Beyond Satisfaction: Toward an Outcomes-Based, Procedural Model of Faculty Development Program Evaluation

D. Christopher Brooks, Lauren Marsh, Kimerly Wilcox, & Brad Cohen
University of Minnesota

In response to the well-documented need for rigorous evaluations of faculty development programs and increasing demands for institutional accountability, University of Minnesota's Office of Information Technology (OIT) researchers have developed an approach to program evaluation that assesses individual level changes to participants' attitudes, values, and behaviors on defined programmatic outcomes. Following an overview of the program in which the new approach was piloted, we detail the process used to develop the instrument, the data collection methods, and the analysis of the results. Additionally, results of reliability tests are presented with suggestions for how the instrument may be improved in future iterations.

In 2009, EDUCAUSE – the leading organizational advocate for the thoughtful application of educational technologies in higher education – identified faculty development programs geared towards encouraging innovative teaching and learning practices as one of the most pressing challenges facing higher education today. In addition to the basic instructional strategies that comprise the vast majority of faculty development programs in the area of information technology (IT), EDUCAUSE has recommended that new areas of focus, such as deep integration of technology into instructional design and assessment of the impact of delivery modes on student learning outcomes, take priority to create a faculty with twenty-first century competencies. Given that these resource-intensive recommendations coincide with increased austerity measures due to severe budget cuts in higher education, the demands for accountability render systematic and rigorous assessment of faculty development program outcomes more necessary and urgent than ever before (Diaz, et al. 2009).

Those tasked with developing robust and reliable measures of faculty development program outcomes quickly discover that there is a dearth of both scholarly work and practical examples of best practices, assessment instrumentation, and empirical research upon which they can draw. Finding the early literature on evaluating faculty development programs lacking, Hoyt and Howard (1978) feared that such programs would "suffer the fate of other educational fads which are born, mature, and die without fair trial or serious study" and urgently called for improvements in and expansions of evaluation protocols (p. 36). Levinson-Rose and Menges' (1981) survey of the field found the number and quality of studies on the evaluation of faculty development programs much improved, but still lacking in the theoretical and methodological rigor required for the advancement of the field. Nearly thirty years after the explosion of faculty development programs, Chism and Szabó (1997) found that while program evaluation was used ubiquitously to evaluate services and document successes, such instruments overwhelmingly gathered data on user satisfaction instead of substantive measures related to program effectiveness in meeting stated objectives. More recently, Hines (2009) found that little had changed, in that program assessment continues to be pervasive but with a primary focus on readily accessible measures such as user satisfaction. Kuscera and Svinicki (2010) concur stating that, "faculty development has not progressed in honing its evaluation practices much beyond the early 1990s" (p. 9).

The lack of progress in the field of faculty development evaluation appears to result from issues related to both supply and demand for innovative change and methodological rigor. On the supply-side, lack of time, resources, and expertise on issues related to instrument development and measurement are cited frequently as reasons that more simple measures of response rates, program satisfaction, and self-reported change are presented as evidence of success. On the demand-side, it appears that a lack of cooperation among participants and an underdeveloped culture of assessment tend to preclude the possibility of making significant strides towards linking changes that occur *within* a faculty development program

directly to actual changes in teaching practices and student learning outcomes (Chism and Szabó, 1997; Levinson-Rose and Menges, 1981; Hines, 2009).

We think that the solution to this disequilibrium in faculty development program evaluation lies in shifting the theoretical and methodological approaches employed by program managers and researchers. First, it is necessary to use variables that move beyond "superficial levels of experience" towards ones that measure individual differences among participants and are comparable in definition and operationalization across programs (Levinson-Rose and Menges, pp. 418-419). Second, faculty development program evaluation needs to be conducted with a methodological rigor that can withstand the scrutiny of peer review. Such a rigorous approach includes, at a minimum, the best practices identified by Hines (2009) and include 1) taking a systematic approach that is 2) directed by the goals of the program that are 3) measurable with 4) *a priori* benchmarks for determining success and that 5) employ appropriate assessment methods using 6) multiple measures to collect both 7) summative and formative data that 8) can be linked causally to the substance and process of the program itself.

In the summer of 2008, the University of Minnesota's Office of Information Technology (OIT) began the process of developing a systematic and rigorous method of assessing the impact of three of its more resource-intensive faculty development programs. Our approach was designed to respond to the deficiencies identified by previous researchers by embracing the theoretical correctives and methodological rigor of the recommendations mentioned previously. On the one hand, by committing institutional resources to and cultivating in-house expertise on measurement, we developed and tested an instrument that affords the opportunity to track over time the impact of participation on individuals participating in our faculty development programs for stated programmatic goals. On the other, by recognizing that claims of causality between faculty development programs and specific classroom outcomes is theoretically and methodologically tenuous, at best, we de-coupled evaluating individual participant transformation from the implementation of lessons and practices learned. That is, to evaluate the impact of our faculty development programs on individual participants, we assess changes to their individual attitudes, beliefs, and behaviors; to evaluate the long-term impact of our programs on their teaching, we collaborate with individual participants on research projects that assess the impact of pedagogical changes prompted by their experiences in our programs.

This article serves the purposes of 1) delineating how OIT researchers approached developing an instrument to measure the impact of the Faculty Fellowship Program (FFP) on its participants, parts of which could be used in assessing the impact of other OIT faculty development programs, and 2) reporting the results of this piloted evaluation project to a broader audience. The overall results suggest that the FFP was successful in accomplishing its goals with considerable gains in key areas being made by the participants. Furthermore, scale reliability tests reveal that OIT researchers were largely successful in measuring the constructs identified as program goals at the onset of the program. Despite some limitations related to the size of the study and the need to improve upon the piloted instrument for future iterations, we conclude that the processes detailed here make significant progress towards realizing necessary changes to faculty development program evaluation by providing a replicable and rigorous model.

OIT Faculty Fellowship Program Overview

The FFP began in 2000 as a means of cultivating faculty leadership in the area of technology-enhanced learning (TEL) at the University of Minnesota, Twin Cities campus. Faculty and instructors from across the disciplines apply to the program by proposing a TEL project on which they wish to work. The faculty who are accepted into the program receive awards of $10,000 and meet regularly throughout the course of the fellowship in a seminar-like environment. While the cohort experience remains the heart of the program, over the years it has adapted to reflect changes that include the following:

- The development and professionalization of the learning technologies field, evidenced by a growing number of journals, conferences, and professional organizations;
- The convergence of best practices in both technology and disciplinary pedagogy;
- A growing emphasis on the scholarship of teaching and learning and its integration into tenure and promotion protocols;
- A demand for accountability that includes local and national discussions about student learning outcomes for higher education; and
- An increasing emphasis on the alignment of classroom, departmental, and university outcomes.

Throughout its many iterations, the program has remained an opportunity for motivated faculty to meet as a group, share progress and receive feedback on individual projects, develop knowledge and skills around teaching with technology, and cultivate an understanding of the university infrastructure through which technologies are vetted and faculty are supported.

Our current program's structure is motivated by OIT's and the University of Minnesota's desire for

a partnership with informed faculty to provide input around issues of the changing TEL landscape that affect faculty in their efforts to provide the best possible learning experiences for their students. This program fosters a multidisciplinary learning community that explores possibilities and best practices in technology-rich learning environments, produces scholarship in this area, generates organizational awareness, and advances faculty leadership around these issues. The iteration of the program addressed in this paper represents a significant departure from previous programs in that the fellows made a commitment to a collaborative project built around the theme, "What does the University of Minnesota need to do to become a 21st-century leader in the area of emerging learning environments?" To provide additional time to address the individual TEL projects and the collaborative work, the program was expanded from two semesters to 18 months.

This fellowship experience began with an intensive three-day workshop intended to ground fellows in instructional design, introduce the theme and structure of this multifaceted program, familiarize them with the tools that would support us, and challenge them to think about the scholarship of teaching and learning (SoTL) and nontraditional forms of publication supported by technology. Thereafter, we met as a group once a month; consultants met monthly with individual fellows to support work on their individual projects.

It has always been an expectation of the program that fellows would share their knowledge and experience, even in the formative stages, with the university, usually through a public presentation. Changing technologies as well as alternatives to traditional publication continue to present new opportunities for formal and informal presentation to audiences in and beyond the university. As of this writing, the fellows have participated in video interviews that can be purposed in numerous ways; maintained a blog, "Conversations about Emerging Learning Environments"; organized an event for peers and campus leaders called, "Innovation and Transformation in a Time of Limited Resources - A Conversation"; and have published a multimedia article in *EDUCAUSE Quarterly* (Solheim, Longo, Cohen, and Dikkers, 2010).

During the 2008-2009 FFP, OIT faculty development consultants who co-manage the program partnered with OIT research and evaluation fellows to develop the prototype instrument described in this paper and to support the faculty fellows in their efforts to develop and implement evaluation plans for their projects.

Instrument Development

At the onset of the OIT Program Evaluation Project, three primary goals related to the development of objective data collection instruments were identified. We sought to 1) derive instrument components deductively from programmatic goals, 2) move beyond measures of satisfaction to ones that measure individual-level changes in the attitudes, values, and behaviors of participants, and 3) develop instruments that could be used repeatedly to gather longitudinal data within particular programs while affording comparability across the entire spectrum of programs offered by OIT.

The first step in this process was to identify the levels of analysis that instrument items would target for measurement, regardless of the program under consideration. The four basic levels identified included the following: 1) the institutional level (e.g., the entire university system, university campus, or college); 2) the departmental or unit level (often the entity sponsoring and administering the program); 3) the programmatic level (the particular program of interest); and 4) the individual level (the program participant). These levels of analysis serve as conceptual containers into which each individual program goal might be sorted once they have been identified and defined (see Table 1).

Second, we worked to identify and define the goals of the program in measurable terms. This process required considerable attention to details surrounding the history of the program and its evolution into its current incarnation. To acquire an historical perspective, previous calls for proposals (CFPs) were gathered and analyzed to glean common themes and stated goals. Additionally, applications to the program and overviews of individual projects were gathered to assess the consistency of fit between the programmatic CFPs and outcomes, both intended and actual. An iterative process involving input from members of the OIT Research and Evaluation Team, the FFP managers, and the manager of OIT faculty development services facilitated the revision and refinement of the identified

Table 1. Distribution of Goal Dimensions across Levels of Analysis and Programmatic Types

	Common to All Faculty Development Programs	Unique to the Faculty Fellowship Program (FFP)
Institutional Level	Student Learning Outcomes (SLOs)	Organizational Awareness
Departmental or Unit Level	Interdisciplinary Community	Scholarship of Teaching and Learning (SoTL)
Programmatic Level	Technology Enhanced Learning (TEL)	Instructional Culture
Individual Level	Skill Development	Leadership

goals into eight basic dimensions: Student Learning Outcomes (SLOs), Interdisciplinary Community, Technology Enhanced Learning (TEL), Skill Development, Organizational Awareness, Scholarship of Teaching and Learning (SoTL), Instructional Culture, and Leadership. Of these dimensions, the first four are common to all OIT faculty development programs (the "*common dimensions*") and the last four are specific to the FFP (the "*unique dimensions*"), thus fulfilling our requirement that our instruments and items be comparable across programs.[1] Furthermore, each of the program's goal dimensions fit into a mutually exclusive cell in the matrix derived from the combination of the theoretical levels of analysis and goal categories and presented in Table 1, thereby providing the foundation for the instrument's content validity.

While the discussions with FFP program managers to identify the goal dimensions provided the basic definitional parameters of the constructs in question, we also relied on the works of Fink (2003) and Bransford, Brown, and Cocking (2000), which serve as foundational theoretical texts for the FFP, to frame the basic dimensional concepts. From this point, the definitions were operationalized into arrays of individual items that were written in the first-person perspective, were amenable to scalar measures of agreement, and corresponded to each of the eight goal dimensions. After several revisions and edits, we settled on forty items that would comprise the final evaluation instrument. The items were randomized and placed on a four-point Likert scale with responses ranging from Agree Strongly to Disagree Strongly (see Table 2).

Methods

Given the eighteen-month scope of the FFP, we decided to administer the evaluation instrument four times at approximately six-month intervals (pretest at the beginning of the program, the six-month point, the twelve-month point, and the end of the program). The collection of data at two additional time periods beyond a typical pretest and posttest sequence provided FFP program managers with formative data that could be used better respond to participants' needs, allowed us to track individual and group progress iteratively, and doubled the number of data points available for analysis. All responses were obtained confidentially by a member of the OIT research and evaluation team and were de-identified after being matched to previous rounds.

Table 2. Common Program Goals: Dimensional and Aggregated Change

Goal Type	Dimension	Item Number	Item
Common Program Goals	Interdisciplinary Community	18	I seek out the opinions and perspectives of others before making important decisions.
		25	**I think that the costs of collaboration tend to outweigh the benefits.**
		16	I frequently discuss issues related to teaching with others.
		12	**I rarely collaborate on academic projects with individuals from disciplines other than my own.**
	Skill Development	3	I am comfortable working with new hardware and software applications.
		11	I would alter the structure and content of my courses to accommodate new technologies if there were an efficient method for doing so.
		10	**I think that new technologies can be incorporated into my courses without changing the way I currently teach.**
		30	**I think that attempting to evaluate courses and projects continuously is too demanding for both faculty and students.**
	Technology Enhanced Learning	7	I use technology to manage course-related tasks.
		23	I use technology in teaching to communicate with students.
		9	I use technology in the classroom for the purpose of information transference.
		31	I use technology during class sessions to enhance the learning experience of my students.
		19	I alter my teaching as I incorporate new technologies into my courses.
	Student Learning Outcomes	40	Technology can help students identify, define, and solve problems.
		14	Technology can help students locate information.
		22	Technology can help students critically evaluate information.
		38	Technology can help students master a body of knowledge.
		5	Technology can help students master a mode of inquiry.
		37	Technology can help students understand diverse philosophies and cultures.
		8	Technology can help students communicate effectively.
		4	Technology can help students understand the role of creativity, innovation, discovery, and expression across disciplines.
		21	Technology can help students acquire skills for effective citizenship.
		34	Technology can help students acquire skills for effective life-long learning.

[1] In order to develop the items for common program dimensions, each of the other OIT-sponsored faculty development programs were subjected to the same process described above for the FFP.

Results

Common Program Items

The average pretest scores on each of the goal dimensions common to all OIT faculty development programs were considerably higher than the expected average of 2.50, the mathematical midpoint of our four-point Likert scale (see Table 3). Specifically, the FFP participants demonstrated a strong disposition to enhance teaching and learning with technology (TEL = 3.40/4.00) and exhibited a great deal of confidence in technology's ability to advance SLOs (3.40/4.00). Additionally, OIT faculty fellows entered the program with a strong affinity for and belief in the importance of fostering and participating in an Interdisciplinary Community (3.20/4.00) and a highly developed set of skills related to integrating technology into their teaching (Skill Development = 3.15/4.00). These high pretest scores for the common goal dimensions are not surprising given that the FFP is the flagship and most selective faculty development program offered by OIT. However, because of the high starting points, we expected the amount of observable change to be relatively small and difficult to obtain given that improvement becomes more difficult the closer one is to the upper limit of the scale.

Despite these concerns, FFP participants experienced considerable change across all four of the individual common dimensions and experienced aggregated normalized gains of approximately 29%. Two of the largest normalized gains were made on the dimensions of Skill Development (56%) and SLOs (50%), both of which are statistically significant at the .05 level using a paired t-test, an impressive result given the small number of cases under consideration (n = 5). Although not statistically significant, the normalized gains made on the dimension of TEL (53%) are no less remarkable. However, the fellows experienced a negative shift (the only one in the study) on the Interdisciplinary Community dimension.

This latter result proved rather surprising given the relatively stable pattern for this dimension established in the first three cycles of the evaluation. Upon closer inspection of the data, we discerned that the precipitous drop in the Interdisciplinary Community score is resultant of responses to one particular question. In the fourth and final iteration, administered during the last group meeting, fellows responded on average that the costs of collaboration outweigh the benefits, a result that was 50% lower than the pretest score.[2] This result was,

Table 2 (continued)

Goal Type	Dimension	Item Number	Item
Unique Program Goals	Leadership	1	I discuss technology enhanced learning issues with my departmental colleagues.
		17	I serve as an advisory resource for colleagues on issues related to teaching and technology.
		32	I participate in public conversations about teaching with technology.
		2	I am comfortable approaching administrators with my ideas and concerns about university programs and initiatives regarding teaching with technology.
	Scholarship of Teaching and Learning	**39**	**I think that the peer review process is the best mechanism by which new knowledge is produced and disseminated.**
		27	**My research is too narrow and complex for general public consumption.**
		35	One of my primary responsibilities as a university instructor is to serve as a public intellectual in my community.
		6	I think that open-access, collaboratively generated repositories of knowledge are invaluable tools and resources.
		15	I think that published research on topics related to teaching and learning is as legitimate a form of scholarship as substantive disciplinary research.
	Instructional Culture	24	I think faculty should cultivate personal relationships with their students.
		28	**I think that the primary task of instructors is to transfer knowledge and expertise to their students.**
		33	I make the effort to empower my students to actively construct, discover, and transform knowledge in my courses.
		36	I think that class time is best used engaging in problem solving, communicative, collaborative, constructivist, and/or expressionistic activities.
	Organizational Awareness	29	I am familiar with the array of institutional units and forms of infrastructural support available to facilitate my use of technology to enhance my teaching and learning goals.
		20	I actively seek out institutional support and consultation on teaching and learning issues only as needs arise.
		13	I am familiar with the institutional procedures associated with strategic decision-making and capital planning as it relates to teaching and learning with technology.
		26	**I am unfamiliar with institutional policy issues of security, safety, and privacy related to technology in teaching and learning.**

NOTE: Numbers and items shown in bold are negatively worded statements that are recoded for data analysis. Respondents were instructed to "tell us how much you agree or disagree with each [item]. Do you Agree Strongly (AS), Agree (A), Disagree (D), or Disagree Strongly (DS)?"

[2] The pretest score for this item, reversed for comparability, was 3.80/4.00; the posttest score was 1.90/4.00. This -1.90 change represents a 50% loss over the course of the program, a difference that is statistically significant at the .05 level.

by itself, large enough to produce the aggregated negative posttest scores on the Interdisciplinary Community dimension.

Unique Program Items

The average pretest scores for the goal dimensions unique to the FFP are comparatively lower than those for the common ones (see Table 4). Fellows scored the highest on dimensions related to teaching (Instructional Culture = 3.55/4.00) and scholarship (SoTL = 3.16/4.00), the predominant professional activities in which they are engaged. The lowest average scores were on the dimensions measuring their leadership skills (Leadership = 2.80/4.00) and familiarity with resources available in support of using technology in teaching and cross-disciplinary collaboration, as well as institutional policies and procedures that shape technology's use in teaching and learning (Organizational Awareness = 2.50/4.00). Not unlike the higher average starting points for the fellows on the common goals, these lower scores on the unique goals were not surprising given that a recognition of the need and desire to develop one or more of these traits served as a primary motivation behind the fellows' decisions to participate in the program.

Overall, the faculty fellows experienced positive normalized gains on all four unique goal dimensions after eighteen months in the program. The greatest amount of observed change, in both absolute and relative terms, occurred on the dimension of Leadership. This suggests that as a result of their experiences in the FFP, the fellows made substantial normalized gains in terms of discussing with and serving as resources for colleagues on issues related to teaching and technology, participating in public conversations about teaching with technology, and approaching administrators with their ideas and concerns about technology and pedagogy. Fellows experienced similar normalized gains in moving their values toward an even more student-centered Instructional Culture (38%) and expanding their conceptualization of what constitutes legitimate and meaningful forms of scholarship to include as SoTL (36%). In the aggregate, FFP participants made larger normalized gains on the unique goals (34%) than they did on the common goals.

Reliability

While the information gathered by our new evaluation instrument during the course of the program provided both formative data that helped the program managers respond to the needs of the participants and summative data that allow us empirically understand the impact of the program on its participants, it is also necessary to evaluate the reliability of the instrument itself. In other words, we need to assess to what extent the individual items are internally consistent with respect to the respective dimensional constructs that they were designed to measure.[3] This affords us the opportunity to identify particular items that require editing or replacement in order to strengthen the instrument's power and to preserve collections of items that appear to measure their intended constructs.

To assess the reliability of the pilot instrument, we

Table 3. Common Program Goals: Palred T-tests of and Gain Scores for Dimensional and Aggregated Change

Dimension	Pretest †	Posttest †	Difference	Normalized Gain (*g*) ††
Skill Development	3.15 (0.13)	3.63 (0.13)	0.48*	56%
Interdisciplinary Community	3.20 (0.09)	2.93 (0.13)	-0.28	-13%
Technology Enhanced Learning (TEL)	3.40 (0.19)	3.72 (0.15)	0.32	53%
Student Learning Outcomes (SLOs)	3.40 (0.18)	3.70 (0.17)	0.30*	50%
Common Program Goals (Aggregated)	3.29 (0.11)	3.49 (0.03)	0.21	29%

NOTE: *$p < .05$

† Cell entries are means with standard errors in parentheses.

†† Normalized Gain scores represent the percentage of change relative to the possible amount of change from the pretest scores. The formula for calculating gains is as follows: %Change = ((Posttest – Pretest)/(Possible change)) * 100.

Table 4. Unique Program Goals: Palred T-tests of and Gain Scores for Dimensional and Aggregated Change

Dimension	Pretest †	Posttest †	Difference	Normalized Gain (*g*) ††
Leadership	2.80 (0.17)	3.40 (0.22)	0.60	50%
Scholarship of Teaching and Learning (SoTL)	3.16 (0.07)	3.46 (0.21)	0.30	36%
Instructional Culture	3.55 (0.15)	3.68 (0.11)	0.13	38%
Organizational Awareness	2.50 (0.11)	2.83 (0.13)	0.33	22%
Unique Program Goals (Aggregated)	3.00 (0.05)	3.34 (0.14)	0.34	34%

† Cell entries are means with standard errors in parentheses.

†† Normalized Gain scores represent the percentage of change relative to the possible amount of change from the pretest scores. The formula for calculating gains is as follows: %Change = ((Posttest – Pretest)/(Possible change)) * 100.

[3] We also need to assess the instrument's construct validity, but it is not possible to employ factor analysis on our pilot data given that our small number of cases (n = 5) produces a Heywood case with communalities (squared correlations) equal to or greater than 1.0 and/or negative error variances.

calculated the Cronbach's α for each individual goal dimension and the corresponding aggregated dimensions. A Cronbach's α is a measure of scale reliability, or the internal consistency of a group of items thought to measure a single construct, and ranges from -∞ to 1.0. In general, a coefficient of 0.70 or better is required for a scale to be reliable, although a coefficient of 0.60 might mark the threshold for acceptability for a pilot study such as this. The results of our reliability tests appear in Table 5.

Four of the eight individual dimensions (SLOs, Leadership, Skill Development, and TEL) returned acceptable reliability coefficients (ranging from 0.7121 to 0.9096), suggesting that no revisions are required prior to the beginning of the next FFP cycle. For the remaining dimensions that failed to break the threshold of acceptability, some changes are necessary if the development of reliable scale constructs is to be accomplished. However, the aggregated dimensions, which serve as cumulative constructs representing both the *common* goals of all OIT faculty development programs and the *unique* goals of the FFP, appear to be substantially reliable with measures with Cronbach's α values of 0.8759 and 0.7710, respectively. Additionally, the reliability of the overall instrument is outstanding with a Cronbach's α of 0.8814. While these values may be slightly inflated given that the number of items in a scale can affect the size of the coefficient (Kopalle and Lehmann, 1997), we remain confident that the overall instrument and aggregated dimensions are reliable given the theoretically derived content and face validity of the items produced.

In light of these results, OIT researchers revisited the entire instrument in light of the results from the reliability tests and made substantial revisions to a number of items in anticipation of the 2010-11 FFP. First, we reversed the direction of all negatively worded items to correct for measurement artifacts that may undermine the reliability of the scales employed (Bradley, Royal, and Bradley, 2008; Stewart and Frye, 2004). Second, other scales were evaluated on an item-by-item basis and resulted in one of three actions. For the Skill Development, Leadership, and Instructional Culture dimensions, we revised the wording of some items to simplify items for clarity. We removed one item from the Technology Enhanced Learning (item #19) and Organizational Awareness (item #20) dimensions given that doing so both strengthened the Cronbach's α and increased the face validity of the scales (see Table 2). In light of the poor Cronbach's α scores for the Interdisciplinary Community and Scholarship of Teaching and Learning dimensions, we overhauled the entire construct, revising each item extensively. Finally, we made no changes whatsoever to the Student Learning Outcomes dimension given that it performed exceptionally well in its original version. While we are confident that the revisions made shall improve the overall reliability of our new 38-item instrument, we will not know for certain until we have finished collecting the data in January 2012.

Table 5. Scale Reliability Scores, by Goal Dimension

Dimension	Number of Items	Cronbach's α
Skill Development	4	0.7256
Interdisciplinary Community	4	0.4561
Technology Enhanced Learning (TEL)	5	0.7121
Student Learning Outcomes (SLOs)	10	0.9096
Common Program Goals (Aggregated)	*23*	*0.8759*
Leadership	4	0.7667
Scholarship of Teaching and Learning (SoTL)	5	0.3550
Instructional Culture	4	0.5912
Organizational Awareness	4	0.4830
Unique Program Goals (Aggregated)	*17*	*0.7710*
All Dimensions (Aggregated)	*40*	*0.8814*

Conclusion

Although the OIT faculty development program evaluation project at the University of Minnesota has produced empirical evidence of the success of the FFP, the considerable variation in form, function, and duration of programs at other institutions suggests that the model of program evaluation detailed above be emulated primarily in procedural, not necessarily substantive, terms for at least two reasons. First, the constructs identified and the manner in which they are defined are specific to both the field of educational technology and, for the unique dimensions, the OIT FFP. Second, there are considerable limitations on the potential application of our findings to other programs from a statistical perspective (e.g., small-n problem and scale reliability issues). However, institutions sponsoring faculty development programs with goals similar to the ones outlined here might benefit from using some of the items and scales developed for the OIT FFP. Regardless, the formative lessons gleaned from the data and the analysis not only have provided the program managers with valuable information about how to structure future offerings of the FFP, but have provided OIT researchers with information that can be used to improve the instrument iteratively.

The procedural impact of our project is manifest in both theoretical and practical terms. Theoretically, our study demonstrates clearly the value of taking seriously the best practices recommendations that Hines (2009) makes with regard to a rigorous, systematic, and robust approach to program evaluation. Practically, we

have sketched out a basic methodology for creating and implementing a customized data collection instrument that should be applicable to any faculty development program. If designed and executed properly, such an approach should prove invaluable to program managers in refining existing programs, to administrators in terms of allocating resources to effective programs, and to entrepreneurs seeking to create new programs to meet faculty needs.

References

Bradley, K.D., Royal, K.D., & Bradley, J.W. (2008). An investigation of 'honesty check' items in higher education course evaluations. *Journal of College Teaching & Learning*, *5*(8), 39-48.

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

Chism, N.V.N., & Szabó, B. (1997). How faculty development programs evaluate their services. *The Journal of Staff, Program, and Organization Development*, *15*(2), 55-62.

Diaz, V., Garrett, P.B., Kinley, E.R., Moore, J.F., Schwartz, C.M., & Kohrman, P. (2009). Faculty development for the 21st century. *EDUCAUSE Review*, *44*(3), 46-55.

Fink, L.D. (2003). *Creating significant learning experiences: An integrated approach to designing college courses*. San Francisco: Jossey-Bass.

Hines, S.R. (2009). Investigating faculty program assessment practices: What's being done and how can it be improved? *The Journal of Faculty Development*, *23*(3), 5-19.

Hoyt, D.P., & Howard, G.S. (1978). The evaluation of faculty development programs. *Research in Higher Education*, *8*, 25-38.

Kopalle, P.K., & Lehmann, D.R. (1997). Alpha inflation? The impact of eliminating scale items on Cronbach's alpha. *Organizational Behavior and Human Decision Processes*, *70*(3), 189-197.

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

Levinson-Rose, J., & Menges, R.J. (1981). Improving college teaching: A critical review of research. *Review of Educational Research*, *51*(3), 403-434.

Solheim, C., Longo, B., Cohen, B.A., & Dikker, A.G. (2010). Interdependent catalysts for transforming learning environments…and the faculty who teach in them. *EDUCAUSE Quarterly*, *33*(3). Retrieved from http://www. educause.edu/EDUCAUSE+Quarterly/ EDUCAUSE QuarterlyMagazineVolum/InterdependentCatalys tsforTran/213680

Stewart, T.J., & Frye, A.W. (2004). Investigating the use of negatively phrased survey items in medical education settings: Common wisdom or common mistake? *Academic Medicine*, *79*(10), S18-S20.

D. Christopher Brooks serves as a Research Fellow in the Collaborative for Academic Technology Innovation in the Office of Information Technology at the University of Minnesota-Twin Cities. He conducts research on the impact of educational technologies on teaching practices and learning outcomes. His research appears in a number of scholarly journals including the The Journal of College Science Teaching, The American Biology Teacher, the International Journal for the Scholarship of Teaching and Learning, Evolution, and the Journal of Political Science Education. He can be reached by email at dcbrooks@umn.edu.

Lauren Marsh is manager of the Faculty Development and Consultation Services in the Office of Information Technology at the University of Minnesota. Since joining the team in 2003, she has worked with faculty, departments, and colleges to help them plan the thoughtful application of learning technologies. She is co-manager, along with Kimerly Wilcox, of the OIT Faculty Fellowship Program–an 18-month, multidisciplinary program that promotes faculty leadership and scholarship in the area of technology-enhanced learning. She has published and presented on faculty development and technology-enhanced learning. She received her Ph.D. in English from the University of Minnesota.

Kimerly Wilcox received her Ph.D. in the Biological Sciences from the University of Cincinnati. While teaching biology courses at the University of Minnesota, she became intensely interested in the enormous educational potential of new technologies and the World Wide Web and, in 2000, joined the Office of Information Technology (OIT) as a Senior Academic Technology Consultant. In that position, she consults with instructors on effective uses of educational technology, coordinates consultation services, participates in several faculty development programs, and co-manages, with Lauren Marsh, the OIT Faculty Fellowship Program. Her interests include active, student-centered learning; learning/cognitive science; new learning environments; and STEM (Science, Technology, Engineering, Math) education. She can be reached at: University of Minnesota Office of Information Technology, 131 Walter Library, 117 Pleasant St. SE, Minneapolis, MN 55455. Email: wilco001@umn.edu.learning. She received her Ph.D. in English from the University of Minnesota.

Brad Cohen currently services as Director of the Collaborative for Academic Technology Innovation in the Office of Information Technology. The Collaborative fosters strategic academic technology partnerships with colleges and other center service providers to advance the core academic mission of the University of Minnesota. He can be reached by email at cohenb@umn.edu.