Med Educ Online [serial online] 2000;5:9. Available from URL http://www.med-ed-online.org

Fitting the Means to the Ends: One School's Experience with Quantitative and Qualitative Methods in Curriculum Evaluation During Curriculum Change[1]

Ann W. Frye, MEd[*], David J Solomon, PhD[†], Steven A. Lieberman, MD[‡], Ruth E. Levine, MD[§], [*]Office of Educational Development, University of Texas Medical Branch, [†]Office of Medical Education Research and Development and the Department of Medicine, Michigan State University[2], [‡]Department of Internal Medicine, University of Texas Medical Branch, [§]Department of Psychiatry and Behavioral Sciences, University of Texas Medical Branch

Abstract - Curriculum evaluation plays an important role in substantive curriculum change. The experience of the University of Texas Medical Branch (UTMB) with evaluation processes developed for the new Integrated Medical Curriculum (IMC) illustrates how evaluation methods may be chosen to match the goals of the curriculum evaluation process. Quantitative data such as ratings of courses or scores on external exams are useful for comparing courses or assessing whether standards have been met. Qualitative data such as students' comments about aspects of courses are useful for eliciting explanations of observed phenomena and describing relationships between curriculum features and outcomes. The curriculum evaluation process designed for the IMC used both types of evaluation methods in a complementary fashion. Quantitative and qualitative methods have been used for formative evaluation of the new IMC courses. They are now being incorporated into processes to judge the IMC against its goals and objectives.

Curriculum evaluation plays an important role in substantive curriculum change.[1] A comprehensive evaluation of a curriculum undergoing significant change requires addressing attitudes, assumptions, values, and expectations.[2] Evaluation processes must examine not only the "product" (e.g., students' knowledge and skills) and the effects of the changes but also the unintended positive or negative consequences of the curricular change.[3]

The evaluation of curricular change may have several different goals. Sometimes there are specific questions about the curriculum to be answered, for example.[4] :

- "Are curricular objectives being met?"
- "Is the student experience what we intended it to be?"
- "Are the course activities optimally useful to students?"
- "What is the effect of this curriculum policy?"

In other instances, the curriculum needs to be assessed in a general way for its worth or merit.[5] A typical question associated with a general assessment of a curriculum's worth is "How satisfied are students and faculty with the curriculum?". Often both specific questions and a more general assessment may be relevant. Whatever the goals of the evaluation, it is important to fit the evaluation methods to the goals selected for a curriculum evaluation project.

The goals set for the curriculum evaluation process should reflect the nature and magnitude of the curricular revisions. Small changes may call for focused evaluations. Significant revisions, however, such as incorporating problem-based learning or Web-based instructional features into a traditional curriculum or implementing integrated courses in place of discipline-based courses, require more substantial evaluation processes. During the implementation of curricular change, evaluation methods should first assess if the changes are being implemented as intended and should examine the impact of these changes on the affected groups. Once the

[1]Brian Mavis, Ph.D. managed the peer-review process for this manuscript.

[2] Formerly of the Department of Internal Medicine and the Office of Educational Development, University of Texas Medical Branch

changes are thoroughly integrated into the curriculum as intended, the outcomes of the curricular revisions can then be assessed against appropriate standards.

This paper describes the design and development of the curriculum evaluation work at the University of Texas Medical Branch (UTMB). Examples from UTMB's experience are used to illustrate the complementary use of quantitative and qualitative evaluation methods.

Choosing the Evaluation Means to Fit the Ends: Quantitative or Qualitative Methods — or Both?

The choice of evaluation methods flows from the goals of the evaluation process. Curriculum quality has traditionally been assessed by "countable", or quantitative, outcome measures such as course grades or scores on external exams. Students are also typically asked to rate different aspects of learning activities, yielding more numerical data. Quantitative approaches, however, may not adequately assess important features of medical education programs such as new emphases on life-long learning or the development of interpersonal skills. Simple comparisons of numerical data alone do not provide information on all of the important aspects of an educational program. Ratings rarely provide the explanatory information necessary for understanding the relationships between curricular features and outcomes. It may even be difficult in some instances to define important variables well enough to quantify them adequately. The notion of "student satisfaction" is a common example of a definitional challenge.

Qualitative measures, which generate words rather than countable elements, offer rich descriptions and explanations of phenomena to the curriculum evaluation process that complement quantitative information.[6] Focus groups, interviews, and open-ended written questions are some of the methods that are useful for addressing the "how" and "why" questions so important in evaluation of a changing curriculum. For example, "How are students using the written course objectives to direct their studying?" and "Why is collaboration among students so slow to replace competition for grades?" Qualitative data are often disparagingly referred to as "soft" data, but when collected and analyzed with appropriate rigor, they are very useful for looking at curricular problems, processes, and innovations.[2]

UTMB's Integrated Medical Curriculum

The recent curriculum revisions and associated evaluation processes at The University of Texas Medical Branch (UTMB) provide examples of combining both quantitative and qualitative evaluation methods to meet the needs of a faculty and school whose curriculum is undergoing substantive change. UTMB instituted the Integrated Medical Curriculum (IMC) in the fall of 1998.[7] The IMC represented radical change for a school with a relatively traditional approach to medical education. Interdisciplinary courses replaced department-based courses, providing a new emphasis on integration of basic and clinical sciences during the first two years of the curriculum. Faculty members were required to learn new roles as facilitators of small-group learning with problem cases and as lecturers charged with delivering interactive conceptual lectures. In addition, new approaches to student assessment and grading were incorporated into the curriculum to match testing procedures more closely to the learning activities.

The Curriculum Evaluation Plan

As the UTMB faculty worked through the curriculum design period, a task force of faculty and educators was charged with developing a comprehensive curriculum evaluation plan. The task force members included several directors of new courses, some Curriculum Committee members, a student representative, and other faculty interested in the work. Educators from UTMB's Office of Educational Development guided the task force's work and implemented the evaluation processes. Organizationally, the task force reported to the Curriculum Committee, the faculty oversight group for the curriculum.

The comprehensive curriculum evaluation plan was intended to guide short- and long- term decision-making about the new curriculum and to evaluate its overall effects. The plan included elements aimed at documenting the implementation of the curriculum against its goals and objectives, monitoring student progress toward curriculum goals, and providing a rich description of how students, faculty, graduates, and external stakeholders (e.g., residency directors) perceive the curriculum. Specific elements were included to document and compare what Coles and Grant termed the "curriculum on paper" (the intended curriculum) versus the "curriculum students experience".[2] Finally, the plan was also designed to docu-

ment the outcomes of the curriculum as students moved through it and then into advanced training.

The curriculum evaluation plan's design was drawn from several evaluation models. Principles from Guba and Lincoln's *naturalistic inquiry* paradigm[8] guided the choice of assumptions for the plan's approach. Those assumptions were that different groups (e.g., students and faculty) and even different individuals within those groups might experience the curriculum innovations differently. Understanding the variation in their experiences was assumed to be important in understanding how the curriculum was working.[9] For this reason, the decision was made to obtain data from multiple sources. Stufflebeam's *CIPP model*[10], referring to educational *C*ontext, *I*nput, *P*rocesses, and *P*roduct, stipulates that the purpose of evaluation is to improve the program rather than to prove anything about the program. The CIPP model's focus on program context, input, processes of the program, and the program's product is particularly useful for designing evaluations meant to provide information for guiding change and improvements. Finally, Stake's *responsive evaluation* model specifies that evaluators and stakeholders create the evaluation together.[11] In the IMC evaluation plan, the primary stakeholders included the Curriculum Committee (the faculty group charged with curriculum oversight), the course directors, and the students, all of whom participated in some way in the design of evaluation procedures and the collection of evaluation data.

Faculty groups needed information to supplement and put into perspective what they were seeing and hearing about the curriculum from students and faculty on a daily basis. They also needed to make longer-term assessments of curriculum quality and outcomes. To meet these needs, systematic data collection, appropriate sampling of respondent groups, and careful design of questions was required. A mixture of quantitative and qualitative methods was appropriate for this set of needs.

The quantitative measures adopted included close-ended survey questions and rating forms for students, faculty, and graduates; test scores; library records; admissions data; registrar's records; residency match data; and graduation data. Figure 1 lists these measures and their association with the areas covered in the evaluation process. The measures allowed comparisons of variables across time or across curriculum units (e.g., courses), as well as pre- and post-reform comparisons. They also facilitated comparisons to standards set for the curriculum, e.g., expected pass rates on USMLE exams. The difficult challenges of constructing good instruments to yield quantitative data were balanced against the instruments' relative ease of administration and analysis.

The qualitative measures included written and oral responses to open-ended questions by students, faculty, and graduates; focus group reports; and descriptions of students on clinical clerkships (Figure 1). These measures allowed investigation of why something was working the way it was as well as examination of problems, processes, and the internal dynamics of courses and the curriculum. They helped to insure that important evaluation information was not missed and to identify the issues and concerns critical to improvement.[12] The benefits of rich information gain were balanced against the challenge of objective data analysis.

Implementation of the Evaluation Plan

In the initial years of the IMC, data collection processes were aimed primarily at formative evaluation of the curriculum. The approach was largely descriptive in nature in order to provide rich information about what was happening and to identify the immediate consequences of the curricular changes. Student focus group sessions and computer-based surveys with both Likert scale and open-ended questions provided feedback for monitoring progress toward curricular goals, as did similar faculty surveys. Systematic sampling of respondents and carefully constructed questions were employed in order to increase validity of the evaluation results. Data summaries were provided to the Curriculum Committee and course directors. Numeric summaries of responses to Likert-scale items quantified the comparison of student and faculty reactions across courses and the curriculum. In a complementary fashion, the data collected and analyzed with qualitative methods proved useful in describing the quality of the small-group work and other learning opportunities, the balance of different course activities, and the usefulness of new learning activities developed in some courses.

Quantitative and Qualitative Methods: The Users' Perspectives

In broad terms, the quantitative data have been most helpful in identifying major problems and in assessing whether individual criticisms obtained via qualitative methods were representative of the opinions of the whole group. The qualitative data have helped in clarifying specific problems and in formulating plans to address them.

An example from the first year's work demonstrates the benefits of using a combination of quantitative and qualitative approaches to curriculum evaluation. One of the evaluation challenges was to track students' perceptions of the consequences of different approaches to grading student performance in the problem-based tutorial groups. Although a standard rating form was used to assess student performance in the groups, courses differed widely in the way ratings were converted into grades. Some courses averaged the percentage of points earned on the rating form into the course grade (honors/high pass/pass/fail). Some assigned honors/pass/fail to the small group score itself. Others assigned only pass/fail to the small group score.

The effects of those differences were identified as an evaluation question. We compared the frequency distribution of student responses across courses to a rating item, "*The tutorial group assessment was a fair measure of your skills and knowledge.*" Responses were marked on a 5-point scale (strongly agree to strongly disagree). We also asked students in focus groups: "*What, if anything, has the grading approach to the tutorial groups in this course meant to you?*" and "*What is the effect on competitiveness among your classmates of this approach to grading in the tutorial groups?*" Finally, we asked students for their written comments on each course's evaluation system as they completed the course.

The data demonstrated that students did perceive an effect on their approach to the small groups that they attributed to the grading system employed in the course. Several second-year courses used a pass/fail grading system in the small groups rather than the 10-point scale used in most of the first year courses. Initially students reported a favorable decrease in competitiveness and disruptive "grandstanding" by classmates during the small group sessions. However, in subsequent courses, students reported that there was a noticeable decrease in some students' effort and contributions in the small groups attributable to the pass/fail grading system. After three courses in which small groups were graded pass/fail, the use of honors/pass/fail grading in a subsequent course was perceived as beneficial to small group work by students. This observation encouraged the Curriculum Committee to establish a curriculum-wide policy that all small groups should receive greater weight in the final course grade than simply pass/fail. Using a combination of quantitative and qualitative evaluation methods not only allowed identification of the effect of different grading methods but also the tracking of changes in that effect as students moved through the curriculum.

The qualitative and quantitative data have complemented each other in other ways as well. In some instances, the quantitative data have been useful in helping to assess whether strong individual opinions were representative of the opinions of the group as a whole. For example, the clinical skills course received several strongly worded student criticisms about the value of ward visits. The quantitative data, however, indicated that these experiences were viewed positively by most students. The combined use of qualitative and quantitative data indicated to the course directors and the Curriculum Committee that the experience was generally positive but that some areas (i.e., specific ward services) required improvement.

In other instances, the qualitative data illuminated problems identified by the quantitative data. Quantitative ratings data indicated broad student dissatisfaction with some required course texts. The qualitative data were essential in informing the course directors that in some cases, the texts were too general, while in others the texts were too detailed.

Finally, the combination of quantitative and qualitative data has helped to promote continuous quality improvement. During the first year of the curriculum, qualitative feedback from the students indicated a desire for large group experiences to review and summarize the PBL cases after their conclusion. This became a standard feature of all courses. Ongoing evaluation indicated that these wrap-up sessions, generally structured as open question and answer sessions, were not meeting the students' needs. Qualitative data suggested a more structured approach would be preferred. Implementation of changes in subsequent courses' wrap-up sessions has resulted in improvement.

Conclusion

Two years of multi-method curriculum evaluation with emphasis on description and documentation have been completed at UTMB. In the curriculum's third year of implementation, the evaluation focus turns toward activities that provide more summative information. At the same time, the work of evaluating the curriculum at the course level continues. The task force's initial plans projected a decrease in the quantity of data collected for individual courses as the courses matured and an increase in attention to

the long-term effects of the curricular change. Course directors were unwilling, however, to give up the relatively rich data they were getting about their courses.

The original set of student survey questions, both closed- and open-ended, has therefore been retained with very little change. Additional elements of the curriculum evaluation plan have been implemented as outlined in Figure 1. These additions have necessitated some adjustments in the original data collection approaches. To avoid increasing the burden on students that the gathering of more information would involve, students are now randomly assigned to sets of evaluation questions during each evaluation period. Each student therefore answers only a subset of all the questions being asked in a given administration but are invited to comment freely on all aspects of a course. Data analysis and reporting procedures have been streamlined to compensate for the greater quantity of quantitative and qualitative data being gathered. It is anticipated that continued use of mixed quantitative and qualitative methods will allow the measuring of the curriculum against the overall goals and objectives of the Integrated Medical Curriculum and examination of IMC students' performance in subsequent educational activities, including clerkships and post-graduate training.

In summary, the UTMB experience illustrates that curriculum evaluation is an essential component of effective curricular revision. A curriculum evaluation plan, tailored to the specific goals and nature of the curriculum, works best when it is sensitive to the needs and structure of the organization and of the various stakeholders in the curriculum. Organizational needs change over the course of implementing curricular revisions, generally moving from formative questions about the congruence of the intended and actual curriculum to more summative questions about the extent to which the goals of curricular reform have been achieved. Adoption of both quantitative and qualitative evaluation methods is useful to adequately address the multifaceted needs of a comprehensive evaluation of a curriculum reform.

References

1. Bland CJ, Starnaman S, Wersal L, Moorhead-Rosenberg L, Zonia S, Henry R. Curricular change in medical schools: How to succeed. Academic Medicine 2000 June; 75(6):575-594.

2. Coles CR, Grant JG. Curriculum evaluation in medical and health-care education. Medical Education 1985; 19:405-422.

3. Craig P, Bandaranayake R. Experiences with a method for obtaining feedback on a medical curriculum undergoing change. Medical Education 1993; 27: 15-21.

4. Stufflebeam DL, Webster WJ. An analysis of alternative approaches to evaluation. In: Madaus GF, Scriven M, Stufflebeam DL, editors. Evaluation models: Viewpoints on educational and human services evaluation. Norwell, MA: Kluwer Academic Publishers, 1983;14.

5. Scriven M. The nature of evaluation part I: Relation to psychology. Practical Assessment, Research & Evaluation [serial online]1999;6(11). Available from http://ericae.net/pare/getvn.asp?v=6&n=11.

6. Miles MB, Huberman AM. Qualitative data analysis: A sourcebook of new methods. Newbury Park, CA:Sage Publications, Inc., 1984.

7. Bernier GB, Adler S, Kanter S, Meyer W. On changing curricula: Lessons learned at two dissimilar medical schools. Academic Medicine 2000; 75:595-601.

8. Guba ES, Lincoln YS. Epistemological and methodological bases of naturalistic inquiry. In:Madaus GF, Scriven M, Stufflebeam DL, editors. Evaluation models: Viewpoints on educational and human services evaluation. Norwell, MA: Kluwer Academic Publishers, 1983.

9. Fetterman DM. Qualitative approaches to evaluating education. Educational Researcher 1988;17(8):17-23.

10. Stufflebeam DL. The CIPP model for program evaluation. In: Madaus GF, Scriven M, Stufflebeam DL, editors. Evaluation models: Viewpoints on educational and human services evaluation. Norwell, MA: Kluwer Academic Publishers, 1983.

11. Stake RE. Program evaluation, particularly responsive evaluation. In: Madaus GF, Scriven M, Stufflebeam DL, editors. Evaluation models: Viewpoints on educational and human

services evaluation. Norwell, MA: Kluwer Academic Publishers, 1983.

12. Patton MQ. How to use qualitative methods in evaluation. Newbury Park, CA: Sage Publications, 1987.

Figure 1

UTMB Integrated Medical Curriculum Years 1 and 2

Areas for Evaluation and Evaluation Measures

DATA SOURCES

quantitative data | **qualitative data**

An "x" indicates that the data source addresses the topic in that row.

AREAS FOR EVALUATION	CURRMIT (AAMC curriculum database)	Student surveys -- open-ended questions	Student surveys -- ratings	Student course evaluations -- open-ended questions	Student course evaluations -- ratings	Student focus groups	Graduate survey	Faculty surveys	Faculty interviews	Residency directors' survey -- open-ended questions	Residency directors' survey --ratings	Student performance on course assessments	Practice of Medicine Course OSCEs	Clerkship assessments	Clerkship OSCEs	Fourth-Year Clinical Skills Exit Exam	Medical School Learning Environment Survey	Mitchell Cognitive Behavior Survey	RWJ/GPI Survey	Stress level Survey	Step I exam	Step II exam	Registrar's records	Faculty effort tracking system	Residency Match results	School of Medicine admissions data	Library records
Student Outcomes																											
Biomedical knowledge (see curriculum objectives list)			X	x		x	x	x			x	x		x							x	x					
Applying biomedical knowledge to pt problems (clinical reasoning; problem solving)			x	x		x	x	x			x	x		x	x	x					x	x					
Clinical skills (H&P)			x	x		x	x	x			x	x	x	x	x	x											
Preparation for required clerkships			x			x	x						x	x													
Working effectively in a healthcare team (teamwork skills)			x			x	x	x			x			x													
Identification of, addressing knowledge deficits			x			x	x	x			x	x															
Career choice			x				x												x								
Professional behavior			x			x	x				x		x	x	x	x											
Residency choices (specialty area, competitiveness of program)			x																						x		
Residency match success			x																						x		
Skills for self directed learning			x			x	x	x			x																
Communication skills (with faculty, peers, patients)			x			x	x	x			x	x	x	x	x	x											
Impact on Faculty																											
Faculty time/effort requirements								x	x															x			
Impact on other faculty duties (research/clinical)								x	x																		
Faculty satisfaction								x	x																		
Faculty enjoyment								x	x																		
Other (e.g., new collaborative relationships)								x	x																		
Effectiveness of Curricular Structure																											
Organ system organization			x	x		x	x	x	x			x															
Small group sessions			x	x		x	x	x	x			x															
PBL cases			x	x		x	x	x	x																		
Sequencing of content			x	x		x	x	x	x																		
Distribution of content across the curriculum	x		x			x	x	x	x																		
Integration of basic with clinical sciences	x		x	x		x	x	x	x			x															
Integration across disciplines	x		x	x		x	x	x	x			x															
Understanding of PBL objectives and student/faculty roles in curr.			x	x		x		x	x																		
Content coverage	x		x			x	x	x	x																		
Match of learning objectives to learning activities			x	x		x		x	x																		
Coordination of learning activities			x	x		x		x	x																		
Student Evaluation System																											
Satisfaction			x	x		x	x	x	x																		
Impact on student learning and study habits			x			x	x	x	x																		
Validity indicators for evaluation system			x			x	x	x	x																		
Learning Environment																											
Student satisfaction with the curriculum		x	x			x	x	x									x										
Student stress levels		x	x			x														x							
Changes in study habits, focus		x	x			x											x	x									
Perception of positive learning environment						x		x	x								x										
Other outcomes																											
Students' assuming responsibility for learning			x			x	x	x	x		x			x													
Impact on applicants' qualification levels																										x	
Impact on number of applicants																										x	
Impact on applicants' choosing UTMB																										x	
Impact on attrition																							x				
Library use			x			x																					x