Integrated FCAR Model with Traditional Rubric-Based Model to Enhance Automation of Student Outcomes Evaluation Process

Dr. Fong K. Mak, Gannon University

FONG MAK, P.E. received his B.S.E.E. degree from West Virginia University in 1983, M.S.E.E. and Ph.D. in Electrical Engineering from the University of Illinois in 1986 and 1990. He joined Gannon in 1990. He was the Chair of Electrical and Computer Engineering at Gannon University from 2001 till 2014 and the Program Director for the professional-track Gannon/GE Transportation Embedded System Graduate Program for 2001-2014. He is now the professor of the department.

Dr. Ramakrishnan Sundaram, Gannon University

Dr. Sundaram is a Professor in the Electrical and Computer Engineering Department at Gannon University. His areas of research include computational architectures for signal and image processing as well as novel methods to improve engineering education pedagogy.
Integrated FCAR Model with Traditional Rubric-Based Model to Enhance Automation of Student Outcomes Evaluation Process

Abstract: The Electrical and Computer Engineering (ECE) department at Gannon University has been through two successful ABET accreditations, in 2005 and 2011, with the use of the Faculty Course Assessment Report (FCAR) model. In the 2005 cycle of accreditation review, essential FCAR methodology was used; whereas, in the 2011 cycle, the concept of key assignments, with the well-defined process to generate justifiable objective evidence, was used to augment and further improve the FCAR assessment model adopted. In either cycle, student outcomes (SO) are directly assessed with supporting evidence for the well-defined performance vector termed EAMU where E stands for Excellence, A for Average, M for Minima, and U for Unsatisfactory. However, in either cycle of processes, there were no refined performance indicators (PI) defined for each SO. In the assessment model for the current cycle, a set of PIs are defined for each SO. However, we rapidly realize that if, for example, a set of three PIs are defined for each SO, the evaluation effort will be at least three times more time consuming.

To further improve the assessment model used, the traditional rubric-based assessment model is augmented by classifying courses in the curriculum to three levels: introductory, reinforced, and mastery. It is customary for the traditional rubric-based assessment model to include only the courses in the mastery level for the program outcomes assessment. The drawbacks of looking only at courses at the mastery level are: (1) lack of information needed at the lower level to identify the root cause of the deficiency when the symptom occurs at the higher level courses; (2) lack of the mechanism to compute a clear indicator such as the Student Outcomes (SOs) performance index based on Performance Indicators (PI) of that SO in order to facilitate the automation of the evaluation process.

In this paper, a novel approach is presented to demonstrate how a traditional rubric-based approach can be integrated with the FCAR assessment approach to allow computation of the SO performance index from roll-up data. The performance index is calculated based on the weighted average of relevant PIs for the three different levels of courses. Analytic results on how the SO performance index measured up against the heuristic rules used previously are discussed. Last but not least, results of how the SO performance index can be used to address the overall attainment of the SO expectation are shown.

I Introduction

The ECE department at Gannon University has been through two successful ABET accreditations, in 2005 and 2011, with the use of a well-documented Faculty Course Assessment Report (FCAR) model [1-3]. In the 2005 cycle of accreditation review, the essential FCAR methodology report in [1] was primarily the model adopted; whereas, in the 2011 cycle, the concept of key assignments with the well-defined process reported in [4] was used to generate justifiable objective evidence. This objective evidence was used to augment and further improve the FCAR assessment model. In either cycle, student outcomes (SO) are directly assessed with supporting evidence for the well-defined performance vector termed EAMU where E stands for Excellence, A for Average, M for Minima, and U for Unsatisfactory. However, in either cycle of processes, there were no refined performance indicators (PI) defined for each SO. In order to prepare for the next cycle of accreditation and to better refine our assessment of the program
effectiveness, it is necessary to include performance indicators (PIs) for each SO. This would let us with the identification of areas for improvement. However, we rapidly realize that if, for example, a set of three PIs are defined for each SO, the evaluation effort will be at least three times more time consuming. In addition, there are some significant challenges such as (a) how to assess and augment the PIs into the existing process (b) making the PIs acceptable to both, the faculty within the department and, the external evaluator of the program.

The traditional rubric-based assessment model which is widely used by universities is studied first [5-15]. The essence of the traditional rubric-based assessment model lies in classifying the courses in the curriculum to three levels: introductory, reinforced, and mastery. It is customary for the traditional rubric-based assessment model to include only courses in the mastery level for the program outcomes assessment. The drawbacks of looking only at the courses at the mastery level are: (1) lack of information needed at the lower level to identify the root cause of the deficiency when the symptom occurs at the higher level courses; (2) lack of mechanism to compute a clear indicator such as the Student Outcomes (SOs) performance index based on Performance Indicators (PI) of that SO in order to facilitate the automation of the evaluation process. As a small department with limited resources, we need automation of data collection and compilation, which, therefore, enables us to focus more on data evaluation for improvement. Better yet, if an ultimate clear indicator for the state of the SO can be easily derived and readily available to aid in the data evaluation process, then that will be tremendously helpful in streamlining the data evaluation process.

In this paper, a brief summary of the essence of the traditional rubric-based methodology is first discussed, followed by the comparison to the essence of the FCAR methodology. Thereafter, a novel approach is presented to demonstrate how a traditional rubric-based approach can be integrated with the FCAR assessment approach to allow computation of the SO performance index from roll-up data. The performance index is calculated based on the weighted average of relevant PIs for three different levels of courses. Ultimately, each SO is assessed to determine whether the performance meets expectations, exceeds expectations, or is below expectations. Customarily, even for the FCAR methodology, heuristic rules are used [5] to gauge results on how the SO performance is measured up for the final three expectations. In this paper, results of how the SO performance index can be used to address the overall attainment of the SO expectation are shown.

II GR Rubric-Based Assessment Model

The traditional rubric-based assessment model has been used widely by many universities in various formats. By and large the major contribution, in particular in the area of engineering accreditation, is attributed to Dr. Gloria Rogers’ work and workshops [9,14]. Although there are many major contributors to the rubric-based assessment, in tribute to Dr. Rogers’ significant contribution role, this paper will term the traditional rubric-based assessment model as the GR Assessment Model. Figure 1 illustrates the general flow of a typical GR assessment assurance process.
GR assessment model begins with aligning Program Education Objective (PEO) to the Mission of the University. Student Outcomes align with PEOs. Performance criteria or Performance Indicators (PI) are a set of criteria to be measured for the success of each SO. Education practices and strategies, assessment of the data collected and analysis of the evidence, and eventually evaluation of evidence to generate action items are adapted or developed based on the needs and circumstances of the program. The action items generated then feedback to the program to further investigate the necessary changes required to be made to PIs, strategies, assessment and/or evaluation processes. The education practices/strategies, assessment and evaluation processes are the main source of variation among universities.

But, in essence, the GR assessment model begins with the curriculum outcomes-mapping matrix, a sample of which is shown in Figure 2. Figure 2 shows a sample curriculum outcomes-mapping matrix. This matrix correlates all those courses that have learning components which contribute to the PIs and the corresponding SO being measured. In general, for the GR assessment model, courses are classified into introductory, reinforced and mastery levels. By doing so, it is clear from the curriculum-map matrix how courses are distributed among each PI being measured. The GR assessment model focuses on courses that are at the mastery level for assessment instead of looking at results from every course in the matrix to simplify the process of collecting vast amount of data for evaluation. The assessment results from data collected at the mastery level will generate sets of action items which feedback to the program for improvement. Even with data collected only at the mastery level, for instance at a large university, the amount of data collected will still be an issue for a timely evaluation. Figure 3 shows a commonly used process of how the GR assessment model is implemented\[^{[9,15]}\].
Since there are vast amounts of data collected even for the mastery level courses, a sample of data, 10% for example, are actually being assessed by an independent multi-rater team. The independent raters are in general selected to be individuals different from the faculty members.
who actually taught those courses to be evaluated. The intent is to give unbiased evaluations of the evidence collected. Furthermore, a set of rubrics is used for each PI so that the raters employ consistent criteria when an article from the same student is being evaluated. Action items generated from different raters are then summarized as a report and feedback to the program director or chair to take corrective actions for improvement. The program director or chair may then choose to disseminate or generate sub-action items for faculty to act on or to change the curriculum as a result. This multi-rater methodology is adopted by BlackBoard® [16].

Augmented with the above is the assessment plan. The assessment plan defines when, the frequency, and the number of SOs to be evaluated. This evaluation is of the corresponding SOs’ own cycle of assessment before the next accreditation.

The GR assessment model has the following characteristics:

- Since only mastery-level courses are being assessed, even without dedicated toolsets, the process can be achieved manually with commonly available tools like Words, Excel, etc. in a timely manner.
- Independent raters remove the involvement of faculty teaching the courses during the evaluation process.

The process is particularly time effective if the assessed results at the end meet the expectations, since laterally you could justify meeting an outcome by investigating evidence from one course at the mastery level. However, what if the assessed results indicate there are concerned areas that need to be improved. From the independent raters’ perspectives, they can only suggest areas to be looked at, but not precisely which course in the curriculum or rather which content in a course needs improvement in instructional delivery. In addition, the following are the challenges we face if the GR assessment model were to be implemented:

- Owing to limited resources, multi-raters are not a practical approach for our program. It is difficult to form independent raters who have the necessary subject-matter expertise to evaluate students’ work in core engineering subjects. For general education subjects or soft skills assessment, forming multi-raters team is feasible, but also not desirable for the following additional reason: faculty directly involving in teaching the subject matter are excluded during the evaluation process which also cause faculty member not to own the assessment/evaluation process right at the beginning.
- Without including courses at introductory or reinforced level in the assessment and evaluation processes, critical information on the root causes for issues which arise at the mastery level cannot be captured.
- It is difficult, if not impossible, even with the use of BlackBoard® to fully automate the GR assessment model for data collection, evaluation, and generating action items back to the appropriate courses for improvement. BlackBoard® is available to us. It also lacks the mechanism to give a full and ultimate evaluation of any specific SO that consists of many PIs. This is due to the difficulty to coordinate multi-raters teams for different PIs in order to provide summative evaluation on the corresponding SO. Most likely, the SO evaluation is performed separately after the summative results can be tabulated for the SO.
III FCAR assessment model

FCAR was first created by Dr. John Estell in 2001 [1]. The FCAR methodology has been evolving since then and implemented in various formats [3-6] in many universities. FCAR allows instructors to write assessment reports in a concise, standardized format conducive for use in both course and student outcomes assessment. The most recent FCAR methodology consists of the FCAR which is generated by faculty members at the end of the semester. The FCAR provides one or two pages of summative information related to the courses taught by each faculty member during that semester. The FCAR generally contain the following information:

- Course Description
- Course Outcomes
- Class Grade Distribution
- Course Outcomes Assessment
- Student Outcomes Assessment
- Reflection
- Proposed Action Items

The main idea is to capture the reflection and proposed action items for improvement of courses taught at the grass-roots by the responsible instructors. Hence, the assessment information is processed by the instructor who is most closely associated with the data, so that any observed difficulties or extenuating circumstances affecting performance can be readily documented as part of the FCAR. In general, instructors are asked to directly contribute to the assessment of student outcomes by collecting evidence such as scores recorded from assignments. These are then documented in the FCAR to support a set of predetermined ABET SO performance criteria for that course. The FCAR uses the performance vector, conceptually based on a performance assessment scoring rubric developed by Miller and Olds [17], to categorize aggregate student performance. Table 1 shows the performance vector termed as EAMU.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent (E)</td>
<td>Student applies knowledge with virtually no conceptual or procedural errors</td>
</tr>
<tr>
<td>Adequate (A)</td>
<td>Student applies knowledge with no significant conceptual errors and only minor procedural errors</td>
</tr>
<tr>
<td>Minimal (M)</td>
<td>Student applies knowledge with occasional conceptual and/or procedural errors</td>
</tr>
<tr>
<td>Unsatisfactory (U)</td>
<td>Student makes significant conceptual and/or procedural errors when applying knowledge</td>
</tr>
</tbody>
</table>

In general, the specification of the performance criteria in each category is up to the instructor. The performance criteria can be refined according to the relevance to the course. It is a choice for the instructor to use either rubrics or to specify scoring levels for each category, typically

In general, the specification of the performance criteria in each category is up to the instructor. The performance criteria can be refined according to the relevance to the course. It is a choice for the instructor to use either rubrics or to specify scoring levels for each category, typically
defined by 90% and above for Excellent, 75%-89% Adequate, 60%-74% Minimal, and below 60% Unsatisfactory.

In the FCAR assessment model, the FCARs from each course are further processed into a performance vector table (PVT) for each student outcome. Figure 4 shows a sample of a PVT for an SO. The data presented in each PVT is evaluated through the use of two heuristic rules.

![Figure 4: A performance vector table (PVT) for a SO](image)

The first heuristic rule is captured in Table 2 and is used to generate flags of appropriate color to aid in the identification of concern areas, if any. The second heuristic rule is given in Table 3. The second rule applies to the PVT that consists of courses belonging to the same matrix corresponding to each SO. This classification is used to examine the performance level of how each SO is finally met by the aggregate contribution from the selected set of courses.

### Table 2: First heuristic rule for color flags

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Flag</td>
<td>Any performance vector with an average below 3.3 (on a 5-point scale) AND a level of unsatisfactory performance that exceeds 10% in the U column.</td>
</tr>
<tr>
<td>Yellow Flag</td>
<td>Any performance vector with an average below 3.3 OR a level of unsatisfactory performance (U) that exceeds 10%</td>
</tr>
<tr>
<td>Green Flag</td>
<td>Any performance vector with an average that is at least greater than 4.6 and no indication of unsatisfactory performance (U)</td>
</tr>
<tr>
<td>No Flag (white)</td>
<td>Any performance vector that does not fall into one of the above categories</td>
</tr>
</tbody>
</table>
Table 3: Second heuristic rule for expectation classification

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below Expectation</td>
<td>When there are (1) any red or yellow flags raised in the final year of data, or (2) multiple red flags are raised prior to the final year of data, or (3) a majority of red and yellow flags are raised prior to the final year of data</td>
</tr>
<tr>
<td>Exceeding Expectation</td>
<td>When there are a majority of green flags raised</td>
</tr>
<tr>
<td>Meeting Expectation</td>
<td>The default condition when the application of the heuristic rules does not fall into either of the Below Expectation or Exceeding Expectation categories</td>
</tr>
</tbody>
</table>

An assessment process using FCAR methodology

The assessment process using FCAR methodology has been through several iterations in our department. For example, to streamline and maintain consistency in the selection of key assignments as supporting evidence, the systematic process of constructing a syllabus with well-written justification for identifiable key assignments as reported in [4] is used. The idea is to set in place the process of “plan-teach-assess” in every core course. A well-constructed syllabus is the planning stage. During teaching is when the course portfolio with supporting evidence is collected. Finally, the FCAR is the assessment stage of the course. Figure 5 illustrates the assessment process that involves each faculty not only at the course level, but also to call the meeting for the SO that he/she is responsible for based on the courses listed in the PVT.

Figure 5: SO assessment process using FCAR methodology
In this case, we adopted EvalTools® [18], which is the only online tool that facilitates the FCAR assessment model. EvalTools® is our primary toolset for data collection, tracking, and compilation of the FCAR data into PVT. The discussion and suggested action items in the review meeting are captured in EvalTools® as well. Since the time consuming portion of the assessment process is performed by EvalTools®, faculty members can focus more on the evaluation process. The evaluation process comprises review of the proposed action items from the FCAR and the preparation of a brief executive summary for the SO under review. As indicated in Figure 5, alumni survey and senior-exit survey are also used in the evaluation process. Even though alumni survey data is not required by the recent changes in ABET requirements, we keep it to gain insight and to gauge if we indeed meet our PEOs. EvalTools® tracks the survey data and generates the curriculum outcome-mapping matrix.

III Essentials of GR and FCAR assessment models

There are strengths and deficiencies in both methodologies. In particular, the following are observed:

- The GR assessment model classifies courses into three levels: introductory, reinforced, and mastery, but the assessment is only at the mastery level to give clear indication of final attainment of results and time effectiveness by looking only at a smaller set of data.
- The FCAR methodology can also be used to focus only on courses that are at the mastery level. In addition, assessment of course outcomes as well as the assessment of mapped student outcomes is possible. Course-outcomes assessment is not required for ABET accreditation purpose, but is required by the Middle-States accreditation process. Course outcomes assessment is also mandated by our university-wide assessment process. Hence, it makes sense to track the results for both course and student outcomes.
- The FCAR methodology involves instructors working on the FCAR in the assessment process right from the beginning. It is the primary reason we adopted the FCAR methodology for our assessment process.
- The FCAR review process on the PVT is also a multi-raters process, but not independent multi-raters process that is normally used in the GR assessment model.
- The FCAR methodology gives a more direct way of evaluating SO from PVT using heuristic rules.

The challenge now is how to integrate the strengths of the GR assessment model into the FCAR assessment model to give appropriate course level classification, and yet retain/enhance the existing FCAR assessment model to achieve the following goals:

1. Track each performance indicator (PI) assessment with the PVT comprising courses including introductory and/or reinforced in addition to mastery level. This is to help capture the root causes of concern if these should occur at the introductory or reinforced level.
2. Evaluate the SO from the roll-up data of its corresponding set of PIs, and yet be able to have an aggregate average for each PI and the aggregate average as the performance index for the SO that is consistent with the heuristic rules that were used in the FCAR assessment model.
IV FCAR assessment model with course-level classification

The ECE department has successfully obtained assistance from MAKTEAM Software that owns EvalTools® to develop our proposed features and requirements.

PI Definition
The department organized several sessions of discussion and examined what has been done in this area. Defining PIs that make sense is not an easy task. Each PI can be defined in a broad sense or in depth. There are merits and demerits for having PIs defined in breadth or in depth. If the PI is defined in depth, it naturally leads to more precise measurement of performances. However this is at the expense of time necessary to complete the evaluation. The added advantage of the PI defined in depth is that PIs can be further classified according to Bloom’s taxonomy to give very precise knowledge of which domains of learning are being assessed in the curriculum. With the PI defined in breadth, the major advantage is reduction in the number of PIs needed to be measured. The department has decided to have PIs defined in breadth and test out its effectiveness first. Table 4 shows a portion of the PIs defined. As indicated in Table 4, there are at least three PIs defined for each SO. The effort for the evaluation process has been expanded at least three fold.

Table 4: Snapshot of PIs defined

<table>
<thead>
<tr>
<th>SO1: Ability to apply knowledge of mathematics, science, and engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 1_1</td>
</tr>
<tr>
<td>PI 1_2</td>
</tr>
<tr>
<td>PI 1_3</td>
</tr>
<tr>
<td>PI 1_4</td>
</tr>
<tr>
<td>PI 1_5</td>
</tr>
<tr>
<td>PI 1_6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SO4: Ability to function on multi-disciplinary teams</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 4_1</td>
</tr>
<tr>
<td>PI 4_2</td>
</tr>
<tr>
<td>PI 4_3</td>
</tr>
<tr>
<td>PI 4_4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SO9: Recognition of the need for, and an ability to engage in life-long learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 9_1</td>
</tr>
<tr>
<td>PI 9_2</td>
</tr>
<tr>
<td>PI 9_3</td>
</tr>
</tbody>
</table>

Course-level classification
It should be re-iterated that course-level classification into introductory, reinforced, and mastery levels are based on the skill sets intended to be measured in the course, and not the academic year of that the course. For example, Electromagnetism is a junior-level course, but if it were used to gauge the performance indicator on “use principles of Newtonian and Maxwellian
physics to solve problems in computer and electrical engineering” as part of the PIs for the SO1: “ability to apply knowledge of mathematics, science and engineering”, this course is classified as “mastery” for this SO. Hence, the same course can be classified at different levels for different SOs intended.

Figure 6 shows an enhanced version of a PVT from roll-up data of the FCAR from each course and the result after review. As indicated in Figure 6, there are four courses which span introductory through reinforced to mastery and are included for assessing this PI_1_2. In general, the introductory and reinforced courses are not mandated to be included for each PI. It is the decision of the department to see what information needs to be derived from this set of data. Figure 6 also shows the weighting factor (WF) for different course levels. In this case, we choose WF of 1 for introductory, 2 for reinforced and 5 for mastery as indicated in the “red” box. How the EAMU average and the aggregate average of PI being computed is discussed next.

Figure 6: Enhanced PVT with review results

EAMU average
How the EAMU average is being computed is reported in [5]. It is repeated here for clarity. Originally, each performance vector (EAMU) carried a weight of 3 for E, 2 for A, 1 for M, and 0 for U; the average would have been between 0 and 3. With EvalTools®, however, the EAMU vector has been scaled from 0-3 to 0-5. As an example, let’s look at how EAMU vector for Electronics 1 in Figure 6 yields an average of 4.55 as indicated in the blue box. The EAMU is (28,2,2,1). The average is computed as:

\[
EAMU_{avg} = \frac{\sum E \times 3 + \sum A \times 2 + \sum M \times 1 + \sum U \times 0}{\sum E + \sum A + \sum M + \sum U} \times \frac{5}{3}
\]
\[ = \frac{28 \times 3 + 2 \times 2 + 2 \times 1 + 1 \times 0 \times 5}{28 + 2 + 2 + 1} \times 3 = 4.55 \]

For the Electronics 1 course, the color flag is white (no flag) according to the heuristic rule in Table 2.

**PI aggregate average (PI performance index)**

The PI average can be computed in two ways. One is in the regular sense of weighted average by taking each EAMU average weighted by its WF for the aggregate average as shown below:

\[
P_{\text{avg, regular}} = \frac{EAMU_{\text{avg1}} \times WF_1 + EAMU_{\text{avg2}} \times WF_2 + \cdots + EAMU_{\text{avgn}} \times WF_n}{WF_1 + WF_2 + \cdots + WF_n}
\]

\[ = \frac{2.41 \times 1 + 4.55 \times 2 + 4.77 \times 2 + 4.71 \times 5}{1 + 2 + 2 + 5} = 4.46 \]

By computing the weighted average in the regular sense, the focus is to assign more weight to the frequency of occurrence of courses at the same level. In this example, the PI average result will be weighted more on courses that are at reinforced level. Even without the mastery course included, the resulting PI average may still turn out to be good. This, in reality, is problematic, as the performance indicator shall be at least measured for the mastery level courses. Hence, for new programs that may not have sufficient courses at the mastery level to be measured, the regular weighted average gives a good sense of what the program achieves and emphasizes at the instant. However, for programs that have the full set of courses, the regular weighted average of the PI does not necessarily give an accurate indication of the performance achieved. This leads us to look at computing the average by considering different levels of courses in a comprehensive manner as follows:

\[
P_{\text{avg, comprehensive}} = \frac{EAMU_{\text{avg, introductory}} \times WF_{\text{introductory}} + EAMU_{\text{avg, reinforced}} \times WF_{\text{reinforced}} + EAMU_{\text{avg, mastery}} \times WF_{\text{mastery}}}{WF_{\text{introductory}} + WF_{\text{reinforced}} + WF_{\text{mastery}}}
\]

\[ = \frac{2.41 \times 1 + \text{avg}(4.55, 4.77) \times 2 + 4.71 \times 5}{1 + 2 + 2 + 5} = 4.41 \]

where \( WF_{\text{introductory}} = 0 \) if \( EAMU_{\text{avg, introductory}} = 0 \); similarly, \( WF_{\text{reinforced}} = 0 \) if \( WF_{\text{reinforced}} = 0 \).

Courses at the same level are grouped together for their aggregate contribution to the PI average. In this case, the EAMU averages of the two courses at reinforced level are averaged first to give the group average contribution to the overall PI average. The frequency of occurrence for the same level of courses is de-emphasized and replaced by the group contribution from the same
course levels. In addition, by following the same thinking for the EAMU color flag rules, the visual indication of color flags for PI aggregate average is determined by the following rule:

Table 5: Heuristic rule for PI color flags

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red Flag</td>
<td>PI aggregate average less than 3.0 (60%) OR (PI aggregate average less than 3.3 (66%) AND with more than 10% U)</td>
</tr>
<tr>
<td>Yellow Flag</td>
<td>(PI aggregate average greater than 3.3 (66%) AND with more than 10% U) OR (PI aggregate average less than 3.3 AND with less than 10% U)</td>
</tr>
<tr>
<td>Green Flag</td>
<td>PI aggregate average more than 4.6 (92%) AND with 0% U</td>
</tr>
<tr>
<td>No Flag (white)</td>
<td>PI aggregate average does not fall into one of the above categories</td>
</tr>
</tbody>
</table>

In this example, the percent U is 7.22%. The percent U is computed by the following formula:

\[
\%U = \frac{\text{sum of } \%U \text{ from each class}}{\text{total students}}
\]

Since the aggregate is 4.41 with 7.22% of U, the aggregate average has no flag (white) as indicated by the green highlighted box in Figure 6.

Based on the color flag for the aggregate average of PI, the result implies that PI meets expectation. If the color flag is yellow or red, it implies that the PI does not meet expectation (Below Expectation). Similarly, if it is green, it implies that the overall performance for PI exceeds expectation. Is this result consistent with the result obtained if heuristic rule of Table 3 applies to this example? From Table 3, if the rules apply, the end result of analysis will conclude that this example meets expectation and is consistent with the results implied by the aggregate average of PI. To this end, we achieve our goal by looking at the aggregate average of the PI or a PI performance index to give a quick visual result on the PVT that is consistent with the past practice of applying heuristic rules.

It should be made clear that if there are no courses at the introductory or reinforced level, the corresponding WF is not taken into account. However, mastery level courses are required for computing the aggregate average for PI. In this example, if the mastery level course were not included due to error, the end result is a low aggregate average for PI with red flag because both the introductory and reinforced WFs are intentionally set to lower values than that of the mastery WF. By doing so, it allows us to visually spot the source of the error and take corrective actions accordingly.

SO aggregate average (SO performance index)
Could a similar approach be applied to compute SO aggregate average and give good performance index for identifying which level of expectation the SO achieves? Figure 7 gives a consolidated view of all the roll-up data for the PIs reviewed for SO1 in spring 2015. In this example, PI_1_5 was not listed. This calls for an assessment plan update to lay out when and
how often this group of PIs needs to be reviewed. The final version of the consolidated PVT should include all roll-up data from all PIs. However, the computation approach presented here for SO aggregate average is still applicable.

Since the PIs are indicators, they should be of equal importance for measuring the performance of SO1, for example. Hence, PIs are given equal weight. As a result, the aggregate average for the SO should just be the average of all roll-up PIs involved as follows:

$$SO_{avg} = \frac{PI_{1,avg} + PI_{2,avg} + \cdots + PI_{n,avg}}{\text{total number of PIs}}$$

Figure 7: Roll-up data of PIs for SO1
In this example, the SO average turns out to be 4.0. In this example, it also turns out that all PIs have no color flag (white). In general, all PIs may have different color flags. To capture this information, the EAMU vector for SO is formed by capturing the total number of each category of E, A, M, and U that each PI belongs. Hence, in this example, the EAMU is (0,5,0,0) indicating that there are 5 PIs and all of them have no flags (white) which is in the “A” category. Please be reminded that E is green, A is white, M is yellow and U is red. The ultimate color flag for SO is based on the following rule:

Table 6: SO Color flag rules

<table>
<thead>
<tr>
<th>Color Flag</th>
<th>Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>SO aggregate average is less than 66% or 3.3/5.0 or 2/3</td>
</tr>
<tr>
<td>Green</td>
<td>SO aggregate average is more than or equal to 92% or 4.6/5.0 or 2.75/3.0</td>
</tr>
<tr>
<td>Yellow</td>
<td>(SO aggregate average is between 66% and 92%) and (sum of yellow and red &gt; sum of white and green)</td>
</tr>
<tr>
<td>White</td>
<td>If none of the above</td>
</tr>
</tbody>
</table>

Although, four color flags are used for SO, red and yellow flags imply SO is Below Expectation, white implies Meeting Expectation, and green implies Exceeding Expectation. In this example, the color flag is white for the SO as indicated in the green box in Figure 7. As shown in Figure 7, the concern areas are Circuit I and Computer Architecture. Altogether, the data suggests that the overall SO meets expectation which is consistent with what the SO aggregate average implies.

Figure 8 gives another consolidated view of all SOs with roll-up data from their own sets of PIs.
In this top-level view, the SO1 results are expanded to show the detailed results of its PIs. Let us take PI_1_2 highlighted in the red box as an example. Its PI aggregate average is 4.41 with 7.22% in U and 97 students total involved in 4 classes. Two classes have PIs in E (green) category, one in A (white) category and one in U (red) category. The PIs were reviewed on 2015-10-30. In fact, by looking at this table, there is still a lot of work ahead as SO3, SO6, and SO8 are red in color.

It should be emphasized that the enhancement made to our assessment/evaluation process did not change the existing data collection process for each faculty member. The only new step for the faculty is to map key assignments to PIs instead of directly mapping to SOs as has been done in the past. To this end, we have achieved our goal of integrating the merits of the GR assessment model into the FCAR assessment model. In so doing, we take course-level activities into account for computing the PI and SO performance index, thereby providing a clear indication of the attainment of PI and SO, respectively.

V Conclusions
In this paper, a novel approach is presented to demonstrate how a traditional rubric-based approach can be integrated with the FCAR assessment approach to allow computation of the SO performance index from roll-up data. The performance index is calculated based on weighted averaging of relevant PIs in three different levels of courses. By doing so, the enhanced approach achieves (1) capture of possible root causes of deficiency at the lower level courses when the symptom occurs at the higher level courses; (2) providing the mechanism and performance indices for PIs and SOs that clearly indicate their attainment and/or concern areas for improvement; (3) analytic results on the SO performance index indeed measured up against the heuristic rules previously used to gauge the attainment of SOs; (4) last but not least, the performance indices that facilitate further automation of the evaluation process.

Bibliography


[14]. Workshops information available at http://www.abet.org/workshops-and-evnts/


[16]. Information available at http://www.blackboard.com
