Data-driven Iterative Refinements to Educational Development Services: Directly Measuring the Impacts of Consultations on Course and Syllabus Design

Chad Hershock
Carnegie Mellon University, hershock@andrew.cmu.edu

Laura Ochs Pottmeyer
Carnegie Mellon University

Jessica Harrell
Carnegie Mellon University

Sophie le Blanc
Carnegie Mellon University

Marisella Rodriguez
University of California, Merced

Authors

Chad Hershock, Laura Ochs Pottmeyer, Jessica Harrell, Sophie le Blanc, Marisella Rodriguez, Jacqueline Stimson, Katherine Phelps Walsh, and Emily Daniels Weiss

Data-driven iterative refinements to educational development services: directly measuring the impacts of consultations on course and syllabus design

Chad Hershock, Laura Ochs Pottmeyer, Jessica Harrell, Sophie le Blanc, Marisella Rodriguez, Jacqueline Stimson, Katharine Phelps Walsh, and Emily Daniels Weiss

Abstract

Evidence-based practice in educational development includes leveraging data to iteratively refine center for teaching and learning (CTL) services. However, CTL data collection is often limited to counts and satisfaction surveys rather than direct measures of outcomes. To directly assess impacts of consultations on course and syllabus design, we analyzed 94 clients' syllabi (32 faculty, 62 graduate students and postdocs) before and after consultations. Faculty and non-faculty clients demonstrated significant change following consultations (6% and 10% gains in syllabus rubric scores, representing 50% and 31% of possible gains and effect sizes of 0.73 and 1.04 standard deviations, respectively). We compared faculty clients to quasi-experimental control groups that did not receive consultations. Syllabi from non-clients scored lower and did not demonstrate similar changes across semesters. Attendance at a CTL seminar on course and syllabus design did not explain variation in clients' syllabi. We discuss implications for assessment of CTL services and how we leveraged

formative assessments to inform and iteratively refine our educational development practices.

Keywords: educational development, formative outcomes assessment, centers for teaching and learning

How can centers for teaching and learning (CTLs) measure the outcomes of their services to inform their educational development practices? For example, CTL services frequently disseminate effective course and syllabus design (C&SD) principles (e.g., see Ambrose et al., 2010; Fink, 2013; Wiggins & McTighe, 2005). Common approaches include one-on-one consultations, stand-alone seminars, orientation events, multi-day institutes, and/or web resources. Our CTL allocates significant resources to all of the above but disproportionately more to consultations. Naturally, we wondered whether our consultation services enhance clients' C&SDs. Furthermore, how could we leverage outcomes assessments to continuously improve our educational development practices? This article describes how we used syllabi analyses as direct outcomes measures to formatively assess and reflect on our C&SD consultation services. It also contributes to a gap in the scholarship of educational development (SoED) literature on outcomes assessment.

Scholarship on assessing CTL impacts identifies multiple possible levels of outcomes analyses (Haras et al., 2017; Hines, 2015; Kreber et al., 2001):

1. counts of instructors served;
2. instructor satisfaction with CTL services;
3. instructor beliefs about teaching and learning;
4. instructor perceptions of institutional culture;
5. instructor learning gains;

6. instructor course and syllabus designs;
7. instructor teaching practices during and between class sessions; and
8. student outcomes, including learning, persistence, or attitudes.

Previous theoretical and empirical SoED aggregates Outcomes 6 and 7 broadly as teaching behaviors or practices. We intentionally separate and delineate these outcomes because CTL services may target one or both, and different data sources may be required for each.

Typical CTL evaluation practices focus on Outcomes 1 and 2 (Beach et al., 2016; Haras et al., 2017; Hines, 2011). These data are easy to collect and can indicate the reach and relevance of CTL services. Survey data can provide useful formative feedback (e.g., perceived strengths and weaknesses of services and suggestions for change). Nevertheless, neither client counts nor typical CTL evaluation forms *directly* measure educational development impacts. For instance, feedback surveys may ask instructors to self-report changes regarding Outcome 7. Unfortunately, studies comparing instructors' self-reports to classroom observations suggest self-reports are unreliable as indirect measures of the impacts of professional development programs (Ebert-May et al., 2011). Similarly, self-reports of confidence or perceived ability to implement evidence-based practices encountered via CTL services are indirect measures of outcomes. They do not directly measure changes in knowledge, skill, or practice. Because counts, satisfaction data, and self-reports alone fail to adequately inform how best to iteratively refine CTL services or fully demonstrate the value added by CTLs, recent reports argue for incorporating more and broader data sources directly measuring educational development outcomes (Beach et al., 2016; POD Network, 2018). Historically, SoED studies omit Outcomes 5–8 above (Chism et al., 2012; Stes et al., 2010; but for recent exceptions, see Palmer 2016; Tomkin et al., 2019; Wheeler & Bach, 2021). Here, we focus on CTL outcomes analyses regarding changes in instructor course designs as documented in syllabi (Outcome 6). Our study also illustrates our use of syllabi as data sources to formatively inform and refine educational development practices.

Syllabi analyses may reveal the current state of C&SD elements, such as learning objectives, assessments, pedagogical methods, course policies, and alignment among course design features (e.g., Cullen & Harris, 2009; Doolittle & Siudzinski, 2010; Homa et al., 2013; McGowan et al., 2016; Palmer et al., 2016; Stanny et al., 2015). Syllabi may also provide data on non-contextually dependent teaching and learning constructs, such as time on task and level of expectations (Campbell et al., 2019) or learner centeredness (Palmer et al., 2016). Syllabi can be readily compared across CTL clients, regardless of their discipline and teaching contexts. Of course, syllabi have limitations as data sources. Instructors may use practices not documented in syllabi or fail to use practices documented in syllabi. For classroom teaching practices, specifically, another data source, such as an observation or analyses of other teaching artifacts, may be more appropriate. Regardless, educational developers can still gain valuable insights from syllabi analyses regarding impacts of CTL services, especially those targeting elements of C&SD (see above) rather than, or in addition to, Outcome 7 above. Furthermore, when services target adoption of evidence-based teaching strategies, the work of explicitly and intentionally articulating related C&SD elements in writing can be an important step. It can support instructor commitment to adoption and foster reflection on implementation, especially if instructors perceive syllabi as a public agreement with students. Additionally, syllabi analyses are amenable for pre/post analyses of interventions and outcome assessments at scale, especially when resources are limited.

Surprisingly, few studies analyze syllabi to measure the impact of an intervention on instructor C&SD (Chism et al., 2012; Stes et al., 2010). Palmer et al. (2014) developed and validated a rubric to measure the degree to which syllabi are learning centered. Researchers used this rubric to compare changes in syllabi before and after a week-long course design institute (Palmer et al., 2016). This approach illustrates the value of capturing pre/post measurements. However, including a comparison group (e.g., instructors not enrolled in the program) would strengthen inferences regarding causality. Research

on CTL impacts rarely includes comparison groups (Chism et al., 2012; Stes et al., 2010; but see Meizlish et al., 2018; Wheeler & Bach, 2021).

Our study directly assesses the impact of a CTL consultation service on C&SD as documented in syllabi via pre/post analyses and includes a comparison group. We analyzed syllabi from 94 clients (32 faculty, 54 graduate students, and eight postdocs) before and after each received a CTL consultation on course and syllabus design. Using our institution's syllabus registry, we generated two quasi-experimental comparison groups for faculty who did not receive a consultation: (a) individual course syllabi from 32 faculty members, and (b) pairs of syllabi from 10 faculty members who taught the same course in consecutive semesters. We matched all faculty client and comparison group syllabi on discipline and course level and, when possible, employment track and faculty rank. Our analyses included data on prior attendance at CTL CS&D seminars to statistically "control" for effects on pre-/post-consultation data.

Research Questions and Significance of Study

Our study contributes to the literature on CTL outcomes assessments by exploring the following research questions. To what extent:

1. do CTL consultation services influence instructor course design, as documented in their course syllabi?
2. do the impacts of CTL consultations on C&SD differ between faculty and graduate student/postdoc clients?
3. does participation in a CTL C&SD seminar, prior to a CTL consultation, influence:
   a. instructors' C&SD practices, as documented in their pre-consultation syllabi?
   b. the impact of a consultation, as documented by pre/post changes in syllabi?

These questions are relevant for educational developers for several reasons. First, our study helps fill two well-documented gaps in the literature, the need for (a) additional evaluations of the impacts of CTLs services beyond counts of clients and satisfaction data (Beach et al., 2016; Haras et al., 2017; Kucsera & Svinicki, 2010) and (b) assessments of CTL impacts using direct, outcomes measures, pre/post data, and comparison groups (Chism et al., 2012; Stes et al., 2010; Wheeler & Bach, 2021). Second, we compare CTL impacts between current and future faculty. Some CTLs offer consultation services to both constituencies, such as consultations on C&SD, classroom observations of teaching, or Small Group Instructional Diagnosis (SGID; see Finelli et al., 2008; Finelli et al., 2011). However, scholarship comparing impacts across types of clients is lacking (Chism et al., 2012; Stes et al., 2010). Action research (Hershock et al., 2011) regarding how each constituency responds to consultation services can help CTLs tailor their service models to best meet clients' needs. Third, we provide a data-driven model for how to formatively evaluate and iteratively refine the delivery of CTL services.

Methods

We conducted this study at the Eberly Center for Teaching Excellence and Educational Innovation, the CTL at Carnegie Mellon University (CMU). CMU is a research-intensive, private institution with approximately 14,000 graduate and undergraduate students and 1,400 faculty. Consultations represent a large component of the CTL's portfolio. During the academic year 2019–2020, the CTL consulted with 470 faculty and 151 graduate student clients. This study addresses the extent to which clients' syllabi change following a C&SD consultation.

C&SD Consultations

In this section, we compare and contrast our CTL's C&SD consultation service models for faculty clients and graduate students/

postdocs participating in our Future Faculty Program (FFP). Similarities are numerous. All consultations are voluntary, confidential, and provided by full-time educational developers. Consultants receive the same training, including shadowing and being shadowed by more experienced colleagues on consultations and participating in monthly professional development sessions to compare strategies and discuss challenging consultation scenarios. Clients meet with consultants one or more times. Discussions typically focus on learning objectives, assessments, instructional strategies, alignment of course design features, and/or course policies. Clients may be designing a course from scratch, revising a syllabus they inherited from previous instructors, or inspired by syllabi encountered elsewhere. Consultants provide resources, as needed, such as a syllabus template, a set of heuristic questions, and/or recommendations from our institution's Faculty Senate (Eckhardt, 2017). Rather than follow a strict protocol, consultants draw from the same, established C&SD and learning principles (e.g., Ambrose et al., 2010; Fink, 2013; Wiggins & McTighe, 2005) regarding learning objectives, assessments, alignment, inclusive teaching, course policies, and how students learn to provide feedback on clients' C&SDs. However, to strategically "meet the client where they are," including what to prioritize, how much feedback to provide, and when to gently push (or not), consultants also have to rely on intuition and be flexible in their approach. After receiving feedback, clients may revise syllabi and receive additional feedback. Given that contextual, client-specific factors influence the exact content of a consultation, one expects some variation in the client experience. Given this variation, our tests of the impacts of consultations are conservative. Detecting a strong signal regarding syllabi improvements after consultations or relative to a comparison group, in spite of this potential variation, actually strengthens interpretations of the benefits of the fundamental, shared parameters of the C&SD service model.

C&SD consultations may differ between faculty and FFP clients. Time on task regarding C&SD tends to be greater during FFP

consultations. FFP consultations focus exclusively on C&SD, because it's a specific requirement of the FFP program. Faculty consultations often include other CTL services and have shorter timelines for iteration. Many faculty clients request a C&SD consultation proximal to the start of a semester when they will teach the targeted course. Fewer FFP clients are (or will be) teaching the courses they are designing at CMU. Thus, consultants may need to prioritize feedback differently for faculty, based on time constraints and what's possible or most critical.

Additionally, FFP consultations tend to be more scaffolded. Unlike faculty, FFP clients participate in a structured program designed to support skill development and success as educators in a faculty career (Eberly Center, 2021). FFP participants must (a) complete a course and syllabus (re)design project; (b) attend at least eight CTL seminars on evidenced-based teaching and learning; (c) receive at least two teaching feedback consultations; and (d) write a teaching philosophy statement. FFP clients receive feedback on syllabi drafts using a rubric assessing course descriptions, learning objectives, description of assessments, assessment criteria for evaluation (if applicable), course policies, and more (see Appendix A). Faculty clients may also receive feedback on their syllabi, but it is not scaffolded using the FFP rubric. Typically, FFP clients do not receive the rubric prior to submitting their first draft. However, consultants may conduct a generative interview with FFP clients before they draft their syllabi. Generative interviews do not "workshop" elements of course or syllabus design in detail. Instead, they pose guiding questions to help clients get started. Finally, after FFP syllabi pass minimum requirements (on the first draft or after iteration), all FFP clients complete a written reflection on their course designs. This assignment challenges clients to communicate the alignment of course design features, especially regarding instructional strategies, assessments, and learning objectives, because this information is rarely explicit in a traditional course syllabus. Consultants do not request this reflection document from faculty clients.

Study Design and Data Sources

We compared syllabi for 94 CTL clients before and after they received a C&SD consultation. Syllabus pairs came from the same course.

Two-thirds of our sample came from FFP clients. From January 2017 through June 2018, 62 FFP participants (16 master's students, 38 doctoral students, and eight postdocs) received a C&SD consultation and submitted syllabi before and after consultations (Table 1).

One-third of our sample came from faculty clients. We identified 57 unique faculty-course combinations that requested a C&SD consultation between August 2017 and June 2019 and shared syllabi with consultants at the beginning of the process. We emailed these clients to request the syllabi they implemented after the consultation and received syllabi from 43 client-course combinations. Eleven respondents submitted syllabi for multiple courses receiving consultations. For repeat clients, we only analyzed syllabus pairs from their earliest consultation, resulting in a final sample of 32 faculty clients (Table 2).

To better assess CTL impacts, we also analyzed syllabi from two comparison groups of faculty who did not receive a C&SD consultation. In 2018, CMU established a course syllabus registry to better support its students. We identified registry syllabi from faculty who did not receive CTL consultations from 2016 through June 2019. We

Table 1. FFP Clients Sample

Discipline/college	Proportion of clients (%)
	n = **62**
Engineering	18 (29.0)
Humanities and social sciences	14 (22.6)
Fine arts	13 (21.0)
Computer science	8 (12.9)
Science	5 (8.1)
Business	3 (4.8)
Information systems and public policy	1 (1.6)
Participation in additional CTL C&SD services	
C&SD seminar	36 (58.1)

Table 2. Faculty Clients Sample

Discipline/college	CTL clients (*n* = 32)	Syllabus registry, 1 syllabus (*n* = 32)	Syllabus registry "drift," 2 syllabi (*n* = 10)
	Proportion of faculty (%)	Proportion of faculty (%)	Proportion of faculty (%)
Humanities and social sciences	11 (34.4)	11 (34.4)	1 (10)
Engineering	6 (18.8)	6 (18.8)	–
Fine arts	5 (15.6)	5 (15.6)	3 (30)
Information systems and public policy	4 (12.5)	4 (12.5)	2 (20)
Science	2 (6.3)	2 (6.3)	2 (20)
Business	2 (6.3)	2 (6.3)	1 (10)
Computer science	1 (3.1)	1 (3.1)	–
Other	1 (3.1)	1 (3.1)	1 (10)
Employment track			
Tenure track	11 (34.4)	13 (40.6)	3 (30)
Teaching track	9 (28.1)	8 (25.0)	2 (20)
Research track	1 (3.1)	1 (3.1)	–
Adjunct, special, or visiting	11 (34.4)	10 (31.3)	5 (50)
Rank			
Assistant Professor	15 (50.0)	6 (18.8)	1 (10)
Associate Professor	4 (12.5)	8 (25.0)	2 (20)
Full Professor	1 (3.1)	7 (21.9)	2 (20)
Adjunct, Special, Visiting, or Research Professor	12 (37.5)	11 (34.4)	5 (50)
Participation in additional CTL C&SD services			
C&SD seminar	13 (40.6)	0	0

closely matched client and registry syllabi regarding course discipline and level as well as faculty employment track and rank (Table 2). When multiple registry matches occurred, we randomly selected one course. Our first comparison group included a single syllabus from 32 unique faculty. Ten faculty matches in the course registry posted syllabi for the same course taught in consecutive semesters. These faculty formed our second comparison group, accounting for "syllabus drift," ambient changes in syllabi across semesters from instructors independently revising courses. Together, these comparison groups provide reference points for comparing faculty clients' syllabi before versus after a consultation. We could not procure a viable comparison group for graduate students and postdocs who did not receive a consultation.

Additionally, we included data on clients' attendance at CTL seminars on C&SD prior to the completion of consultations. Seminar attendance data spanned July 2016 to the time of consultation. For faculty, we used attendance at the CTL's annual Incoming Faculty Orientation event, which contains a seminar devoted to C&SD. We excluded faculty attending these seminars from our syllabus registry comparison groups. No other CTL events for faculty focused exclusively on C&SD during this period. For FFP participants, we used attendance data from the CTL's annual Graduate Student Seminar Series, which offers at least one seminar on C&SD each year.

Syllabi Coding

The FFP consultation rubric was not designed as a research instrument. Therefore, to assess the quality of syllabi in our study, we created a different rubric informed by course design and evidence-based teaching literature (Ambrose et al., 2010; Palmer et al., 2014; Wiggins & McTighe, 2005) and CMU Faculty Senate recommendations (Eckhardt, 2017).

Six of the authors (C&SD experts) served as syllabus coders and calibrated by individually coding two anonymous syllabi not included in this study. Two of the remaining authors (assessment experts) compiled the codes, identified areas of disagreement, and facilitated a discussion in which the coders revised the rubric. With the revised rubric, the coders independently scored a third anonymous syllabus, followed by another round of rubric revision. The final rubric included six broad categories—*course description*, *learning objectives*, *assessments*, *assessment and grading policies*, *other policies*, and *organization*—each with subcategories for specific features (see Appendix A).

Given the consensus among raters following the norming activities, the six coders independently scored a unique subset of 220 syllabi in the study. Syllabi were randomly assigned, but we ensured that coders did not analyze a syllabus from one of their clients or analyze

both pre- and post-consultation syllabi from the same client. Prior to coding, we de-identified syllabi regarding instructor identity and study conditions.

With a maximum of 56 points possible on the rubric, coders rated each syllabus on 17 of the 20 rubric items (Appendix A). However, coders only rated the remaining three items (9 total points possible) if they were present in the syllabus. Coders agreed that these items (participation grading, other policies, and explanation for other policies) were not necessarily essential in syllabi and thus "optional" from a consultant's perspective. Consequently, we did not count their absence against the score of a syllabus and reduced the total possible points for that syllabus. We converted rubric scores to a proportion to account for the variation in possible point totals and to compare across rubric categories.

Data Analysis

To investigate the impact of consultations on C&SD practices (Research Question 1), we conducted separate analyses for faculty and FFP clients because we were unable to obtain an FFP comparison group. We analyzed FFP syllabus scores using within-subjects, repeated measures ANOVA with syllabus condition (pre-consultation, post-consultation) as a fixed factor. We applied a general linear model to faculty client data and the first syllabus registry comparison group ($n = 32$ matched course syllabi). Independent variables included syllabus condition (pre-consultation, post-consultation, or syllabus registry Comparison Group 1) as a fixed factor and unique faculty identity (dummy coded) as random effects variables. The faculty identity variables account for repeated measures across pre- and post-consultation syllabi for clients. Because we identified only 10 syllabus registry matched pairs meeting selection criteria (Comparison Group 2), we deemed this sample too small for a reliable statistical test. Instead, we include that data below as an observational comparison.

To determine the extent to which syllabus scores and rates of change differed between faculty or FFP clients (Research Question 2) as well as the influence of CTL seminars on the consultation outcomes (Research Question 3.b), we conducted a three-way mixed ANOVA on syllabus rubric scores. Independent variables included syllabus condition (pre-consultation, post-consultation) as a repeated measures factor as well as client type (faculty, FFP) and prior seminar attendance (yes, no) as between-subjects factors. We included all possible two-way and three-way interactions in the model. The *syllabus condition × client type* interaction term directly tests the null hypothesis of Research Question 2—that is, impacts of consultations on syllabi are independent of the type of client. The *syllabus condition × seminar attendance* interaction term directly tests the null hypothesis of Research Question 3.b—that is, impacts of consultations on syllabi are independent of whether clients previously attended a CTL seminar on C&SD. The three-way interaction of *syllabus condition × client type × seminar attendance* directly tests whether impacts of consultations are independent of both client type and previous seminar attendance.

To measure the influence of CTL workshops on the initial condition of clients' syllabi (Research Question 3.a), we conducted a two-way ANOVA on pre-consultation syllabus scores. Independent variables included client type (faculty, FFP) and previous seminar attendance (yes, no) as between-subjects factors.

In all statistical analyses, we checked model assumptions by evaluating boxplots of residuals to identify outliers, the Shapiro-Wilk test and Q-Q plots of residuals to verify normality, Levene's test for homogeneity of variance, and Box's M test for homogeneity of covariances. We did not identify outliers or violations of homogeneity of variance or covariance in our data sets. We only detected violations of normality in the post-consultation syllabus scores in the three-way ANOVA. We proceeded with the ANOVA because this type of analysis is robust to violations of normality (Norman, 2010).

Results

RQ1: To what extent do CTL consultation services influence instructor course design, as documented in their course syllabi?

Faculty Syllabi

Total syllabus scores differed significantly across the three syllabus conditions (pre-consultation, post-consultation, and registry), $F(1,31) = 17.14$, $p < .001$, partial $n^2 = .356$ (Figure 1). This result indicates that approximately 36% of the variance in total syllabus scores depends on experimental conditions (pre-consultation, post-consultation, or registry). A post hoc Tukey HSD test found that the mean differences between each of the three possible pairwise combinations of conditions are statistically significant (Table 3). Clients' syllabi scored higher after a consultation ($M = .87$, $SD = .08$) than before ($M = .81$, $SD = .10$). Clients' pre- and post-consultation syllabi both scored higher than syllabi from non-clients matched from the syllabus registry, whether non-clients posted a single syllabus (Comparison Group 1: $M = .70$, $SD = .11$) or pairs of syllabi for the same course in consecutive semesters (Comparison Group 2: Semester 1 $M = .74$, $SD = .05$; Semester 2 $M = .74$, $SD = .80$). Due to small sample size, we do not present statistical analyses including the syllabus drift comparison group. Observationally, in contrast to clients, syllabus drift samples from non-clients appear to show little or no change across consecutive semesters, as measured by our syllabus rubric. Faculty clients' syllabi also consistently score higher than syllabus drift samples. After

Table 3. Multiple Pairwise Comparisons (Tukey HSD tests) for Scores Across Combinations of Three Faculty Syllabus Conditions

Condition A	Condition B	Mean difference (A–B)	p	95% CI
Pre-consultation	Post-consultation	-.057	.001*	[-.09,-.02]
Post-consultation	Registry	.167	< .001*	[.13,.20]
Registry	Pre-consultation	-.110	< .001*	[-.14,-.08]

*Statistically significant.

Table 4. Mean Syllabus Scores for Non-Clients Posting a Single Registry Syllabus and Faculty Clients (t-tests performed on client data only)

Category	Registry proportion of points scored (0–1) (SD)	Clients' proportion of points scored (0–1) (SD)		t (client pre/post)	p	Effect size (d)
		Pre-consult	Post-consult			
Course description	.76 (.17)	.76 (.18)	.81 (.18)	-1.54	.13	–
Learning objectives	.78 (.20)	.86 (.13)	.88 (.14)	-0.71	.48	–
Assessments	.74 (.16)	.80 (.16)	.87 (.14)	-3.21	< .001*	-.57
Assessment and grading policies	.72 (.21)	.77 (.19)	.84 (.16)	-3.30	< .001*	-.58
Other course policies	.72 (.19)	.82 (.18)	.89 (.10)	-2.59	.01*	-.46
Organization	.88 (.18)	.93 (.14)	.94 (.13)	-.44	.66	–
Total	**.70 (.11)**	**.81 (.10)**	**.87 (.08)**	**-4.13**	**< .001***	**.73**

*Statistically significant.

consultations, faculty syllabi scored significantly higher in three rubric categories: assessments, assessment and grading policies, and other course policies (Table 4).

FFP Syllabi

FFP clients receiving C&SD consultations scored significantly higher on post-consultation syllabi (*M* = 0.90, *SD* = .066) than pre-consultation syllabi (*M* = 0.80, *SD* = .095), *t*(61) = -8.17, *p* < .001, with a large effect size, *d* = 1.04 (Table 5). Additionally, FFP clients demonstrated statistically significant increases in the following rubric categories: course description, learning objectives, assessments, assessment and grading policies, and other course policies.

RQ2: To what extent do the impacts of CTL consultations on C&SD differ between faculty and graduate student/postdoc clients?

Overall changes in syllabus scores following consultations resulted in strong effect sizes for both faculty and non-faculty clients (Tables 4

Table 5. FFP Clients' Mean Syllabus Scores

Category	Pre–proportion of points scored (0–1)	Post–proportion of points scored (0–1)	t	p	Effect size (d)
Course description	.76 (.18)	.87 (.14)	-3.65	.001*	.46
Learning objectives	.83 (.20)	.94 (.08)	-4.46	< .001*	.57
Assessments	.81 (.15)	.92 (.10)	-5.24	< .001*	.67
Assessment and grading policies	.78 (.17)	.89 (.12)	-4.93	< .001*	.63
Other course policies	.76 (.14)	.88 (.10)	-6.31	< .001*	.80
Organization	.90 (.15)	.93 (.15)	-1.10	.28	–
Total	**.80 (.09)**	**.90 (.07)**	**-8.17**	**< .001***	**1.04**

*Statistically significant.

and 5), representing increases of 0.73 and 1.04 standard deviations, respectively. However, changes in syllabus scores following consultations differed between faculty and FFP clients. The three-way mixed ANOVA on all client syllabus scores did not exhibit a statistically significant three-way interaction among syllabus condition (pre-consultation, post-consultation), client type (faculty, FFP), or previous seminar attendance (yes, no), $F(1,90) < 0.0$, $p = .995$. However, this ANOVA revealed a statistically significant interaction between client type (faculty, FFP) and syllabus condition (pre-consultation, post-consultation), $F(1,92) = 5.109$, $p = .026$, partial $n^2 = .053$ (Figure 2). While pre-consultation syllabus scores did not differ between client types (FFP $M = .80$, $SD = .09$; faculty $M = .81$, $SD = .10$), FFP clients received higher post-consultation scores ($M = .90$, $SD = .07$) than faculty clients ($M = .87$, $SD = .08$). While total syllabus scores improved for both types of clients, FFP clients improved more than faculty (10% vs. 6%, respectively). When calculated as a relative growth rate,

$$(F – I) \div (Max – I)$$

where F = final syllabus score, I = initial syllabus score, and Max = maximum possible syllabus score, FFP and faculty clients demonstrated 50% and 31% gains, respectively, relative to possible growth. Faculty clients' growth was driven primarily by categories related to assessments
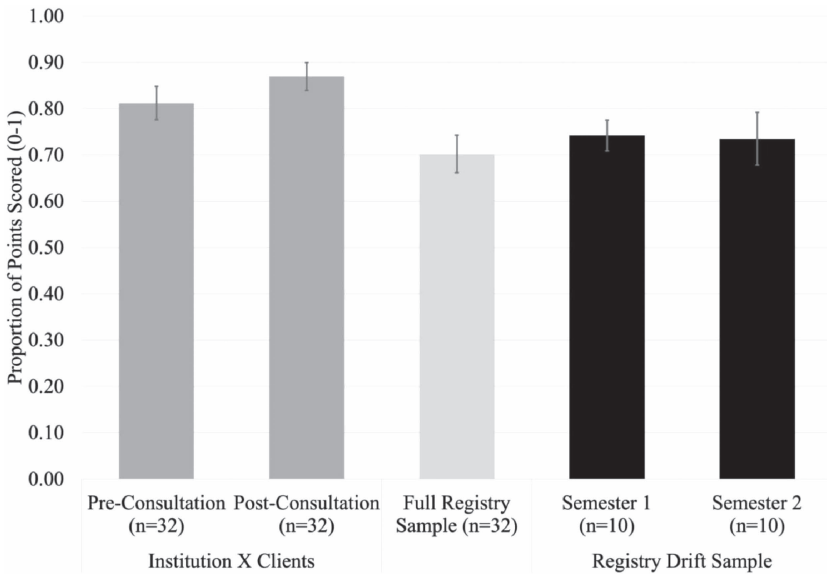
Figure 1. Mean Scores for Faculty Clients' Pre- and Post-Consultation Syllabi and Matched Non-Client Samples (error bars represent 95% confidence intervals)

and policies (Table 4), whereas FFP clients' scores increased in almost every category (Table 5).

RQ3: To what extent does participation in a CTL C&SD seminar, prior to a CTL consultation, influence:

   a. instructors' C&SD practices, as documented in their pre-consultation syllabi?
   b. the impact of a consultation, as documented by pre/post changes in syllabi?

Seminar attendance prior to a consultation did not impact clients' initial syllabus scores. The two-way ANOVA testing the influence of seminar attendance on clients' pre-consultation syllabus scores found no significant interaction between prior attendance at a C&SD seminar
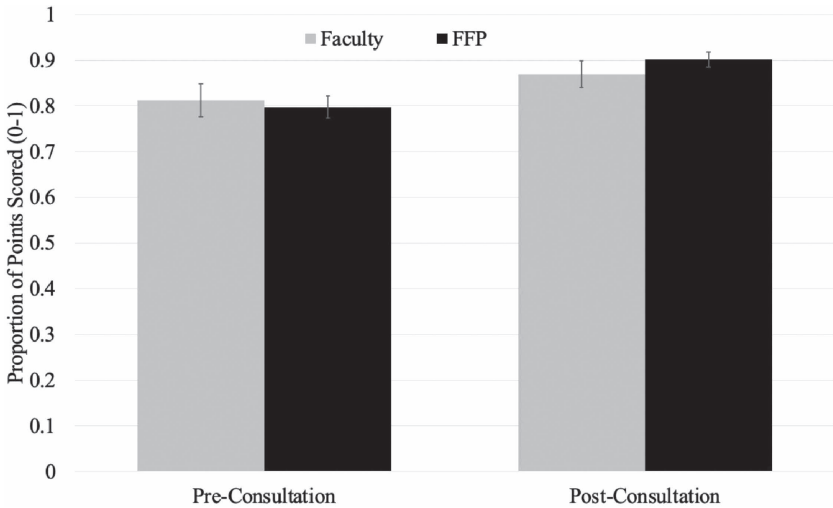
Figure 2. Faculty and FFP Clients' Mean Syllabus Pre- and Post-Consultation Scores (error bars represent 95% confidence intervals)

and client type (faculty, FFP) on pre-consultation scores $F(1,90) = .23$, $p = .63$ (Figure 3). Additionally, there were no significant main effects for either prior seminar attendance or client type on pre-consultation syllabus scores. Similarly, as reported above, prior seminar attendance did not explain the clients' changes in syllabus scores, either within or across client types (i.e., no statistically significant main effects or two- or three-way interactions including seminar condition).

Discussion

Our study directly investigated impacts of CTL services on current and future faculty's C&SD practices, as documented in their syllabi before and after a C&SD consultation. Below, we discuss the implications of our data and assessment approach for educational developers. We also discuss alternative data sources for assessing CTL impacts on instructional practices.
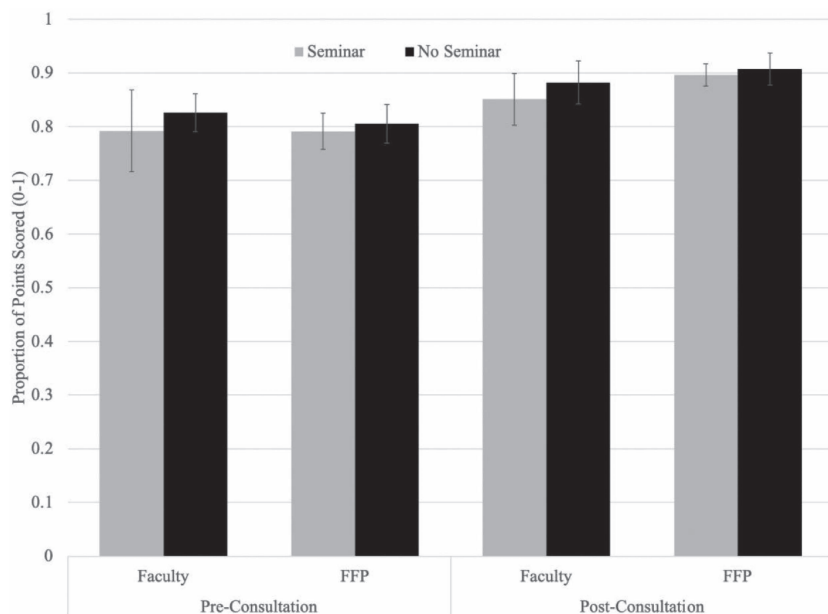
Figure 3. Faculty and FFP Clients' Mean Syllabus Pre- and Post-Consultation Scores and Seminar Attendance (error bars represent 95% confidence intervals)

Measuring Impacts of CTL Consultations on C&SD

What Changed After a Consultation?

Following a CTL consultation, both faculty and FFP clients demonstrated significant changes in C&SD practices (Tables 4 and 5), as measured by our syllabus rubric (Appendix A). Faculty and FFP clients' syllabi scored 31% and 50% higher after a consultation, respectively, relative to the ceiling of possible improvement, with FFP clients demonstrating significantly greater gains. We observed large effect sizes, suggesting changes of 0.73 and 1.04 standard deviations in faculty and FFP clients' syllabi, respectively. Both types of clients' syllabi changed significantly regarding assessments, assessment and grading policies, and other course policies. Following consultations, instructors

included more assessment details (e.g., format, deliverables, time-lines) and provided greater clarity regarding expectations and evaluation criteria. Furthermore, in post-consultation syllabi, instructors were more likely to provide a rationale for their policies. FFP clients' course descriptions and learning objectives also changed significantly. After consultations, course descriptions contained more information on the types of instructional methods students might experience and how the course might contribute to student development, their discipline, and/or their future career. Learning objectives also become more learner centered and measurable and represent a broader range of cognitive skills (e.g., application, synthesis and evaluation, rather than recall or comprehension alone).

Is the Consultation the Cause of the Observed Changes?

Together, our data from controls and comparison groups suggest that CTL consultations on C&SD can directly and positively impact instructors' C&SD practices. Specifically, our results suggest that observed client gains are not likely explained by ambient syllabus drift across semesters or by previous participation in a CTL seminar on C&SD.

Does the magnitude of change observed in clients' syllabi differ from that in syllabi of faculty who are not clients? Yes. Unfortunately, our comparison group for ambient syllabus drift was not large enough for a rigorous statistical analysis. However, non-clients' syllabi appear to exhibit little change across semesters compared to faculty clients' syllabi before and after a consultation (Figure 1). It is unlikely that the observed magnitude of changes in faculty clients' syllabi ($SD = 0.73$) are due to independent, ambient instructor revisions alone, even if they are a self-selected group.

Is the improvement observed in clients' syllabi after a consultation caused by the prior participation in related CTL programs? No. Prior attendance at a CTL seminar on C&SD did not explain variation in pre-consultation syllabi or changes in syllabi.

Do Our Data Suggest That CTL Seminars Do Not Impact Instructors?

No. Our data only suggest that a stand-alone C&SD seminar may not directly translate to syllabi impacts, especially after a delay. Our CTL's stand-alone 60–90 minute C&SD seminars are NOT designed to develop deep mastery but to introduce basic principles and (hopefully) lay a foundation for subsequent consultations. Seminars combine didactic and hands-on instruction. Following active learning exercises highlighting effective practices regarding learning objectives, assessments, alignment, policies, and inclusive climate, seminar participants may experience knowledge gains that do not transfer to syllabi creation without additional interventions, like those in a C&SD consultation or institute. Unlike course design institutes (Palmer et al., 2016) and consultations, seminar participants are not necessarily concurrently working on their C&SD, do not receive feedback on their syllabi, or collaboratively workshop C&SD elements with a trained consultant. Moreover, Palmer et al. (2016) collected syllabi generated by participants at the end of their week-long program. Our clients often (re) design syllabi and request a consultation more than one week after a seminar. Given these two differences between C&SD seminars and institutes, application of principles from seminars represents non-trivial transfer of learning. Based on this article's results, via authentic pre/post assessments, we plan to directly measure instructor learning gains regarding C&SD knowledge and skills in future iterations of C&SD seminars to better understand their impacts.

Why Did Syllabi Change More for FPP Than for Faculty Clients?

To our knowledge, this empirical study is the first to compare CTL consultation outcomes between current and future faculty. Two conspicuous differences between the service models are the most likely explanations for the observed differences (Figure 2). First, FFP consultations are more scaffolded, leveraging a rubric to provide feedback. FFP clients must then use this rubric-driven feedback to iterate upon

drafts until they meet minimum program requirements. The desire to complete program requirements may strongly motivate FFP clients. In contrast, faculty consultations do not necessarily leverage a rubric or involve iterating on drafts based on consultants' feedback. Faculty may demonstrate less change due to differing time constraints, priorities, or receptivity to adopting evidence-based practices. However, we hesitate to generalize regarding such differences among faculty and FFP clients without more data. Second, FFP consultations tend to focus exclusively on C&SD, but faculty consultations do not and likely exhibit less time on C&SD. One or both of these factors may drive the observed differences in consultation impacts on clients' syllabi. Third, for various reasons, some faculty clients have limited agency to alter course descriptions or learning objectives when they inherit a teaching assignment, possibly explaining why syllabi changed less for faculty than for FFP clients for these two syllabus elements.

Limitations

Our study has several limitations. First, we could not identify a viable comparison group for FFP clients. Second, for faculty, given the young age of our institution's syllabus registry, we identified only 10 courses meeting matching criteria that posted syllabi in consecutive semesters. Ideally, we would have preferred to include syllabus drift matches for all 32 faculty clients. Given the observed effect sizes for Research Question 1 and the observational data from the syllabus drift comparison group, we believe this limitation does not invalidate our conclusions. Third, because faculty rank was the last of four criteria in our matching protocol, our syllabus registry samples overrepresent senior faculty compared to our client sample (Table 2). In our opinion, this makes our exploration of Research Question 1 a conservative test. Although it is likely that our faculty client sample was less experienced than the registry sample, clients' syllabi still demonstrated higher initial quality and improvement over time. Finally, as discussed above, documentation of instructional practices in syllabi does not necessarily

guarantee implementation fidelity (Campbell et al., 2019). Further-more, syllabi may not capture certain features of evidence-based instructional practices (e.g., active learning) that may be influenced by CTL consultations, especially regarding classroom teaching strategies. In future studies, another data source, such as an observation of teaching or an alternative teaching artifact (see below), could help mitigate this limitation.

Please note that our study did not attempt to measure classroom teaching practices. Instead, we intentionally focused on C&SD as documented in syllabi to inform our services focused on C&SD and syllabi (as deliverables). Despite these limitations, we believe our data has implications for educational developers regarding outcomes assessments, resource allocation, and consulting practices.

Using Formative Assessment Data for Iterative Refinement

Our approach to formative assessment is motivated by our collective desire to iteratively refine our programs and services, informed by data from direct measures of outcomes. The data collected are not used to evaluate the performance of individual consultants. Instead, the primary purpose is to foster a team-based, growth mindset and data-informed collaborative search for areas for improvement. To establish and maintain buy-in as a team, CTL leadership sought consensus and staff input at every step, from rubric development to study design to data collection, analysis, and interpretation. Both rubric norming and data debrief sessions led to rich exchanges of consulting techniques and experiences as well as discussions about the implementation of service models.

To illustrate, the data from this study instigated deliberations on several changes to one or both service models. Foremost, discussions of the study rubric led to changes in how FFP consultations are scaffolded. The revised rubric more clearly aligns with our evidence-based objectives and priorities for clients' development. We made similar minor changes to our generative interview protocol often used in FFP consultations. The limitations of our data source (syllabi) inspired revisions

to the course design reflection memo that is part of the required FFP C&SD project. We clarified prompt questions to foster clients' reflections on instructional methods and their alignment with learning objectives, two important components of course design that are rarely captured fully in a syllabus. In future formative assessments, this additional deliverable for FFP clients can complement the data available in syllabi. Furthermore, we reevaluated when to provide the rubric during the FFP service model to best support clients and mirror what we recommend to instructors in general (i.e., provide the rubric in advance, rather than after receiving the client's first draft, as standard practice).

The difference in faculty and FFP clients' gains suggests that components of the FFP service model may be beneficial to incorporate into the faculty consultations. In particular, we discussed using the rubric or other resources to scaffold faculty consultations. Time constraints and/or competing priorities often prevent holistic feedback and iteration on faculty syllabi. Previously, consultants varied in what resources they used to support faculty clients. Some employed recommendations from our Faculty Senate regarding suggested syllabus components (Eckhardt, 2017). None reported leveraging the FFP rubric with faculty. Increasing the scaffolding in faculty consultations may enhance faculty clients' gains. However, increasing the focus of faculty consultations on particular growth areas highlighted in our data (e.g., assessments) may also be advantageous for "meeting clients where they are" and impacting C&SD, given the constraints discussed above.

Without this project, these team-wide conversations likely would not have happened, at least not in the same inclusive, systematic, data-driven way. Conspicuous benefits of the data-driven approach to reflective practice presented here include the ability to (a) more precisely measure the impact of a CTL service on instructors' course design practices; (b) leverage the power of group collaboration to generate data-informed, actionable steps to iteratively enhance a consultation service model; and (c) normalize dialogue across CTL staff members to share and enhance consultation practices and techniques. We acknowledge that this model requires both time and resources

and may need to be adjusted to be feasible in other CTL contexts. Nevertheless, we believe the potential benefits merit consideration as one option for formatively evaluating CTL services.

Recommendations for CTLs and Educational Developers

When evaluating CTL services using direct measures of outcomes, it can be difficult to attribute causality without a comparison group. Without comparison groups, outcomes assessments document "what is/happens" but not "what causes." And, as with our FFP client sample, finding a comparison group is non-trivial. Even without a comparison group, pre/post measurements can provide rich, evidence-based fodder for reflection and iterative refinement of CTL practices. However, one must interpret data cautiously regarding causality. Given these challenges, we advise investigating multiple data sources and/or outcomes measures to demonstrate the added value of CTL services.

What if a CTL has limited resources, including staff with assessment expertise? The approaches reported here would likely provide actionable, formative data with a smaller sample or when only using descriptive statistics or qualitative methods rather than parametric statistical analyses. Using study designs with large samples and inferential statistics to evaluate CTL impacts is rigorous. However, effective, data-driven, reflective practice is not limited to only that particular analytical "way of knowing." We recommend starting small with direct outcomes assessments and adopting a formative lens, if possible. CTLs can prioritize efforts by asking two questions: What would you most like to know about client outcomes that you don't? And what would you be willing to change in response to that data, if you had it?

Future Directions: Leveraging Alternative Teaching Artifacts as Data Sources

Our formative assessments using syllabus analyses greatly informed our educational development practices regarding C&SD. Consequently,

our reflective dialogues also explored other sustainable options for measuring impacts of CTL C&SD services that provide insights unavailable via syllabus analyses.

Classroom observations can provide systematic, objective data absent from syllabi (Campbell et al., 2019; Smith et al., 2013; Stanny et al., 2015; Wheeler & Bach, 2021), especially regarding classroom teaching strategies. For instance, observation data can directly measure frequency, duration, and timing of particular teaching strategies (e.g., active learning, inclusive teaching techniques) during synchronous course sessions. Moreover, trained observers can document fine-grained teaching and learning behaviors, including patterns of participation, the nature of instructor-student interactions, how feedback is delivered, student time on task, and more. Observational data is rich and actionable. However, to measure CTL outcomes, especially at scale, observations are a logistically daunting commitment when resources are limited, even with efficient and reliable observation protocols.

Analyses of other teaching artifacts, above and beyond syllabi, may provide complementary and less resource-intensive alternatives. Collecting assignment prompts and/or grading rubrics from clients before and after participating in CTL services could measure alignment of assessments with learning objectives in syllabi. A similar approach could be taken for analyzing the impacts of CTL services targeting the adoption of multimedia learning principles (Clark & Mayer, 2016) in the design of online learning modules, instructional videos, or lecture slides. Likewise, one could analyze exam item construction (Parkes & Zimmaro, 2016; Piontek, 2008), exam item analytics (Livingston, 2006), or rubric construction before and after CTL assessment design services. Regarding inclusive teaching, one could measure patterns in the representation of authors, perspectives, or ways of knowing in assigned course readings or materials. For all of the examples above, one could potentially develop evidence-based rubrics for formative outcomes assessment. To our knowledge, no SoED studies have included these data sources in CTL outcomes assessments. We hope future research will explore these possibilities and more.

Conclusion

Comprehensive CTL assessment plans should consider evaluating not only outcomes from programs and services but also organizational structure, resource allocation, and infrastructure (Ellis et al., 2020; POD, 2018). Furthermore, when strategically planning for CTL outcomes assessments, multiple sources of data from direct measures of outcomes are desirable because each tells a different part of the story and has its own pros and cons. Our study results suggest that (a) direct outcomes assessments based on artifacts of teaching are viable assessment approaches for CTLs; (b) CTL consultations appear to directly and positively impact clients' C&SD, as documented in syllabi, above and beyond natural, ambient changes across semesters; and (c) CTL one-off seminars do not conspicuously impact the starting point of C&SD consultations or their impacts. While finding viable comparison groups may be challenging, we argue that periodic, pre/post analyses of teaching artifacts are broadly transferable, sustainable approaches for formative, iterative refinements of CTL services. We hope that future research will build on our results, developing practical approaches to outcomes analyses of teaching artifacts as well as how best to directly measure other types of CTL impacts.

Acknowledgments

Biographies

Chad Hershock is the Director of Evidence-Based Teaching & DEI Initiatives at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Prior to 2013, he was an Assistant Director and Coordinator Science, Health Science, and Instructional Technology Initiatives at the Center for Research on Learning at the University of Michigan. His PhD is in Biology from the University of Michigan.

Laura Ochs Pottmeyer is a Data Science Research Associate at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in Science Education from the University of Virginia.

Jessica Harrell is a Senior Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in Rhetoric from Carnegie Mellon University.

Sophie le Blanc is a Senior Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in Political Science from the University of Delaware.

Marisella Rodriguez is a Senior Consultant at the Center for Teaching and Learning at the University of California, Berkeley. When data from this study was collected, she was a Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in Political Science from the University of California, Davis.

Jacqueline Stimson is a Senior Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie

Mellon University. Her PhD is in Classical Studies from the University of Michigan.

Katharine Phelps Walsh is a Senior Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in History from the University of Pittsburgh.

Emily Daniels Weiss is a Senior Teaching Consultant at the Eberly Center for Teaching Excellence and Educational Innovation at Carnegie Mellon University. Her PhD is in Chemistry from Carnegie Mellon University.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.

Beach, A. L., Sorcinelli, M. D., Austin, A. E., & Rivard, J. K. (2016). *Faculty development in the age of evidence: Current practices, future imperatives*. Stylus Publishing.

Campbell, C. M., Michel, J. O., Patel, S., & Gelashvili, M. (2019). College teaching from multiple angles: A multi-trait multi-method analysis of college courses. *Research in Higher Education*, *60*(5), 711–735. https://doi.org/10.1007/s11162-018-9529-8

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*(1), 129–145. https://doi.org/10.1002/j.2334-4822.2012.tb00678.x

Clark, R. C., & Mayer, R. E. (2016). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). Wiley.

Cullen, R., & Harris, M. (2009). Assessing learner-centredness through course syllabi. *Assessment & Evaluation in Higher Education*, *34*(1), 115–125. https://doi.org/10.1080/02602930801956018

Doolittle, P. E., & Siudzinski, R. A. (2010). Recommended syllabus components: What do higher education faculty include in their syllabi? *Journal on Excellence in College Teaching*, *21*(3), 29–61.

Eberly Center. (2021, December 10). *Learn about our Future Faculty Program*. Eberly Center for Teaching Excellence and Educational Innovation. https://www.cmu.edu/teaching/graduatestudentsupport/futurefacultyprogram.html

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

Eckhardt, D. (2017, April 4). *Syllabus Best Practices Group, resolutions*. Faculty Senate Meeting 9. Details omitted for reviewing.

Ellis, D. E., Brown, V. M., & Tse, C. T. (2020). Comprehensive assessment for teaching and learning centres: A field-tested planning model. *International Journal for Academic Development*, *25*(4), 337–349. https://doi.org/10.1080/1360144X.2020.1786694

Finelli, C. J., Ott, M., Gottfried, A. C., Hershock, C., O'Neal, C., & Kaplan, M. (2008). Utilizing instructional consultations to enhance the teaching performance of engineering faculty. *Journal of Engineering Education*, *97*(4), 397–411. https://doi.org/10.1002/j.2168-9830.2008.tb00989.x

Finelli, C. J., Pinder-Grover, T., & Wright, M. C. (2011). Consultations on teaching: Using student feedback for instructional improvement. In C. E. Cook & M. Kaplan (Eds.), *Advancing the culture of teaching on campus: How a teaching center can make a difference* (pp. 65–79). Stylus Publishing.

Fink, L. D. (2013). *Creating significant learning experiences: An integrated approach to designing college courses* (Rev. and updated ed.). Jossey-Bass.

Haras, C., Taylor, S. C., Sorcinelli, M. D., & von Hoene, L. (Eds.). (2017). *Institutional commitment to teaching excellence: Assessing the impacts and outcomes of faculty development*. American Council on Education.

Hershock, C., Cook, C. E., Wright, M. C., & O'Neal, C. (2011). Action research for instructional improvement. In C. E. Cook & M. Kaplan (Eds.), *Advancing the culture of teaching on campus: How a teaching center can make a difference* (pp. 167–182). Stylus Publishing.

Hines, S. R. (2011). How mature teaching and learning centers evaluate their services. *To Improve the Academy*, *30*(1), 277–289. https://doi.org/10.1002/j.2334-4822.2011.tb00663.x

Hines, S. R. (2015). Setting the groundwork for quality faculty development evaluation: A five-step approach. *Journal of Faculty Development*, *29*(1), 5–12.

Homa, N., Hackathorn, J., Brown, C. M., Garczynski, A., Solomon, E. D., Tennial, R., Sanborn, U. A., & Gurung, R. A. R. (2013). An analysis of learning objectives and content coverage in introductory psychology syllabi. *Teaching of Psychology*, *40*(3), 169–174. https://doi.org/10.1177/0098628313487456

Kreber, C., Brook, P., & Educational Policy. (2001). Impact evaluation of educational development programmes. *International Journal for Academic Development*, *6*(2), 96–108. https://doi.org/10.1080/13601440110090749

Kucsera, J. V., & Svinicki, M. (2010). Rigorous evaluations of faculty development programs. *The Journal of Faculty Development*, *24*(2), 5–18.

Livingston, S. A. (2006). Item analysis. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 421–441). Lawrence Erlbaum Associates.

McGowan, B., Gonzalez, M., & Stanny, C. J. (2016). What do undergraduate course syllabi say about information literacy? *portal: Libraries and the Academy*, *16*(3), 599–617.

Meizlish, D. S., Wright, M. C., Howard, J., & Kaplan, M. L. (2018). Measuring the impact of a new faculty program using institutional data. *International Journal for Academic Development*, *23*(2), 72–85.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Palmer, M. S., Bach, D. J., & Streifer, A. C. (2014). Measuring the promise: A learning-focused syllabus rubric. *To Improve the Academy*, *33*(1), 14–36. https://doi.org/10.1002/tia2.20004

Palmer, M. S., Streifer, A. C., & Williams-Duncan, S. (2016). Systematic assessment of a high-impact course design institute. *To Improve the Academy*, *35*(2), 339–361. https://doi.org/10.3998/tia.17063888.0035.203

Parkes, J., & Zimmaro, D. (2016). *Learning and assessing with multiple-choice questions in college classrooms*. Routledge.

Piontek, M. (2008). *Best practices for designing and grading exams* (Occasional Paper No. 24). Center for Research on Learning and Teaching, University of Michigan.

POD Network. (2018). *Defining what matters: Guidelines for comprehensive center for teaching and learning (CTL) evaluation*. https://podnetwork.org/content/uploads/POD_CTL_Evaluation_Guidelines__2018_.pdf

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

Stanny, C., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education*, *40*(7), 898–913. https://doi.org/10.1080/02602938.2014.956684

Stes, A., Min-Leliveld, M., Gijbels, D., & Van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the

research. *Educational Research Review*, *5*(1), 25–49. https://doi.org/10.1016/j.edurev.2009.07.001

Tomkin, J. H., Beilstein, S. O., Morphew, J. W., & Herman, G. L. (2019). Evidence that communities of practice are associated with active learning in large STEM lectures. *International Journal of STEM Education*, *6*(1), 1–15. https://doi.org/10.1186/s40594-018-0154-z

Wheeler, L. B., & Bach, D. (2021). Understanding the impact of educational development interventions on classroom instruction and student success. *International Journal for Academic Development*, *26*(1), 24–40, https://doi.org/10.1080/1360144X.2020.1777555

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Pearson.

Appendix A: Future Faculty Program Course & Syllabus Design Rubric

Category	Category description	Ratings		
Course description – What will students learn (i.e., knowledge, skills, attitudes, as opposed to topics)? Why will this matter to students? How will the course help students develop as scholars, learners, future professionals? What will students experience in the course? What are the instructional methods, and how will they support student learning?				
Course Description 1 (CD1)	Basic information about what students will learn in this course	Excellent (3)	Needs Improvement (2)	Poor (1)
Course Description 2 (CD2)	How this learning is significant to student development, their discipline, and/or their future career	Excellent (3)	Needs Improvement (2)	Poor (1)
Course Description 3 (CD3)	Brief information about instructional methods and what students can expect to experience in the course	Excellent (3)	Needs Improvement (2)	Poor (1)
Learning objectives – What, specifically, will students be able to do or demonstrate once they've completed the course?				
Learning Objective 1 (LO1)	Are student centered	All (3)	Some (2)	None (1)
Learning Objective 2 (LO2)	Are measurable	All (3)	Some (2)	None (1)
Learning Objective 3 (LO3)	Represent a range of cognitive skills and complexity		Present (2)	Absent (1)
Assessments				
Assessment 1 (A1)	The alignment between assessments and learning goals is clear.	Excellent (3)	Needs Improvement (2)	Poor (1)

Category	Category description	Ratings		
Assessment 2 (A2)	Opportunities for formative (e.g., quizzes, homework, in-class exercises, reading questions) and summative assessment are included (e.g., high stakes/low stakes, drafts, scaffolding).	Present (2)		Absent (1)
Assessment 3 (A3)	Information about the timing of the assessments is included.	All (3)	Some (2)	None (1)
Assessment 4 (A4)	Expectations for student deliverables (what are they actually going to do) are explicit.	Excellent (3)	Needs Improvement (2)	Poor (1)
Assessment 5 (A5)	Criteria for evaluation are included.	Present (2)		Absent (1)
Assessment Element 6 (A6)[+]	Participation grade, if present, is described such that a student would understand what they need to do in order to achieve full participation credit.	Excellent (3)	Needs Improvement (2)	Poor (1) N/A

Assessment and grading policies – How will grades be calculated for the course? Will students have the opportunity to drop any scores or submit revised drafts? How does recitation/lab/discussion/etc. factor into grading? Will different types of assignments be graded differently? What are your policies for late work, re-grades, makeups?

Category	Category description	Ratings		
Grading Policies 1 (GP1)	Grade breakdown is clear and easy to understand.	Excellent (3)	Needs Improvement (2)	Poor (1)
Grading Policies 2 (GP2)	Relevant policies including makeups, late work, re-grades, etc. are described.	Excellent (3)	Needs Improvement (2)	Poor (1)

Other course policies

Category	Category description	Ratings		
Other Policies 1 (OP1)	Accommodations	Excellent (3)	Needs Improvement (2)	Poor (1)
Other Policies 2 (OP2)	Student wellness	Present (2)		Absent (1)
Other Policies 3 (OP3)	Academic integrity	Excellent (3)	Needs Improvement (2)	Poor (1)

(Continued)

(*Continued*)

Category	Category description	Ratings			
Other Policies 4 (OP4)[+]	Other policies: If present, the policy and procedure are clear (e.g., attendance, laptops, email, recording lectures).	Excellent (3)	Needs Improvement (2)	Poor (1)	N/A
Other Policies 5 (OP5)[+]	Other policies: If present, has an explanation.	Excellent (3)	Needs Improvement (2)	Poor (1)	N/A
Organization					
Organization 1 (O1)	Is the essential information easy to locate? Are the sections of the syllabus presented in cohesive order?	Excellent (3)	Needs Improvement (2)	Poor (1)	

[+]This category is optional and was only scored when applicable.