

# Streamlining Program Assessment for ABET: What to do with all that data

James Allert

*University of Minnesota Duluth*

***Abstract*—One of the most daunting tasks of ABET accreditation is preparation of program assessment reports. Since these are necessarily a distillation of data from numerous sources (including all courses that have been assessed) the process can be laborious and error-prone. This paper presents a software solution called Program Assessor developed by the author at the Department of Computer Science at The University of Minnesota Duluth (UMD) that automates the process of data compilation, analysis, summarization and report generating.**

***Index Terms*—Engineering education, ABET, accreditation.**

The challenge of ABET accreditation is one that no engineering program can take lightly. It involves the collection of direct measures from every course in a program and from a variety of other sources to document and provide evidence to support the claim that course and program objectives are being met. The data stream feeding into this process comes from multiple sources, in multiple formats and must somehow be managed and made sense of. Ultimately it must be condensed down into meaningful summaries of objectives, outcomes and performance criteria satisfaction at both the course and program level.

The stream of data does not end with program accreditation either. Accreditation involves ongoing monitoring of courses and of the program itself. Besides evidence that objectives and outcomes are being met it also requires documentation of the process of continuous quality improvement. This entails an endless cycle of assessment and reassessment at both the course and program level.

Mastering this data streaming process and automating the tasks involved in the use of such data are crucial to the survival of programs and the maintenance of the sanity of those involved. This paper presents one method that we have devised to be particularly easy to employ and a powerful tool for taking control of these tasks.

## *I. Introduction*

The ABET accreditation process is familiar to most US engineering programs and has guided engineering education for over 75 years. Although the particulars of the process are subject to annual changes, the overall thrust of the endeavor remains the same – documentation of processes that assess and continually review how well program outcomes and performance criteria are being met.

As a result of this need for accountability engineering departments have struggled to put in place effective means of program assessment based on direct measures. Direct evidence, in the form of student projects, exams, reports, and other measures, are used to substantiate the program's claims that its outcomes are being satisfactorily met.

ABET requires specific evidence that engineering programs have enabled a series of 11 abilities in its graduates. These abilities (known as “a through k”) form the basis of all engineering program assessments. Providing evidence that a program has satisfied a-k requires a large amount of data gathering, analysis and synthesis. There are a number of ways to approach the data processing tasks. Much of the data can be manually gathered, but without statistical software it is difficult to manage the process. Solutions range from inexpensive desktop software packages that manage some of the critical tasks to expensive, fully-integrated, commercial, web-based solutions.

## *II. Literature review*

A number of assessment-related software solutions have been presented in various engineering education forums. A number of the most recent and well-established ones are mentioned here. Although the list is not exhaustive, it does represent the wide range of solutions that are available. It should be noted that non-engineering disciplines often have their own accreditation boards and are responsible for similar program assessment reports. Many of these disciplines, especially the field of education, have developed tools to automate the process and provide the feedback necessary to foster continuous quality improvement.

There are several major types of automated program assessment tools. Web-based tools are desirable because they are easily deployed to those who need them and because they may allow for collaborative interaction. Heinze et. al. [1] use web-based automated assessment to allow students to take mock FE exams online and, in this way, develop an awareness of their strengths and weaknesses as they prepare for the real thing. These results are not factored into the ABET assessment process but provide students with a solid understanding of the extent to which ABET abilities have been mastered. Such information could be advantageously used by departments to assess the strengths and weaknesses of their seniors prior to taking the FE exams.

An example of a web-based tool that directly addresses the departmental need for ABET reporting is the ECAT system developed by Deborah Trytten [2]. This system summarizes the extent to which each of the ABET outcomes are being met using tallies of assignments which relate to each outcome. ECAT can provide evidence that each outcome was assessed although it cannot facilitate an evaluation of the level of performance relative to each outcome.

The problem of gathering actual student performance measures and integrating them into a system from which ABET outcome evaluation can be made is described by Booth [3]. He proposes the development of a database system to do this. Ultimately an online database, incorporated with student and departmental reporting tools may be the most efficient and effective means to solving the accreditation reporting dilemma.

Rather than using the online database model, Burge and Leach [4] have developed a Microsoft® Excel spreadsheet that is used to condense student performance data into results that apply to ABET outcomes. The results average the scores for each outcome. This is a desktop solution that is easy to use and efficiently produces summary reports for the departmental user.

A more integrated approach, but a costly one, is to have assessment integrated into student e-portfolios. McNair et. al. [5] describe such a system in relation to ABET outcomes that relate to profession skills. Fully integrated systems that consolidate student work, as incorporated in e-portfolios into departmental reports are as yet only implements in expensive proprietary software and would be a good candidate for an Open Source project.

Each of these approaches has its advantages and disadvantages. The software program described in this paper is an Excel spreadsheet, similar in concept to Burge and Leach but much more extensive in its reporting capabilities and without some of the problems inherent in the averaging of student data (discussed later in this paper).

### *III. Program assessment*

ABET program assessment is a complex undertaking. Every educational program has certain objectives in mind. From the ABET standpoint, important evidence that a program is successfully achieving its objectives comes from its graduates. Evidence of the degree to which graduates were adequately prepared for their chosen field comes ideally from those who are 3-5 years into their careers. Surveys of alumni and other similar instruments often are used to provide this information. This information can be incorporated into a final program assessment report for accreditation purposes as indirect evidence of program objective fulfillment.

Such information is valuable but not timely. If a problem exists in an engineering program it is always better to know sooner rather than hearing from graduates 5 years later. For this reason the outcomes of current student learning are the primary focus items of program assessment.

Student learning outcomes are associated, and often synonymous with, ABET-designated abilities (often called the ABET a-k because of their formal listing in ABET literature). Those abilities are:

- (a) an ability to apply knowledge of mathematics, science, and engineering
- (b) an ability to design and conduct experiments, as well as to analyze and interpret data
- (c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability
- (d) an ability to function on multi-disciplinary teams
- (e) an ability to identify, formulate, and solve engineering problems
- (f) an understanding of professional and ethical responsibility
- (g) an ability to communicate effectively
- (h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context
- (i) a recognition of the need for, and an ability to engage in life-long learning
- (j) a knowledge of contemporary issues
- (k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

Direct-measures are used to legitimate a program's claim that each of these is satisfactorily enabled in students upon graduation. This means that, at a minimum, one direct measure would be required for each of the eleven outcomes. One measure may not be considered sufficient to justify a claim that the outcome has been achieved. It is common to 'triangulate' outcomes by requiring several different measures as evidence. Thus, a minimum of 22 or perhaps 33 measures would be more appropriate (two or three specific assessments of student performance per outcome).

Student learning outcomes are broad statements of abilities students are expected to have upon graduation. Each student learning outcome is achieved through the mastery of a set of related skills. These critical skills, known as performance criteria, are monitored in the accreditation process as well. Program performance criteria are specific tasks and capabilities that students must demonstrate proficiency in as they acquire the abilities that define each outcome. These criteria are most often identified by program faculty consensus. For example, a program might decide that there are three performance criteria (a1, a2, a3) related to student outcome a. Direct evidence must be gathered to demonstrate that each performance criteria has been satisfied by graduating students in addition to evidence that outcome a has been met overall.

The number of performance criteria required may be as little as one, but more often consists of a set of three or four items. It is less work if there are fewer of these but it is usually the decision of faculty as to what these essential skills are and, in some instances, faculty may generate a long list. If each outcome had only three performance criteria then a total of 33 direct measures would

be required, at a minimum. If each criterion relies on more than one piece of direct evidence to legitimate the claim that it has been satisfied then the amount of data grows rapidly.

Table 1 indicates the number of direct measures required by different configurations in which there are eleven (a-k) outcomes. Each of the eleven outcomes (a-k) must have 1 or more performance criteria. Each Performance criteria must have one or more direct measures that provide evidence of performance.

Table 1. Direct measure totals (*columns 1,2,3,4 indicate direct measures reported per criteria*)

<i>Criteria per outcome</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
1	11	22	33	44
2	22	44	66	88
3	33	66	99	132
4	44	88	132	176
5	55	110	150	220
6	66	132	198	264
7	77	154	231	308

Table 1 illustrates the data proliferation problem that programs must address in order to submit an accreditation report. Even a small number of performance criteria and a few direct measures for each can add up to hundreds of data values that must be stored and analyzed. These are not the only reports of course. An additional set of reports for each of the student outcomes (a-k) is also required, based on the direct measures that directly support it. This makes ABET data reporting a formidable challenge.

#### *IV. Program assessment data requirements*

Direct evidence of student performance is gathered from a variety of sources, most often coursework, exams, projects, and other graded instruments. These are easily provided by course assessments. Indirect evidence of criteria and/or outcome satisfaction, in the form of student survey responses, faculty course evaluations, outside review and other sources can be used as supplementary evidence. As a result, it is not uncommon for a single course to be able to provide dozens of items of direct evidence.

For some instructors, the more direct measures that can be incorporated into a course assessment the better. When this is done the effect is to greatly multiply the amount of data consolidation work that must be done.

The emphasis on quantity increases the data processing unnecessarily. Although a course might employ numerous quizzes, exercises, homework assignments, and other direct measures, only the

most substantive assessments need to be reported. Using only the best performance indicators, rather than all, or dozens of them, can reduce the number of data items required to a handful in most instances.

In addition to the number of direct measures, the manner in which these results are reported is also worth paying attention to. Ideally, these results should be compiled and reported using percentages derived from rubric category frequencies.

Rubric categories, such as “Unsatisfactory, Minimally satisfactory, Satisfactory and Exemplary” are typically used to group results for each measure based on varying levels of performance. In this way it is possible to determine how many student performances there are in each category. The category frequency counts can then be turned into percentages for the purposes of comparison. Rubric category percentages are much more descriptive than arithmetic means because they provide more than one measure of performance. Table 2 presents a typical course assessment report spreadsheet showing the rubric category frequency tallies for a number of direct measures from several courses.

Table 2. Typical results of direct measure assessments

Assessment	Outcome	Criteria	Unsatisfactory	Minimal	Satisfactory	Exemplary
Exam II	a	a1	30	19	35	16
Exam I (Part I)	a	a1	21	25	41	13
Exam I (Part II)	a	a1	11	17	50	22
Exam I (Part IV)	a	a1	9	29	43	19

The spreadsheet shown in Table 2 is only a small excerpt of a real course assessment data collection with the results of course assessments from fall 2007 to spring 2010. In this case, there were eleven student outcomes (a-k) and 31 performance criteria (a1, a2, b1, b2, etc). One course may have multiple assessments tied to a single performance criterion. A single course may also have assessments tied to more than one performance criterion.

This is the point at which the assessment problem becomes most difficult. In order to determine whether a performance criterion (such as criterion a1) has been met, the direct measures of performance related to criterion a1 must be collected from all courses reporting direct measure results related to criterion a1. Figure 1 illustrates what such a report looks like when generated by the Program Assessor software.

Student Outcome 2: an ability to apply knowledge of computing and mathematics appropriate to the discipline.								
Criteria a1. Students understand information representation								
	Exemplary		Satisfactory		Minimal		Unsatisfactory	
	%	n	%	n	%	n	%	n
Weighted Aggregates	26.5%	107.0	61.8%	176.0	19.1%	77.0	10.7%	46.0
CS-150 - Exam II, Fall, 2007	31.8%	28	40.9%	39	13.0%	12	13.0%	12
CS-150 - Exam I (Part I), Fall, 2007	16.0%	15	61.0%	61	24.0%	24	16.0%	16
CS-150 - Exam I (Part II), Fall, 2007	21.7%	20	52.3%	48	17.4%	16	3.7%	8
CS-150 - Exam I (Part IV), Fall, 2007	20.0%	24	62.5%	76	25.0%	20	4.8%	0
CS-260 - Final, Part I, Spring, 2008	47.5%	19	50.0%	12	12.5%	5	10.0%	4

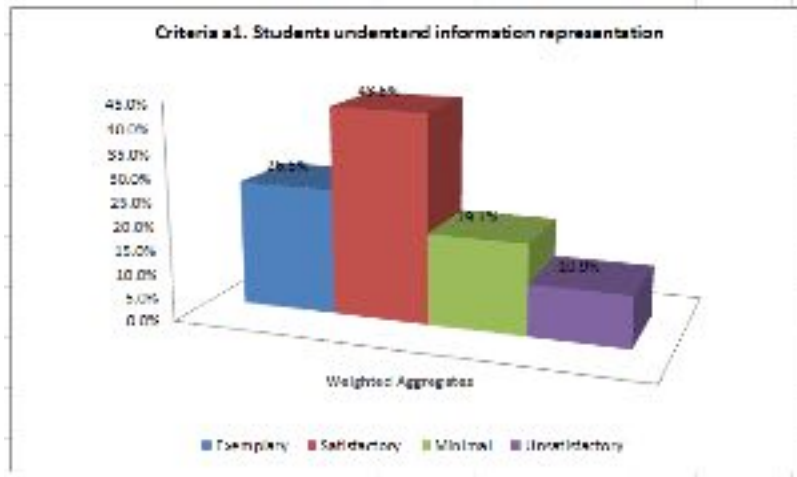


Figure 1. Detailed direct measure report for program performance criterion a1.

Figure 1 lists the direct measures in support of criterion a1 down the left side of the table. Each measure lists the course identification number, assessment name, semester and year. The performance results are broken down by category (Exemplary, Satisfactory, Minimal, Unsatisfactory) both as n values and percentages. The n value is the number of students performing on a particular assessment at the given level. The percentage is the percentage of all students who performed at that level. Thus, for the first direct measure we find that:

- The measure was derived from course CS-150
- The specific assessment event was Exam II
- The semester and year this data comes from is Fall 2007
- The percentage of students performing at an exemplary level was 31.8%
- The actual number of students performing at the exemplary level was 28

Similar percentages and n's are provided for all other performance categories for this measure. All direct measures include these same items of information. In the example shown in figure 2 there are 4 assessments from CS-150 and one from CS-260 that relate to criterion a1. The overall results are shown at the top of the table. These are percentages based on the total n's for each performance category unless some direct measures have been weighted differently from others, in which case the results reflect the weighting.

From Figure 1 we learn that overall the level of performance demonstrated by the direct measures of criterion a1 is:

- Exemplary – 26.5% (n = 107)
- Satisfactory – 43.6% (n = 176)
- Minimal – 19.1% (n = 77)
- Unsatisfactory – 10.9% (n = 44)

Note that these are the results for one performance criterion only. There will be a separate report like this for each performance criterion. Program Assessor automatically generates these reports, drawing from the original course assessment spreadsheet and grouping by performance criterion.

Performance data drawn from individual direct measures must also be used to support claims that student outcomes have been met. This report is similar to that shown in Figure 1 with the difference being that it uses all of the direct measures pertaining to all of its criteria. In other words, the data from direct measures used to justify criterion a1 may be combined with data from all of the direct measures used to support criteria a2, a3, and all other a-related criteria. As a result, this compilation could be very large. One way to reduce the size is to be more selective about which direct measure data is used. It is not true that all direct measures used to support criteria need to be included in the report that supports student outcome satisfaction.

Figure 2 illustrates, in a manner similar to that used in Figure 1, how the supporting data is presented. In this case the student outcome is ABET outcome e.

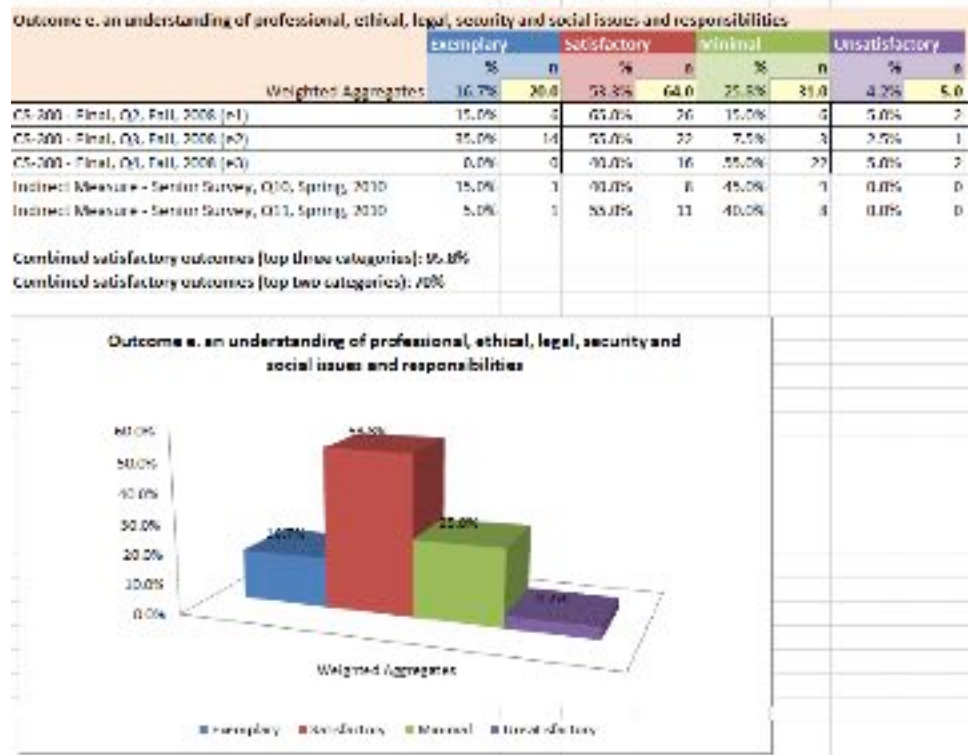


Figure 2. Detailed direct measure report for student performance criterion e.



Figure 2 lists the direct measures in support of student outcome e down the left side of the table. Each measure lists the course identification number, assessment name, semester and year. The performance results are broken down by category (Exemplary, Satisfactory, Minimal, Unsatisfactory) both as n values and percentages. The n value is the number of students performing on a particular assessment at the given level. The percentage is the percentage of all students who performed at that level. Thus, for the first direct measure we find that:

- The measure was derived from course CS-300
- The specific assessment event was the final exam, question 2
- The semester and year this data comes from is Fall 2008
- The percentage of students performing at an exemplary level was 15.0%
- The actual number of students performing at the exemplary level was 6

Similar percentages and n's are provided for all other performance categories for this measure. All direct measures include these same items of information. In the example shown in Figure 2 there are 3 direct assessments from CS-300 (an ethics course). The overall results are shown at the top of the table. These are percentages based on the total n's for each direct measure performance category unless some direct measures have been weighted differently from others, in which case the results reflect the weighting.

From Figure 2 we learn that overall the level of performance demonstrated by the direct measures of criterion a1 is:

- Exemplary – 16.7% (n = 20)
- Satisfactory – 53.3% (n = 64)
- Minimal – 25.8% (n = 31)
- Unsatisfactory – 4.2% (n = 5)

Note that these are the results for one student outcome only. Program Assessor generates a separate report like this for each student outcome. It is also important to note that, although only direct measures are used in the computations, the performance results for indirect measures are also included. This allows the reviewer to incorporate other perspectives into an evaluation of how well this particular outcome has been achieved. In Figure 2 there are two lines identifying indirect measures. These were the answers to senior survey questions in which students were asked to rate their understanding of ethical, legal, security or social issues.

#### *V. Program assessment outputs*

We have seen how direct measure data are used to describe the proficiency of students in a single performance area (Figure 1) and how a similar approach is used to give evidence for proficiency in a student outcome area (Figure 2). At the program level, it is also instructive to know how each of the performance criteria for a particular student outcome match up. Figure 3 indicates what such a report looks like. The data in each line of the table comes from the results of different criterion reports (such as Figure 1). For example, if student outcome a has two

performance criteria (a1 and a2) then the results for a1 and a2 can be displayed in relation to one another for comparison purposes. This is shown in Figure 3.

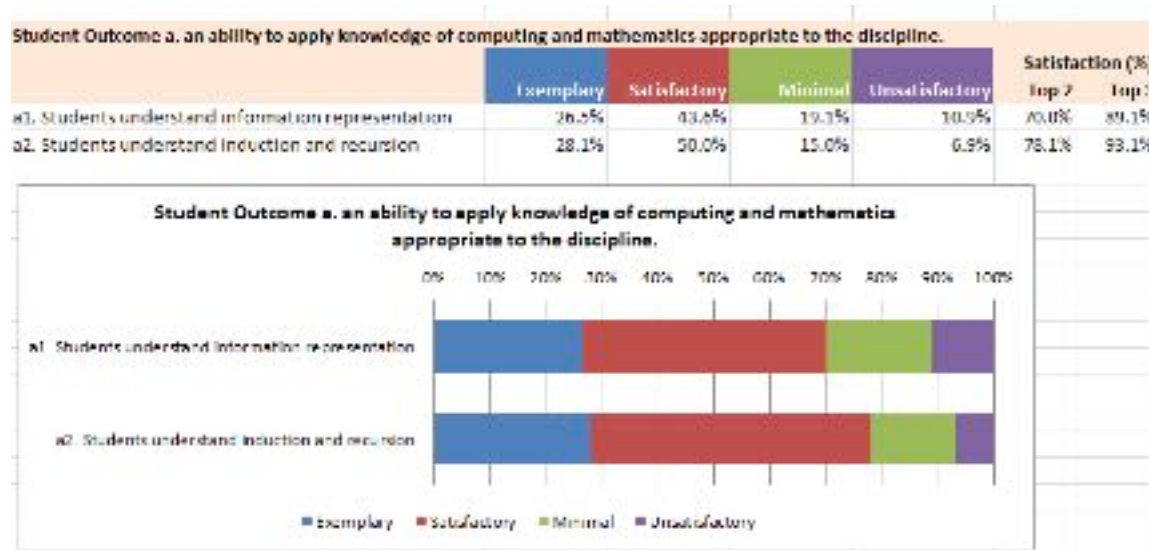


Figure 3. Comparison of performance criteria results for student outcome a.

Figure 3 provides side-by-side profiles of a1 and a2. By comparing the corresponding percentages for each category the distribution of achievement can be evaluated for the group as a whole as well as individually. Strengths and weaknesses are easily recognizable when data is displayed in this manner. The percentages used in the comparison derive from the individual results (ie. Figure 1).

At the heart of ABET accreditation is the evidence that each of the abilities (a-k and others) is being enabled. These student learning outcomes are displayed together in Figure 4.

Program Outcomes					Satisfaction (%)	
	Exemplary	Satisfactory	Minimal	Unsatisfactory	Top 2	Top 3
a. an ability to apply knowledge of computing and mathematics appropriate to the discipline	27.0%	48.4%	27.0%	0.0%	74.2%	89.2%
b. an ability to analyze a problem and identify and define appropriate computing requirements	19.4%	54.4%	16.4%	10.4%	71.4%	84.4%
c. an ability to design, implement, and evaluate a computer-based system, process, component, or program	25.1%	52.0%	28.7%	0.2%	74.1%	89.8%
d. an ability to function effectively on teams to accomplish a common goal	19.4%	54.4%	16.4%	10.4%	71.4%	84.4%
e. an understanding of professional, ethical, legal, security and social issues and responsibilities	16.7%	51.3%	25.8%	4.2%	70.0%	85.8%
f. an ability to communicate effectively with a range of audiences	19.4%	54.4%	16.4%	10.4%	71.4%	84.4%
g. an ability to assess the local and global impact of computing	30.5%	40.0%	3.0%	6.0%	80.5%	89.5%
h. recognition of the need for and an ability to engage in continuing professional development	15.0%	50.0%	17.5%	17.5%	60.0%	87.5%
i. an ability to use current techniques, skills, and tools necessary for computer practice	15.4%	50.7%	12.1%	11.7%	77.2%	84.0%
j. an ability to apply math for calculations, algorithms, proofs, and to theory that derives into software	15.7%	54.5%	3.4%	10.4%	80.2%	89.8%
k. an ability to apply design and development principles in the construction of software systems	19.4%	54.4%	16.4%	10.4%	71.4%	84.4%
l. an ability to understand the scientific method and apply it in a laboratory setting	20.1%	55.5%	8.0%	1.0%	80.7%	88.6%

Figure 4. Comparison of student outcome results for student outcomes a-l.



categories (exemplary and satisfactory). The bar chart illustrates the top two performance sum for each of the two data collection cycles for this student outcome. The first cycle produced a performance measure of 70%. In other words, 70% of students achieved either exemplary or satisfactory performance in this area. The second cycle produced 76.1%. Although we cannot determine if this increase is statistically significant we have some evidence that the trend is proceeding in the desired direction and will continue to monitor the differences between the values as more data for this outcome is acquired in the second data collection cycle.

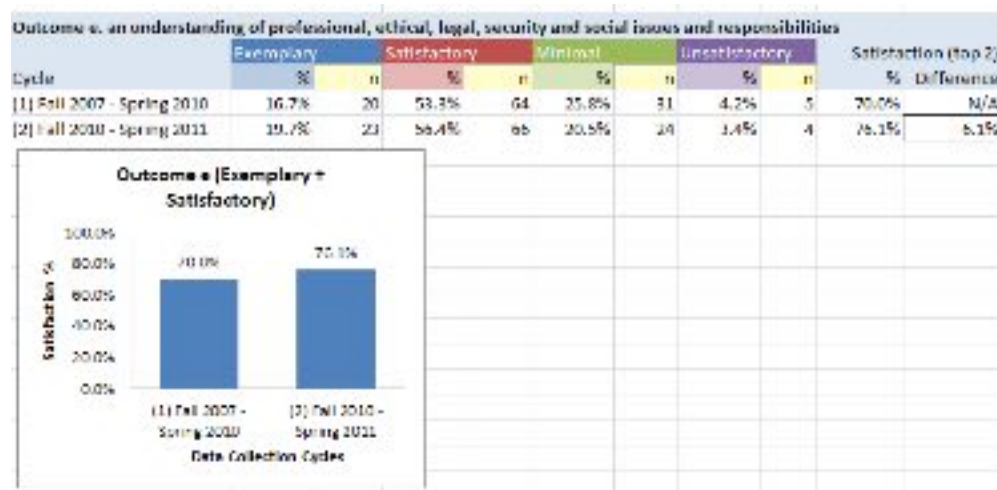


Figure 6. Continuous improvement monitoring by comparing data collection cycle results

## VI. Conclusion

Program Assessor has proven to be a valuable resource for ABET accreditation reporting in the Department of Computer Science at UMD. It requires a minimal amount of up-front data entry (course assessment rubric performance results as shown in Figure 1) and no further data entry. It requires no Excel skills or coding. Once the data is in, reports such as those shown in Figures 2-6 are generated automatically at the touch of a button. In our department we have adopted 12 program student-learning outcomes and a total of 31 various performance criteria. In the most recent ABET report prepared by our department almost 300 direct measures of student performance were provided by faculty for inclusion in the accreditation analysis. Compiling detailed reports of how well the computer science program achieved each student outcome and performance criteria would be incredibly time-consuming if done by hand. All these reports are virtually instantaneous with the Program Assessor software however.

The ability to avoid laborious data tasks is important when program assessment is at stake. It not only removes a huge faculty and administrative burden, it also allows both faculty and administrators more time to reflect on the results. Automated reporting serves the accreditation process best by facilitating reflection and improvement. We have found that the real value of Program Assessor is that it moves the assessment agenda along from the mundane to the important issues, allowing more time to discover the strengths and weakness of the program and much less time crunching the numbers that reveal them.

## Bibliography

- [1] L.R. Heinze, J. M. Gregory, and D. J. Bagert, "Web-based Automated Assessment for ABET Programs" *Proceedings of the 32<sup>nd</sup> ASEE/IEEE Frontiers in Education Conference, Boston, MA, Nov. 6-9 2002, Session TIB.*
- [2] D. A. Trytten, "ECAT: An Electronic Data Analysis Tool for ABET Display Materials", *Proceedings of the 40<sup>th</sup> ASEE/IEEE Frontiers in Education Conference, San Diego, CA, Oct. 27-30 2010, Session TIH-1.*
- [3] L. Booth, "A Database to Promote Continuous Program Improvement" *Proceedings of SIGITE'11, Minneapolis, MN, Oct. 19-21 2006, pp. 83-88.*
- [4] L. L. Burge and R. J. Leach, "An Advanced Assessment Tool and Process" *Proceedings of SIGCSE'10, Milwaukee, WI, March 10-13 2010, pp. 451-454.*
- [5] L. McNair, M. Paretto, M. Knott and M. L. Wolfe, "Work in Progress: Using e-Portfolio to define, Teach, and Assess ABET Professional Skills" *Proceedings of the 36<sup>th</sup> ASEE/IEEE Frontiers in Education Conference, San Diego, CA, Oct. 28-31 2006, Session TIJ-13.*