

LEARNING FROM THE FYE PILOT PROJECT

Carlos Benito, Ph.D.

Art Warmoth, Ph.D.

The GE Task Force has engaged a significant group of faculty, academic staff, and students, to reflect on initiatives to improve the freshman year experience. One outcome of these activities has been a proposal to reform the existing freshman year curriculum by incorporating a First Class as the anchor for a new Freshman Year Experience. After explaining their plan within various forums, and then organizing a GE Fair, they have proposed to try it with a pilot project. This is a commendable approach, in line with modern decision-making: learning by doing, from a scientifically designed experiment. That is, they have decided to learn: if their alternative program will improve student learning and development; and what quantity and quality of resources will be necessary for the task. A pilot project will allow the faculty and administration to increase knowledge and to reduce uncertainty about student learning in FYE.

A pilot project is learning-by-doing approach: it will allow its administrators to adjust as they learn through the one year experience. It will also serve as an evaluation opportunity. It will answer questions such as these: Do we need different FYE programs for different groups of students or an improved generic one can do it? What is the cost-effectiveness of the proposed FYE program relative to the existing one?

We think that the faculty, through its governance structures, in collaboration with the administration, should approve the implementation of the pilot project. To accomplish its goals, however, the FYE project has to go along with an evaluation project. Were the evaluation absent or incomplete, one of the functions of a pilot experiment would be missed.

In the case of the Freshman Year Experience, program evaluation should incorporate both ongoing program assessment designed by faculty working for the pilot and a comparative assessment of the outcomes of three major components of the freshman curriculum: 1) students taking the FYE, 2) students taking

the current EMT Program, and 3) students taking only English 101 and Philosophy 101.

The program evaluation methodology should identify and evaluate outcomes consistent with the Mission, Goals and Objectives of General Education. And it would involve the following design elements and steps.

1. Quantitative Methods

- 1.1. Sampling (control group issues)
- 1.2. Standardized Assessment Instruments
- 1.3. Survey
 - 1.3.1. Questionnaire Design
 - 1.3.2. Survey Administration
- 1.4. Data processing
 - 1.4.1. Statistical analysis
 - 1.4.2. Path Analysis
 - 1.4.3. Regression Estimations
- 1.5. Interpretation of results
- 1.6. Report

2. Qualitative Methods

- 2.1. Focus Groups
- 2.2. Embedded Assessment
- 2.3. Other Methods
- 2.4. Report

It is not clear at this point whether the design of an appropriate evaluation strategy can be accomplished using on campus resources or should involve an outside consultant. We propose that a task force be created to further define the parameters of the program evaluation that is needed, and then invite both on campus and off campus professional evaluators to submit bids for carrying out the evaluation program. The methodology and resources for program evaluation should be in place before the Academic Senate authorizes the implementation of the pilot project.

Addenda from Carlos' 3/05

Dear Paul:

I have few questions for you, who in one way or another embody what the GE task force (by now a learning community) has learned. The motivation for my questions is how best to dialogue with colleagues who are pondering about the impacts of the Pilot project as it moves toward implementation.

Here they are:

1) What is the main learning objective (or development goal) of a first year experience at SSU? If I am narrowing it too much, you can list two objectives

2) What is the major contribution of the FYE Pilot toward that objective, which the existing program does not provide?

3) By what means the FYE Pilot will achieve that contribution?

4) What are the additional resources and/or structural changes (say reallocations) necessary for implementing those means?

5) Is the expected contribution (improvements) to learning (or development) worth its opportunity cost (as defined by additional resources and structural changes)?

6) What is going to be the main function of a Pilot (as different than a permanent FYE program): a process for learning by doing and therefore adjusting the program? or a way to appraise the worth and costs of a FYE relative to existing programs?

You may have your own conceptual framework for articulating the issues that I am asking about. You may choose to answer with your own conceptual framework.

Regards, and thank you for your good work.

From E Sundberg:

I just heard that this document was circulated to those individuals listed below (Senate chairs, and APC/EPC Chairs?) and that Elaine McDonald has placed it on the EPC agenda this week for discussion. Paul, have you been invited to attend EPC for this discussion?

I hope that we will address this proposal at our retreat on Sunday and that Rose Bruce can also have opportunity to provide input into this document.

I myself have several issues with this:

1. I don't understand the intention here of having Carlos Benito and Art Warmoth develop an assessment strategy for the FYE Pilot. It does not appear that they have read the assessment section of the FYE Pilot plan, which outlines some excellent strategies for direct assessment of student learning (which this proposal does not).
2. I also have problems with this document coming before EPC and bypassing Paul Draper and the FYE task force. It make sense that if it is intended to answer some of the questions that came up at EPC for Paul to answer, then this should have been given to Paul for incorporation into his responses to EPC.
2. This document is recommending the use of an outside evaluator, which is not the proposal of the GE task force and report. The report clearly states that the assessment development should be done by the coordinator, faculty, and institutional research – that is, by those involved in the pilot. Why is it necessary to have outside evaluators bid on this and add to the cost of this project? We have the expertise and ability to do it ourselves. Why wouldn't we involve our own IR as they have been instrumental in developing survey instruments for our FYE (EMT and Freshman Seminar).
3. This proposal from Carlos and Art only outlines – and sketchily at that -- indirect assessment techniques, when we are being called upon to directly assess student learning – that is, develop forms of assessment that actually look at what students are learning in terms of critical thinking, writing, etc. The FYE proposal proposes a much richer three-pronged assessment approach that could actually get to this important question.
4. I do not support the creation of yet another task force to define the assessment approach. I think we need to organize the faculty who will be teaching in the pilot, appoint a coordinator, and get to work!
5. The introduction to this proposal is biased: it states that the "FYE project has to go along with an evaluation project." This implies that there is not an evaluation proposed or planned when, indeed, there is. From the outset, the pilot task force has incorporated assessment into our plan.

Just some thoughts – how do you wish to proceed? What would you like me to convey to EPC tomorrow?

Elaine

From Paul Draper

Elaine and EPC:

Thanks to Carlos, Art and you for helping to more fully articulate assessment goals and needs for the "First Class" pilot. The pilot will need to discover what succeeds, what fails, and what needs improvement in the pilot in order to determine if it should be expanded, redesigned or curtailed in possible 2006-07 iteration. And that determination will need to be made in a timely fashion –say by January of 2006.

The Benito/ Warmoth memo "Learning from the FYE Pilot Project" is a welcome contribution to the assessment portion of the pilot, and I believe a discussion of it in EPC will have benefits. But I would like to put into writing my two cents, which I spoke about with you earlier today:

- A) The protocol relies heavily upon so-called "proxy" assessment methodologies (surveys). While these are good for corroborating other so-called "authentic" methods, they should not be the primary tools for assessing either learning outcomes or program effectiveness. Methodologies such as focus groups, embedded assessments, portfolio reviews, etc. must be aligned with specific learning outcomes and program goals in the "First Class" pilot, and would also need to be applied to the control groups suggested in the protocol ["1) students taking the FYE, 2) students taking the current EMT Program, and 3) students taking only English 101 and Philosophy 101."] (Benito) Only by working with the same tools in the control groups as the subject group can authentic assessment data be considered reliable—and point us in the right direction.
- B) Attached is a questionnaire that Carlos asked me to address regarding an assessment protocol for the First Class pilot. I believe it should be part of your assessment discussion in EPC. Carlos' questionnaire is very helpful in helping define "First Class" goals and objectives, and I will be responding to it in the near future. However, it must be the actual instructional faculty who are selected to teach in the pilot who will ultimately be charged with determining many of its specific learning outcomes. Responses to many of Carlos' questions can and will be addressed in the near term.
- C) The assessment protocol outside the GE Subcommittee FYE and/or Initiative Task Force. As such, it does not feel fully reflective of the "First Class" proposal itself, or organic to the process. There are certainly ways to close those loops that will undoubtedly strengthen the suggested protocol, and create an effective means for deepening the communication between EPC and its GE Subcommittee. This is particularly vital in that GE continues to operate without an EPC liaison this semester.

Best regards,

Paul

3/9/05

From Rose Bruce:
3/10/05

Hello All. I agree that Path Analysis can be an effective tool to understand student characteristics. However for the pilot and subsequent studies I believe the questions is: are the student learning outcomes different (or the same) across the sections (Pilot, Fr. Sem., Non Fr. Sem). This could be accomplished with a common final exam scored according to a rubric generated by the faculty to discern writing and critical thinking skills. Then it's simply an ANOVA to compare means across groups. If we wanted to account for entry verbal skills we could covary by SAT-verbal or EPT scores. Rose

-----Original Message-----

From: Carlos Benito [mailto:crlbenito@netscape.net]

Sent: Wednesday, March 09, 2005 10:51 PM

To: paul.draper@sonoma.edu

Cc: Elaine McDonald; 'Elaine Sundberg'; Warmoth, Art; 'Eduardo M. Ochoa'; 'L. Rose Bruce '

Subject: Re: FW: Appraising FYE

Dear Paul:

I share with you that those who will be more involved with the FYE pilot have to decide the relative importance and functions of alternative assessment instruments.

I also see the validity of your argument to keep assessment within the FYE Task Force. The evaluation of the project as a whole, however, could well be done by another group.

Nevertheless, without intending to change your preferences O:-) , I would like to invite you to a presentation that I will make on March 22nd, at noon. I will discuss a research about the behavior of high school Latino-youth, regarding community participation. It was based on a large survey that we conducted for this purpose. I will present it in the Brown-bag Lectures of the School of Social Science. Hopefully you will observe other functions for survey-based research. Using hypothesis about student behavior, and data collected via surveys, regression analysis allows to go beyond mean values, or the behavior of the average student. Indeed, Mr. and Ms. Average do no exist, but specific individuals, say students, with different behaviors. Being able to explain those differences is very enlightening for designing or redesigning programs. This requires to gather information about the characteristics of students--once we introduce characteristics we make room for diversity and for tailoring programs to different students' needs.

Were your choice to generate data with portfolios, and were you interested in the ex-post evaluation of the Pilot, still it will be necessary to gather data about the characteristics of each student, such as socioeconomic position, motivations, ethnicity and culture, gender, high school GPA or other scores, etc. The data will need to be generated for students in the Pilot and for students in the existing programs. It is about the control group.

Best regards,

Carlos

Dear colleagues:

Carlos' comments resonate with the question that I raised today at the Senate: what are the differences between **assessment**, **evaluation**, and **program review**? My question was to a certain extent rhetorical, since I have certain view on the matter.

While I am by no means an expert on assessment, I have learned enough about it in the last few years to realize that **assessment** in the evolving technical sense (as in “the assessment movement in education”) is about (1) identifying learning outcomes spanning a range of knowledge, skills, and abilities; (2) devising diagnostic tools that can provide information about how well these outcomes are being accomplished (preferably real-time and embedded in the learning process) in a course and a program; (3) using the results to guide program improvement in a continuous loop of delivery, assessment, and improvement. Thus, in the current technical usage, assessment really refers to the continuous quality improvement framework in place for a particular program. Assessing a program is not a one-time event based on a static picture of achievement in this usage.

Program review—which of course has a longer lineage—entails periodic (typically 3 to 5 years) inspection of all aspects of an academic program to judge the level of quality achieved and the level of inputs supplied to it. It identifies areas to be improved over the time frame until the next review. In the past, quality was measured by proxy variables such as the demand for the program, the scholarly productivity of its faculty, and the success of the graduates in the marketplace or in follow-up graduate placement. More recently, with the advent of the assessment movement, the reviews also include coverage of the quality improvement framework (with assessment and feedback of the results into curricular and pedagogical change as key components). (Accreditation reviews are in fact a mega version of program review.)

Evaluation is not a term that has a generally accepted meaning distinct from the two previously described terms. However, in the context of the emerging discussion, I propose that we agree to use it to designate a specific activity. The activity in question could also be described as a controlled test of effectiveness, where two or more programs are compared side-by-side at a point in time to see whether the null hypothesis (of no difference in effectiveness between the programs) can be rejected with an acceptable degree of confidence. This type of evaluation is more like the tests of effectiveness performed for new pharmaceutical drugs or for medical treatments than anything we have seen regularly in academia.

I believe it is important to distinguish sharply between such evaluation protocols and assessment. Assessment mechanisms are not designed to look for a trigger level of effectiveness for a go-no go decision on a program. They are designed for continuous improvement of an existing program.

Once the distinction is clear, we can make an informed decision on whether assessment or evaluation is the appropriate methodology for a program.

Going beyond these observations which I believe should not be too controversial, I advance the opinion that evaluation in the sense defined above may not be the appropriate

methodology for a pilot program, at least not unless the outcomes looked for are defined differently than head-to-head test between a mature program and a fledgling one. And here I would draw on an analogy. Let's assume that the learning outcomes for the current GE courses and the pilot alternative are the same (which is not self-evident). Then a valid analogy could be made between a mature mainframe computer of the 1970's and an early version of the Intel CPU chip. Clearly the mainframe at that time would have won hands down. However, the engineers at the time knew better than to use this fact to abandon work on CPU chips. They knew that a solid state design was inherently superior in principle given the underlying physics, and that through continuous incremental improvement the chip would overtake the old transistor-based mainframes. And indeed, now chips have the power of what used to be considered supercomputers.

The moral of this story is that a qualitatively different design (for a chip or an academic program) based on newer, more advanced science or pedagogy will eventually win over initially more effective mature designs based on older paradigms, but not right away. And of course, in the early days of chips, there were plenty of pro-mainframe skeptics.

So this is a plea for us to suspend disbelief and open ourselves to the possibility that a new paradigm, based on recent gains in knowledge and experience nationally, can generate quantum gains in educational effectiveness after a suitable series of cycles of assessment and improvement. I believe that direct inspection of the soundness of the pedagogical methodology should be deciding factor on whether we embark on what I am convinced will be an exciting and reinvigorating voyage of discovery for our colleagues and our students.

Eduardo M. Ochoa

Provost and Vice President for Academic Affairs

Sonoma State University

Rohnert Park, CA 94928 USA

Tel. 707.664.2192

eduardo.ochoa@sonoma.edu