



**FORM FOR SUBMITTING FULL PROPOSALS FOR CONSIDERATION
FOR THE
2014 ROBERT J. MENGES HONORED PRESENTATION AWARD**

Instructions:

- Boxes will expand to accommodate text
- Total word count must not exceed 2000 words for the body of the proposal (excludes appendices).
- Please 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.

YOUR NAME:	INSTITUTION:	EMAIL:
Deborah Meizlish, Mary Wright, Joe Howard, Matthew Kaplan	Center for Research on Learning and Teaching, University of Michigan	debmeiz@umich.edu
TITLE:	Leveraging Institutional Data to Demonstrate our Impact	

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

Outreach to new faculty is considered one of the most important activities that teaching centers undertake (Sorcinelli, Austin, Eddy & Beach, 2006). Through these efforts, centers provide important support to new faculty in their teaching roles, helping them successfully start their teaching careers while connecting them to sources of ongoing educational development. However, documentation of the value of new faculty programs has been difficult to establish, given the many challenges of doing evaluation of educational development programs, such as time, resources, and nature of the work (Chism, Holley & Harris, 2012). Indeed, lack of good evaluation models has been established as one key challenge to doing effective evaluation (Hines, 2009, 2011). Therefore, the findings of this research will be of interest to other educational developers who seek an evidenced-based justification for developing or sustaining a new faculty program. More importantly, this project presents a model for doing a rigorous faculty development evaluation, with primary reliance on extant institutional data (e.g., teaching center records and student ratings). This evaluation model offers other centers a resource-efficient approach for illustrating the impact of their educational development initiatives. To that end, it is consistent with POD’s strategic plan (IV.2) to “provide resources to help members to regularly assess programs, services, and resources.”

In this project we examine the impact of a year-long new faculty program. Participation in the program (which was established in 2009 in our university’s largest college) is required of all new tenure-track faculty. New faculty participate in a one-and-a-half day orientation program that includes pedagogy sessions, course planning conversations with experienced

faculty, and micro-teaching. Once the academic year is underway, programming varies but always includes a midterm student feedback (or SGID) consultation during each faculty member's first term of teaching.

The goals for the program are many and include:

- Introducing faculty to University students and University resources on teaching and learning;
- Deepening faculty understanding of how students learn; and
- Facilitating discussion about effective teaching strategies.

The program has been very well received by the new faculty. For example, when asked to indicate how much they agreed with the statement "The program was valuable," the mean response was a 4.23 on a 5-point scale (scaled from 'strongly disagree' to 'strongly agree'). That it has been so well received is particularly heartening given the diverse constituencies in this required program, which not only includes faculty from philosophy to physics but also faculty with widely varying amounts of prior teaching experience. However, satisfaction can not be the only measure, especially given the institutional time and resources devoted to the program. Thus we embarked on this research project to answer the following questions:

1. Do faculty who participate in the program report improvements in their preparation for college teaching?
2. Do faculty who participate in the program engage in more educational development activities than comparable faculty who did not participate in the program?
3. Do faculty who participate in the program receive higher student ratings than comparable faculty who did not participate in the program?

WORD COUNT

503

2. DESCRIPTION OF RESEARCH DESIGN:

Our research design has three major features. First, we use a pre/post survey methodology to investigate how participation in the program influenced participants' self-assessments of their preparation for a range of pedagogical tasks. Table 1 presents the scale and items we used to create our composite score.

Second, we took advantage of a unique opportunity to have a "control" group for a faculty development (FD) program. Because this FD program was required of all new tenure-track faculty, we had a natural control group formed by the cohorts that immediately preceded

the establishment of the program. For the purposes of this project, our control group includes all assistant professors who joined the college in the three years prior to the establishment of the program. We compare these faculty to the first three cohorts of faculty in the program. In doing so, we always compare faculty at the same stages in their careers. That is, we compare faculty at the conclusion of their first, second, or third years at University X (see Tables B and C). To check on the validity of this comparison, Table D presents the breakdown of faculty by the disciplinary division and level of the courses they taught. With roughly comparable groups, we felt comfortable deploying this comparison to answer our second and third research questions.

Finally, in examining the impact of the program on participants' student ratings scores, we use a multilevel regression model with mixed effects. This model enables us to examine program impact while controlling for other known factors that influence student evaluations.

Regarding data sources:

Our measure of continued participation in educational development is derived from our Center's own service database. We have a long-standing practice of tracking all faculty who participate in our consulting services, programs, and grants. Admittedly this is an imperfect measure, missing as it does the range of other educational development activities that faculty could be participating in. However, from our Center's perspective, one goal of this program is to connect new faculty to the ongoing professional development opportunities that we offer. At a minimum, this measure helps us to understand whether such a connection is in fact happening.

To measure impact on students, we utilize data from our institution's student ratings system. Although student ratings are one measure of instructional effectiveness and reported student learning, they have been found to correlate with student performance and peer/expert course evaluations (Arreola, 2000; Baleche, Fairris & Marks, 2012). Our examination of student ratings utilized student responses to three items:

- Overall, this was an excellent course;
- Overall the instructor was an excellent teacher; and
- I learned a great deal from this course.

In order to ensure a robust analysis, we calculate what we call "e-scores" for each faculty member. In brief, for each faculty member, we average their scores on each student rating item across all the courses they have taught. But this is not a straight average. Instead, we use a weighted average based on class size. Our rationale for this approach is twofold. First, to ensure independent measures, we want each faculty member to appear only once in each analysis. That requires a method for combining their ratings across courses. Second, from an institutional perspective, the impact of any faculty member in part rests on the number of students they teach. Using class size to weight these averages enables us to

better account for the impact of the program on the entire student body reached by these faculty.

WORD COUNT

580

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

There are limited studies of the impact of new faculty programs, which is striking given these programs' reported importance in teaching center initiatives (Sorcinelli, Austin, Eddy & Beach, 2006). To illustrate, Stes, Min-Leliveld, Gijbels, & Van Petegem's (2010) very comprehensive review of educational development impact research found only one evaluation of an initiative targeting new faculty -- and it was published in 1977! Clearly, documenting the value of these programs has been a challenge for the field (Chism, Holley & Harris, 2012; Hines, 2009, 2011).

Lack of good evaluation models has been established as one key challenge to doing effective evaluation (Hines, 2009, 2011). The study presented here presents a model to overcome these gaps and address key goals for evaluation of educational development initiatives, namely participant satisfaction, impacts on participants and learners, and institutional outcomes (Chism & Szabó, 1998; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010). Given documented challenges to evaluation of faculty development initiatives, our study's emphasis on the analysis of extant data is important. We present a resource-efficient evaluation model that is applicable to other educational development initiatives and centers' evaluations. Trigwell, Rodriguez & Han (2012) give one example of this approach in an evaluation of a year-long certificate program, using student course satisfaction measures, as well as teaching awards and grants, as key metrics of the program's impact. We also use student ratings (albeit items pertaining to reported learning), but we broaden our perspective away from measuring solely "high profile" metrics like awards and grants to more routine, often taken-for-granted, signals of commitment to teaching, like participation in consultations, workshops, and early feedback.

WORD COUNT

267

4. FINDINGS: THEIR SIGNIFICANCE & LIMITATIONS:

Research Question 1: Do faculty who participate in the program report improvements in their preparation for college teaching? Yes. As the data in Table E reports, there is a statistically significant increase in participants' self-reported sense of preparation for teaching. This is particularly noteworthy, given the relatively narrow scale used in this measure.

Research Question 2: Do faculty who participate in the program engage in more educational development activities than do comparable faculty who did not participate in the program? Yes, as the data in Table F illustrates, program participants sought out more Center-sponsored professional development opportunities than faculty who joined the University prior to the program's inception. This finding is statistically significant in Year 1 and Year 3.

Research Question 3: Do faculty who participate in the program receive higher student ratings than do comparable faculty who did not participate in the program? Yes. Tables G through J present our bivariate analyses while Table K presents our multivariate analysis. As Table G reports, program participants' scores on the student ratings items show a statistically significant increase as compared to our control group. While the difference was most pronounced in Year 1 (see Table H), they persist in the right direction in our Year 2 and Year 3 analyses (Tables I and J). Turning to our multivariate analysis (Table K), our model includes program participation as well as other variables known to influence student ratings for instruction. For example, we control for disciplinary division, course level, and prior motivation to take the course (measured by student responses to the question "I had a strong desire to take this course"). A positive and statistically significant impact of the program continues to hold (particularly on Q2: Overall the instructor was an excellent teacher and Q3: I learned a great deal from this course), even once controls for these other factors are introduced.

Discussion:

To recap, these results document positive impacts associated with a program for new faculty. Not only were faculty satisfied with the program but they reported increases in confidence for a range of teaching tasks. When compared with earlier cohorts of faculty, program participants were also more likely to take advantage of professional development opportunities offered through the teaching center and received stronger student ratings.

At our institution, presentations of these data have further strengthened administrative buy-in for this program. But we believe its import stretches beyond our local context. As we discussed in our Introduction, there have been few substantial evaluations of new faculty programs. This research adds an important new element to the literature and can

help our faculty development colleagues further justify similar work on their own campuses. Further, our methodology, utilizing a naturally occurring control group and leveraging existing institutional data, can serve as a model for future evaluations of faculty development work. Finally, this data can help provide benchmarking data on the degree of impact we might expect from faculty development programs. For example, while our multilevel regression indicates a positive and statistically significant association between student ratings and participation in the program, the magnitude of the effect is smaller than several other items in our model (e.g., disciplinary division and student motivation).

This work also suggests important future directions. At a minimum, similar research would be enhanced by leveraging an even longer time horizon. Do these differential effects last or do they erode over time? Thinking more broadly, it is important to note that this new faculty program combines many approaches (e.g., pedagogy seminars, micro-teaching, structured short-term mentoring on teaching, consultations, and midterm student feedback). As designed, this evaluation cannot disentangle the relative impact of each component. Future work could productively address these issue.

WORD COUNT
610

TOTAL WORD COUNT FOR ALL 4 SECTIONS: 1960
--

(Include appendices, below)

References

Arreola, R. A. (2000). *Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system* (2nd ed.). Bolton, MA: Anker.

Beleche, T., Fairris, D., & Marks, M. (2012). Do course evaluations truly reflect student learning? Evidence from an objectively graded post-test. *Economics of Education Review*, 31: 709-719.

Chism, N., & Szabó, B. (1998). How faculty development programs evaluate their services. *Journal of Staff, Program, and Organization Development*, 15 (2), 55-62.

Chism, N.V.N., Holley, M., & Harris, C.J. (2012). Researching the impact of educational

development: Basis for informed practice. In J. E. Groccia & L. Cruz, Eds. *To Improve the Academy* (pp. 129-145). San Francisco: Jossey-Bass.

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(3), 5-19.

Hines, S. R. (2011). How mature teaching and learning centers evaluate their services. In J.E. Miller & J. E. Groccia, Eds. *To Improve the Academy* (pp. 277-289). San Francisco: Jossey-Bass.

Sorcinelli, M. D., Austin, A. E., Eddy, P. L., & Beach, A. L. (2006). *Creating the future of faculty development: Learning from the past, understanding the present*. San Francisco: Jossey-Bass.

Stes, A., Min-Leliveld, M., Gijbels, D., & Petegem, P.V. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review*, 5: 25-49.

Trigwell, K., Rodriguez, K. C., & Han, F. (2012). Assessing the impact of a university teaching development programme. *Assessment & Evaluation in Higher Education*, 37(4): 499-511.

Appendix A: Supporting Tables

Table A: Pre and Post Items Associated with the Question "How prepared do you feel for the following tasks...?" (4=Very Prepared, 3=Prepared, 2=Unprepared, 1=Very Unprepared)

Designing and planning a course
Preparing lectures
Motivating diverse students
Utilizing active learning techniques in lecture
Utilizing course management tools
Supervising graduate student instructors
Teaching graduate students
Mentoring graduate students

Table B: Cohort Year and Population by Group

Control Group	New Faculty Program Group
2006 - 2007 n=32	2009 - 2010 n=40
2007 - 2008 n=36	2010 - 2011 n=29
2008 - 2009 n=31	2011 - 2012 n=20
TOTAL n=99	TOTAL n=89

Table C: Years of Teaching per Cohort; Shaded Areas included in Analyses

Year 6					
Year 5	Year 5				
Year 4	Year 4	Year 4			
Year 3	Year 3	Year 3	Year 3	Year 3	
Year 2	Year 2	Year 2	Year 2	Year 2	Year 2
Year 1	Year 1	Year 1	Year 1	Year 1	Year 1
2006-7	2007-8	2008-9	2009-10	2010-2011	2011-2012
Control Group			New Faculty Program Group		

Table D: Comparison of Courses Taught by Disciplinary Division and Level, by Group

	By Disciplinary Division		By Level		
	Control Group	New Faculty Program Group		Control Group	New Faculty Program Group
Humanities	30%	36%	Lower Division	24%	27%
Social Science	40%	38%	Upper Division	44%	49%
Natural Science	30%	26%	Master's	10%	6%
			Doctoral	22%	22%

Table E: Change in Mean Composite Confidence Score for New Faculty Program Group

	Pre	Post	Difference
Composite Confidence Score	2.8	3.1	+0.3****

Items used to construct scale are presented in Table A.

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table F: Comparison of Average Contacts with Teaching Center per Faculty Member by Year of Teaching (excludes any required contacts)

Year of Teaching	Control Group	New Faculty Program Group	Difference
In Year 1	1.40	1.99	+42%*
In Year 2	0.78	0.83	+6%
In Year 3	0.49	0.86	+76%*

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table G: Comparison by Group of median “E-Scores” across First Three Years of Teaching (Scale: 5=Strongly Agree...1=Strongly Disagree)

Question	Control Group	New Faculty Program Group	Difference
Overall this was an excellent course	4.08 (.30)	4.22 (0.38)	+0.14**
Overall this was an excellent teacher	4.29 (0.37)	4.46 (0.39)	+0.17**
I learned a great deal from this course	4.24 (0.25)	4.39 (0.27)	+0.15***

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table H: Comparison by Group of median “E-Scores” – Teaching Year 1 (Scale: 5=Strongly Agree...1=Strongly Disagree)

Question	Control Group	New Faculty Program Group	Difference
Overall this was an excellent course	4.11 (0.40)	4.29 (0.61)	+.18*
Overall this was an excellent teacher	4.30 (0.46)	4.53 (0.60)	+.23*
I learned a great deal from this course	4.27 (0.32)	4.47 (0.06)	+0.20**

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table I: Comparison by Group of median “E-Scores” - Teaching Year 2
(Scale: 5=Strongly Agree...1=Strongly Disagree)

Question	Control Group	New Faculty Program Group	Difference
Overall this was an excellent course	4.06 (0.37)	4.17 (0.46)	+ .11~
Overall this was an excellent teacher	4.25 (0.46)	4.40 (0.48)	+ .15*
I learned a great deal from this course	4.23 (0.35)	4.35 (0.37)	+ .12*

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table J: Comparison by Group of median “E-Scores” - Teaching Year 3

Question	Control Group	New Faculty Program Group	Difference
Q1: Overall this was an excellent course	4.15 (0.37)	4.28 (0.42)	+ .13~
Q2: Overall this was an excellent teacher	4.37 (0.39)	4.49 (0.40)	+ .12~
Q3: I learned a great deal from this course	4.31 (0.30)	4.42 (0.30)	+ .11*

~p<0.10, *p<0.05, **p<0.01, ***p<0.001

Table K: Multilevel Regression Model with Mixed Effects: Predicted Score on Student Evaluation Items Q1, Q2, & Q3.

Coefficient	Q1: Overall this was an excellent course	Q2: Overall this was an excellent teacher	Q3: I learned a great deal from this course
Intercept	3.92	4.27	4.11
New Faculty Program? (1=Yes)	0.08~	0.09*	0.08*
Disciplinary Division (vs. Humanities)			
-Natural Science	-0.18***	-0.25***	-0.14***
-Social Science	-0.05	-0.07	-0.01
Course Level (vs. Lower Division)			
Upper Division	0.06	0.04	0.02
Master’s (500)	0.07	0.04	0.07
Doctoral (600+)	0.13*	0.11*	0.06
Item Q4 High (1=Yes)	0.49***	0.34***	0.40***

~p<0.10, *p<0.05, **p<0.01, ***p<0.001