the DFW rates for URM students were largest for small courses taught by instructors who have not engaged in CTE interventions (M=10.54) and smallest for courses taught by NFP faculty (M=1.56). In other words, it appeared the % DFW rate for URM students decreased the more contact hours instructors had with our CTE. Interestingly, White students' % DFW rates for courses taught by NFP faculty were higher than URM students. This finding is addressed below as it relates to instructional practices.

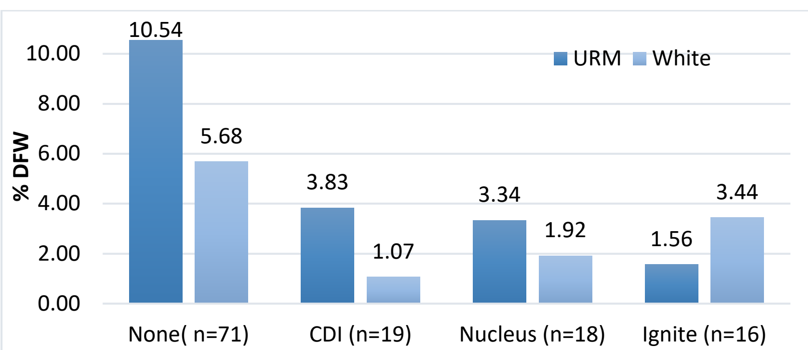

Figure 2. Differences in student failure rates for small course (n<60) by faculty engagement in different CTE interventions

We further separated out the CTL intervention groups by the observed instructional practices (i.e., didactic, interactive lecture, and student-centered instruction) (Figure 3). In courses where didactic instruction was observed, URM student % DFW rates were consistently higher than those of White students, with the exception of courses taught by NFP faculty. In courses characterized by interactive lecture, the % DFW rates for both White and URM students were lower than in courses with more than 80% lecture. The % DFW rate for White students was also lowest in interactive lecture courses, which was not the case for URM students.

[8] We chose this value as these courses would be considered small for our university context. For NFP, the larger courses were taught by only two instructors and we did not include these data in the analysis.

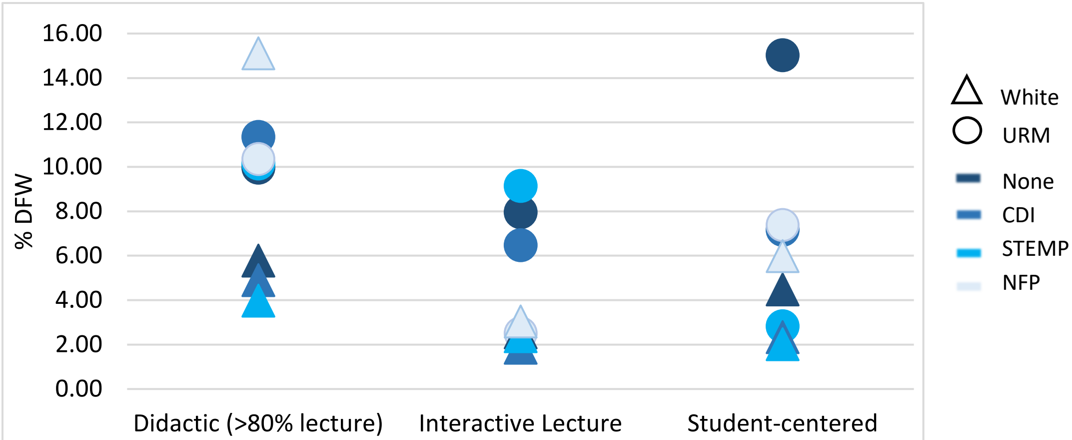

Figure 3. Descriptive differences in URM failure in courses with different observed instructional practices.

When comparing the URM and White student %DFW rate, the gap in URM and White student %DFW rate disappeared when interactive lecture courses were taught by NFP participants. For courses where student-centered instruction was observed, the %DFW rates for URM students in courses taught by non-engaged faculty (15.02%) is almost quadruple that of White student %DFW rates (4.48%). For student-centered courses taught by NFP faculty, the %DFW rate gap between URM and White students was again non-existent (~7%); however, the overall %DFW rate was higher than courses taught by NFP faculty who use interactive lecture (~3%). For student-centered courses taught by STEMP faculty, the %DFW rate gap between URM and White students disappeared *and* was low (~2%) for both groups.

RQ2: Impact of NFP program on new faculty

For courses taught by new faculty, there were significant differences in instructional practices but not in student success (Table 8). In particular, courses taught by NFP participants were observed to have significantly more active learning (Z=-3.05, p<.01) and more learning-focused syllabi (Z=-3.60, p<.001) when compared with courses taught by new faculty who have not engaged in any CTL interventions. There were no significant differences between grade data for courses taught by NFP faculty or new faculty who have not engaged in CTL interventions. There were also no differences between %DFW rates for URM and White students for either group; however, the significantly lower %A rate for URM students compared to White students still existed for both groups.

Table 8. *Differences in New Faculty's Instructional Practices and Student Success*

	Instructional practices, Mean (SD)		Student success, Mean (SD)			
	COPUS Group	Syllabus Score	URM %DFW	White % DFW	URM % A	White % A
NFP	1.97 (.75)**	19. 89 (8.92)**	7.62 (11.33)	6.85 (16.74)	38.34 (28.07)	56.51 (23.83)
None	1.44 (.78)	9.60 (8.76)	9.06 (13.24)	6.82 (9.98)	34.51 (33.07)	50.75 (25.01)

Note. ** significant, p<.01. COPUS group ranges from didactic (1), interactive lecture (2) and student centered (3). Syllabus score ranges from content-focused (0) to learning-focused (46).

Categorizing instructional practices further demonstrated the differences between these two groups. Nearly three quarters of courses (72.4%) taught by new faculty who have not engaged in CTL interventions utilized didactic instruction, whereas nearly three quarters of courses taught by NFP faculty utilized some sort of interactive component to their instruction (i.e., 45.7% interactive lecture, 25.7% student-centered) (Figure 4). Similarly, 75% of syllabi were content-focused for courses taught by faculty who have not engaged in any CTL interventions while two-thirds of syllabi for courses taught by NFP faculty had learning-focused elements (57.6% transitional, 9.1% learning-focused) (Figure 5).

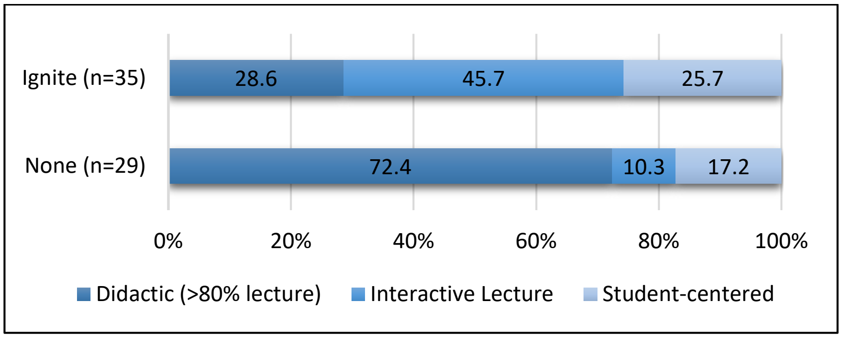

Figure 4. Observed Instructional Practices of Courses Taught by New Faculty

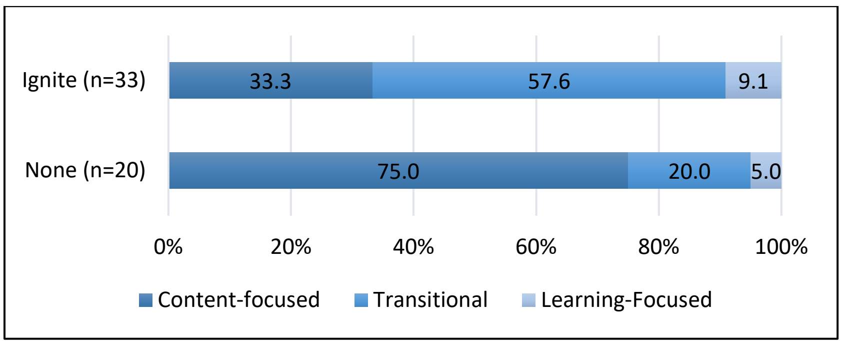

Figure 5. Syllabi Focus for Courses Taught by New Faculty

When exploring %DFW rates for courses taught by new faculty, the gap between URM and White students was negligible, regardless of whether they had participated in NFP (e.g., 7.50% vs. 6.81%) (Figure 6). However, %DFW rates for courses taught by non-engaged faculty was five times higher for URM students (7.50%) than in courses taught by NFP faculty (1.56%).

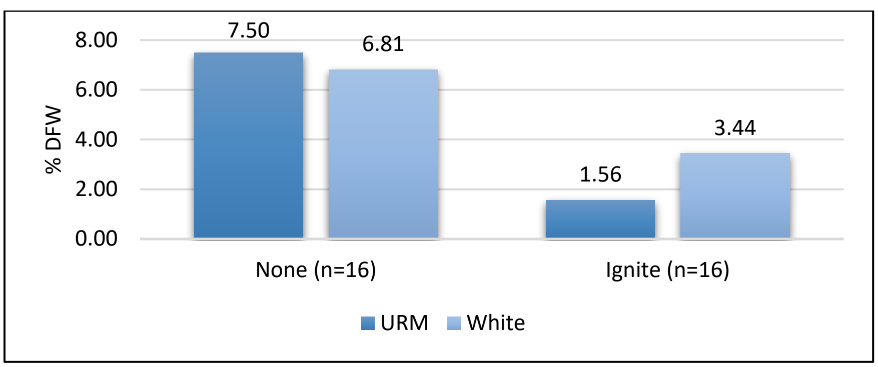

Figure 6. Descriptive differences in %DFW rates and instruction for small courses (n<60) taught by new faculty.

RQ3: Impact of a cohort model within the STEMP program

When comparing Department A, containing a STEMP cohort, to Department B, the control department, there were significant differences in instructional practices and student success (Table 9).

Courses taught in Department A had significantly more observed active learning ($Z=-2.85$, $p<.01$), more learning-focused syllabi ($Z=-3.05$, $p<.01$), lower URM student %DFW rates ($Z=-2.85$, $p<.01$), lower White student %DFW rates (-3.19, $p<.01$), higher URM student %A rates ($Z=-2.11$, $p<.05$), and higher White student %A rates (-2.54, $p<.05$).

Table 9. *Differences in Instructional Practices and Student Success for a Cohort of Faculty*

	Instructional practices, Mean (SD)		Student success, Mean (SD)			
	COPUS Group	Syllabus Score	URM %DFW	White % DFW	URM % A	White % A
Dept A (cohort)	2.53 (.87)**	17.40 (10.03)**	2.81 (7.44)**	2.87 (4.18)**	36.02 (31.80)*	48.91 (18.91)*
Dept B (control)	1.48 (.83)	7.76 (8.05)	21.63 (25.37)	10.83 (8.82)	18.54 (25.16)	34.21 (17.76)

Note. * significant, $p<.05$. ** significant, $p<.01$. COPUS group ranges from didactic (1), interactive lecture (2) and student centered (3). Syllabus score ranges from content-focused (0) to learning-focused (46).

Descriptively, there were differences in the %DFW rates between URM and White students for each Department (Figure 7). The %DFW rates for Department A were virtually non-existent between URM and White students (~3%), while URM students in Department B have a %DFW rate twice that of White students (21.63% vs. 11.44%). The stark differences between Department A and B may be related to the impact of a cohort of faculty engaged in STEMP; however, because there are other confounding factors (e.g., student population, instructor type), this is not a definitive conclusion.

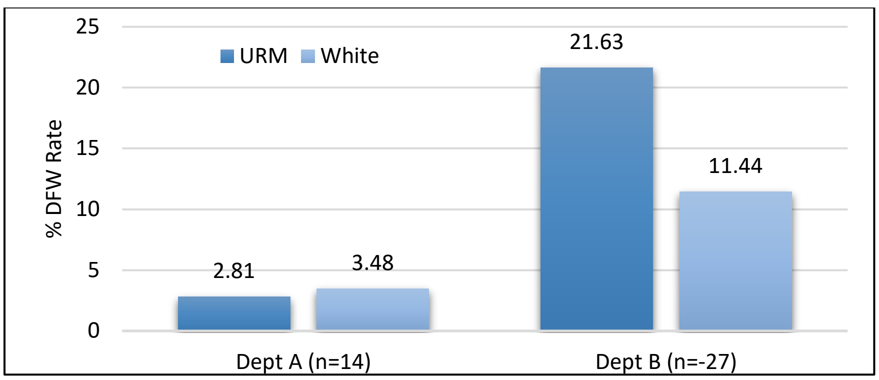

Figure 7. Differences in % DFW rates for courses taught in comparable departments with and without CTL interventions.

We therefore sought to describe the overall grade rates for Department A before and after the cohort of faculty participated in STEMP (37.5% of the departmental teaching faculty) to better understand the impact of the Program (Figure 8). While the departmental %DFW rate for White students did not change pre- to post-STEMP, the URM student %DFW rate for the department did decrease from 15.89% to 9.59%. This suggests that the cohort of STEMP faculty may have had an impact on the URM student failure rate overall in the department.

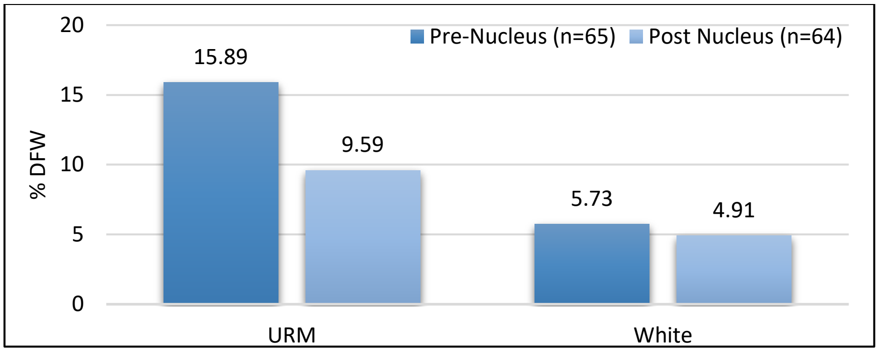

Figure 8. Differences in % DFW rates for courses taught in Department A before and after a faculty cohort participated in STEMP.

Implications and Future Work

There are some limitations to the present study that include the small sample size of the subgroups of courses used to answer research questions 2 and 3 and the self-selection of instructors for participation in CTL programming. Despite the limited ability to make inferential claims about these CTL interventions when comparing to similar ‘control’ groups (i.e., new faculty without CTL interventions, Department B without CTL interventions), the results are promising. Further, comparing pre- and post-data for Department A eliminates self-selection as a possible explanation for the reported finding. We have demonstrated that it is possible to leverage both direct measures of instruction and institutional data to better understand the larger picture of our CTL work with faculty, helping move the field of the educational development forward.

We do not aim to draw definitive conclusions from these data or generalize them beyond our institution, but the potential implications of our work may be useful to other institutions in considering how to move beyond satisfaction and self-report data to understand the impact of their own interventions. It may also provide avenues of future research for scholars seeking to add to the scholarship of educational development.

Our main implications are:

1. Our present study adds to the work on the impact of CDI on faculty syllabi (Authors, 2016) and suggest that CDI has a lasting impact on the way instructors design courses. Of the 147 STEM instructors who have participated in CDI since 2008, 99 participated in the present study (67.3%) as either CDI, STEMP, or NFP participants. These three intervention groups had significantly more learning-focused syllabi than the non-engaged CTL group, which suggests that faculty continue to utilize and implement learning-focused syllabi beyond the course they designed in CDI. Future work could examine how faculty translate their course design to different courses and over time.
2. The large %DFW gap between URM students and White students in student-centered courses taught by non-engaged faculty may be the result of poor implementation of student-centered instruction. Conversely, it appears that student-centered instruction has the potential to close the % DFW gap between White and URM students when faculty engage in intensive CTL interventions. These results add to the recent literature on undergraduate STEM courses that suggested the ways in which active learning is implemented may differentially impact students (e.g., Cooper, Downing, and Brownell, 2018; Snyder, Sloane, Dunk, & Wiles, 2016). Further research is needed not only to explore whether active learning is use, but *how* it is implemented. Our study may suggest that faculty indeed need support in implementing active learning in a way that supports all students.

3. The particularly low %DFW rates and closing of the failure gap between URM and White students observed for student-centered courses taught by STEMP faculty could potentially be explained by a cohort effect. When participants engage in CTL interventions as a cohort (i.e., STEMP), they may be interacting and discussing their course implementation more so than instructors who participate solo. Qualitative exploration of these interactions beyond the FLC may help better understand the impact of a cohort on faculty instruction and student failure.
4. Regardless of interventions, differences in URM and White student %A rates were always significant. This suggests a potentially larger systematic issue that needs to be addressed. Further work exploring institutional data and better understanding student preparation for STEM courses may be needed to explain this phenomenon.