ASSESSMENT AT THE PROGRAM LEVEL: USING ASSESSMENT TO IMPROVE UNDERGRADUATE STATISTICS PROGRAMS

PECK, Roxy & CHANCE, Beth
Cal Poly, San Luis Obispo
USA

With the growing focus on assessment and accountability, programs at many universities are now expected to define learning goals and objectives at the program level (as opposed to the course level) and to devise strategies for assessing whether these goals and objectives are being met. This paper will look at the role of assessment at the program level and how the resulting information about student learning can be used to make informed decisions about curriculum.

INTRODUCTION

While university faculty are certainly comfortable assessing students through exams, homework, and other graded assignments, and may be somewhat familiar with the notion of classroom assessment, assessment at the program level is still not very well understood. In spite of this, increasing pressure from both internal and external sources to understand how programs are performing and how they can be improved is prompting departments to develop and implement assessment plans. In this paper, we will consider the what, why, and how of program assessment, and then will describe the fledgling program assessment efforts for the undergraduate statistics major at Cal Poly, San Luis Obispo.

THE "WHAT" OF PROGRAM ASSESSMENT

Program Assessment is not about assessing individual students or even individual courses, and it certainly isn't about assessing professors. The primary objective of program assessment is to try to find out what students (as a group) are learning as they complete the course work for their major. Its purpose is to improve the quality of student learning through improving curricula and discovering ways to make pedagogy more effective.

"We already do assessment—we grade final exams" is a common response when departments are first asked to consider developing an assessment plan. While there is something to be learned from student performance on exams, assessment at the program level encompasses much more than grading individual students on a course by course basis. If a major is viewed as a collection of courses, each with its own content specific objectives, assessment at the program level asks the hard question—what does it all add up to? What is it that we are hoping this particular collection of courses will produce in the end? Does it add up to more than just the sum of its content parts? Not easy questions, but ones that are definitely worth asking. Program assessment is about trying to provide some answers to these questions.

Program assessment, properly implemented, is a cyclical process that is similar to the familiar scientific process. It begins with the identification of the important educational goals, such as student learning objectives for the program as well as skills and attitudes appropriate to the major. This is one of the most important steps in the process and it may be the most difficult step as well. But, it is also the most critical step because the student learning outcomes identified drive the subsequent steps. These learning outcomes become the basis for ongoing investigation. The second step consists of the planning for assessment—identifying how a student might demonstrate that he or she has achieved a particular learning objective. The final two steps involve collecting and analyzing the appropriate assessment data and then reflecting on what the data reveal about how the program might be improved. Proposed improvements to either the curriculum or pedagogy can then be implemented, and the cycle begins again with the evaluation of the impact of changes made.

THE "WHY" OF PROGRAM ASSESSMENT

As statisticians, we advocate using data to make informed decisions, to monitor processes, and to improve quality. We sometimes even mock those who base policy on anecdotal evidence. It is surprising then that so many of the decisions that we make as academics are based on personal opinion and anecdotal evidence, and that departments of statistics are often skeptical of attempts to measure the quality of our academic programs. Granted, when we do take the time to carefully articulate the characteristics and abilities that we hope to see in the graduates of our programs, it becomes obvious that many are intangible and difficult to measure. But, if we believe what we tell our students about the value of data, the fact that obtaining good information may be difficult shouldn't stop us from trying to measure the things we think are important. We need to put what we preach into practice.

There is an old adage that says "If it ain't broke, don't fix it" that is often used as an excuse to avoid change. And, it is especially easy to avoid change if we also just *assume* that it ain't broke! Of course this strategy ensures that "it" will also never get any better. Perhaps a more reasonable approach, and one consistent with the goals of program assessment, would be "If it is absolutely perfect, leave it alone."

THE "HOW" OF PROGRAM ASSESSMENT

Program assessment must be clearly focused if it is expected to lead to program improvement. A good beginning is an open and frank discussion of the following question: When your students graduate, what is something observable that you think they ought to be able to do? In particular, when developing a statement of learning outcomes, try to stay away from specific content-related descriptions and focus on the big picture. Once comfortable with the list generated, the next step is to consider how you would be able to tell if your students can actually do the things you think your curriculum prepares them to do. Start with one or two of the outcomes that you think are most important, and then think about designing assessment tasks (e.g., exam type items, portfolios and other performance-based assessments like senior projects, exit interviews, etc.) that would allow students to demonstrate whether or not they are achieving a particular learning outcome. Think about when in the program it is most appropriate to conduct the assessment and about the best way to collect the necessary data.

Often assessment items can be embedded in course work—in an assignment or as a question on an exam. Keep in mind though that the data from such embedded assessments needs to also be considered in a more global sense (as compared to the more traditional use of assigning grades to students) to learn what it reveals about students as a group.

Meaningful program assessment requires a significant investment of time. Developing program learning outcomes, designing appropriate assessment items, collecting and evaluating the data, and reflecting on implications for curriculum and pedagogy are not tasks that can be accomplished without thought and effort. Because of this, it is critical that the data collected in assessment efforts be informative and useful to the department. Make sure that assessment items address important learning outcomes and that the items will produce information that will be useful in understanding strengths and weaknesses in the curriculum and in the pedagogies used to deliver the curriculum.

Ultimately, program assessment allows a department to articulate what they value in terms of student learning and to consider the match between what is valued and what the program as taught actually delivers. It is through the reconciliation of discrepancies revealed that programs evolve and student learning improves.

ONE DEPARTMENT'S EXPERIENCE

The Statistics Department at Cal Poly, San Luis Obispo, California, USA is in the early stages of implementing an assessment plan for the undergraduate statistics major. This process has not been an easy one, and we share our experiences with assessment here in the hope that departments that are just now engaging in serious program assessment efforts or who are struggling to get started might find it useful.

Step 1: Developing Learning Objectives

Perhaps the greatest challenge is articulating desired student learning outcomes at the program level. When we first tried to do this, we began with a description of what a statistics major should know that had been developed in the early 1990s. This list was comprehensive and contained 77 bulleted items. In retrospect, we now realize that it was really just a list of the content topics for all of the courses we taught—things like statistics graduates should be able to make inferences about the parameters in simple, polynomial and multiple regression models. Hardly big picture outcomes. An almost identical list could have been obtained by just running the catalog descriptions of all the courses in the major together. But, it did give us a starting point that, along with the ASA Guidelines for Undergraduate Statistics Majors (2000), enabled us to collectively identify some common and higher priority threads and eventually come up with some overall program objectives—outcomes that we thought (hoped) would emerge as students completed the statistics curriculum as a whole. The key was to view the curriculum as more than just a collection of independent and unrelated courses. After several iterations and revisions, the department finally came up with the following statement of student learning outcomes.

Statement of Student Learning Objectives - Statistics Department

The graduate will:

1. Have good working knowledge of the most commonly used statistical methods including
    a. statistical modeling
    b. efficient design of studies and construction of effective sampling plans
    c. exploratory data analysis
    d. formal inference procedures
2. Have background in probability, statistical theory, and mathematics, including especially calculus and linear algebra
3. Be able to synthesize and apply this knowledge and to tailor methods used to the problem at hand, understanding the limitations of the procedures and the appropriate scope of conclusions
4. Communicate effectively (written and oral) with skills in collaboration (within and between disciplines) and teamwork, and in organizing and managing projects
5. Have good mastery of several standard statistical software packages and facility with data management strategies
6. Have a focused concentration in an area of application outside the discipline of statistics

The graduate will have received:

1. Experience with real data and authentic applications
2. Frequent opportunities to develop communication skills
3. Capstone experiences
4. Frequent interaction with faculty and timely advising
5. Exposure to statisticians and statistical applications outside the Cal Poly community

Even though it was time consuming and an activity that we probably would not have undertaken if it had not been mandated by the university administration, developing these student learning objectives did turn out to be a worthwhile exercise for the department. The iterative process used to move from a long list of content topics to a more global description of student learning and experiences at the program level helped to create a shared vision of program goals. The statement of desired student learning objectives is something that can be shared with incoming students, and it also guided our first serious efforts at program assessment and the self-study for our most recent internal and external program review.

Step 2: Collecting Data

After taking most of an academic year to develop the program learning objectives, we started to think about a plan for collecting baseline data. The university mandate gave departments a year to implement an assessment plan that would generate some baseline data and that would then become the foundation for ongoing assessment efforts.

Since we were inexperienced at program assessment, we decided to focus first on learning outcomes that we thought were very important, but for which we had little real evidence to indicate whether or not they were being achieved by students. In particular, we chose to design an assessment that would provide information about learning objectives 3 and 4 on the first list above. To gauge the extent to which we were providing students with the desired experiences in the second list above, a student exit survey was also designed.

Our initial instrument, given to graduating seniors as part of an assignment in a capstone course, included two questions designed to assess learning outcomes at the program level. The first item was written to see if students in their senior year could demonstrate that they could choose appropriate methods of analysis and that they understood the limitations of these methods as well as what kinds of conclusions can be drawn from the analysis. This item is actually similar to one that we have used in our statistical literacy service courses! The second item was designed primarily to assess communication skills. In the second year, an item that examined students' understanding of sampling variability was added, and in the third year, data were collected on a question that required students to explain the meaning of a P-value in the context of a research scenario.

To provide an idea of what a sample assessment item might look like and what can be learned from the assessment effort, consider the item shown below.

The Microsoft Excel(2000) menu provides this information for an unpooled two-sample t-test:

> t-Test: Two-Sample Assuming Unequal Variances analysis tool and formulas
> This analysis tool and its formulas perform a two-sample student's t-test. This t-test form assumes that the variances of both ranges of data are unequal; it is referred to as a heteroscedastic t-test. You can use a t-test to determine whether two sample means are equal. Use this test when the groups under study are distinct. Use a paired test when there is one group before and after a treatment.

Would you recommend any modifications to this description? If so, rewrite the description.

We thought this was a great question and we were also pretty confident that our students would do well on this task. We even thought we were throwing them a softball by asking them the easier task of critiquing a statement rather than generating their own! Well, this is why it is important to collect data, even when you think you know how it will turn out. We expected that students would rewrite to correct the obvious statistical errors and also transform the description into something coherent. We were *so* wrong! Of sixteen respondents over two years, five indicated that the description was fine as is and several only added statements about sample size or population normality. Only one student corrected the erroneous statement about using a t-test to determine whether two sample means were equal.

We fared better on the question that assessed understanding of sampling variability, but were unpleasantly surprised at how students responded to the other two assessment items. We had been relatively sure that a senior statistics major would be able to explain what a P-value means to a non-statistician researcher, but in fact most students gave a less than stellar interpretation. Responses ranged from reasonable (The probability of getting a sample like the one used if the null hypothesis were true) to somewhat garbled (probability of getting sample data deviating as much or more than the expectation) to just plain wrong (P-value is the power of the test; P-value is % of area under curve outside of the rejection region). A bit depressing, but it was a real eye opener.

These examples illustrate the real value of assessment activities. It showed clearly that some of the broad program level learning objectives that we thought to be really important and

that we believed students should be able to do upon completing our curriculum were not being achieved.

The exit survey was also revealing, and did provide valuable data on the extent to which students were satisfied with the experiences provided by the curriculum and on student perceptions of their ability to apply the tools they had learned. Using the survey data and the data from the assessment items, we also discovered that there was not always a strong correspondence between a student's confidence in his or her ability and his or her actual performance on such assessment items.

Step 3: Reflect and Respond

After a bit of departmental soul-searching, a number of changes to the curriculum were made, including the addition of a senior-level consulting course and some revisions to our senior project capstone experience. We also decided to implement a freshman level course for statistics majors that would provide them with an early exposure to some of the fundamental concepts of statistics before they became immersed in courses focused more on statistical methods and theory. In addition, the department is currently considering curriculum-wide changes in order to place more emphasis on communication and to provide more consistent repetition of basic concepts (such a p-value and sources of variability, for example), while also helping students see the connections and distinctions between methods across courses (for example, because regression and ANOVA are covered in separate courses and the ANOVA course primarily used examples where the data were from designed experiments, many students did not realize that experimental data might also be analyzed using regression and tended to choose between regression and ANOVA as a method of analysis based solely on whether the data were observational or experimental).

As the next step, we have been embedding assessment items in exams and assignments in the new consulting course. We are hopeful that students will be a bit more serious in their attempts to answer the assessment items if the responses count toward a course grade than if they are just seen as part of the exit survey. Student performance can then be evaluated by a separate panel of faculty reviewers. We also now have five years of data on the senior exit survey through which to monitor changes and trends. Results of these assessments are discussed at the departmental retreat each fall.

We are hopeful that these changes will address some of the deficiencies observed, but only time (and a continued commitment to assessment!) will tell.

ADVICE TO THOSE PLANNING FOR PROGRAM ASSESSMENT

In the spirit of sharing what we have learned (sometimes the hard way!), we offer the following advice.

- Get departmental buy-in. Work to ensure that the faculty understands the value of program assessment and work to involve the entire department in the assessment effort. Program assessment is not something that can be accomplished by a single person in response to an administrative edict!
- Make sure that the data you collect is on something that is important to the department. Spend some time figuring out what the interesting questions are. If you are going to have to invest time and effort in program assessment, you may as well make it meaningful. Don't collect data in order to answer a question you don't care about!
- Resist the temptation to trivialize program learning objectives just to make assessment easy. It is better to have an imperfect measure of a real and valued desired outcome than a perfect measure of a trivial one.
- Use embedded assessments where possible. Students take them more seriously and it simplifies the data collection process by eliminating the need for separate assessments.

- You don't have to assess every student. If you have a small number of majors, it may not be a problem to include all students in your program assessment, but with a large program, use what you know about sampling!
- Start with small steps. Just as it is not necessary to assess every student, it is also not necessary to assess every student learning objective every year. The task will seem overwhelming if you attempt to do it all from the outset. The most important point here is to take that critical first step. Start by focusing on one or two outcomes that the department feels are really important. Make these the focus of your assessment efforts until you are satisfied that you understand student learning with respect to these objectives. Then you can turn your attention to other learning objectives.
- Make use of local and online resources. Find out what other departments at your university and statistics departments at other universities are doing. You may find it useful to meet occasionally with representatives of other departments to share ideas (and sometimes vent!). Many universities have program learning objectives online and some even have their entire assessment plans on the web.

CONCLUSION

Program assessment is often implemented as a consequence of administrative edicts, and can add to faculty workload. But, it is likely that departments will be required to engage in this activity at least to some degree, and if faculty can get past the usual initial skepticism, program assessment can provide real opportunities for reflection and for improvement of student learning. To be successful, program assessment requires broad departmental cooperation, and it is also important to approach this activity with realistic expectations. It will not be possible to measure student learning precisely and issues of sampling, confounding, and authenticity will come up. Perfect data and clear cut answers are not the norm in program assessment. But, this should not keep us from embracing the challenge of understanding the impact of our curriculum and pedagogies on student learning.

This paper is based on a presentation made at the 2005 Joint Statistics Meetings and referenced below.

REFERENCES

Assessment References for Undergraduate Mathematics. Retrieved from: http://www.maa.org/SAUM/saumbib4.htm

Committee on the Undergraduate Program in Mathematics (CUPM) (1995). Assessment of Student Learning for Improving the Undergraduate Major in Mathematics.*Focus: The Newsletter of the Mathematical Association of America*, *15*(3), 24-28.

Curriculum Guidelines for Undergraduate Programs in Statistical Science (2000). Retrieved from: http://www.amstat.org/education/index.cfm?fuseaction=Curriculum_Guidelines.

Peck, R., & Chance, B. (2005). Assessing Effectiveness at the Program Level: Undergraduate Statistics Program Evaluation, *Proceedings of the 2005 Joint Statistics Meetings*, Alexandria, VA, USA: The American Statistical Association.