Assessment *in* Practice

Using Chocolate Chip Cookies to Develop a Rubric for Assessing Non-Academic Units

Patty Q. Flowers & Stephanie Kolitsch
University of Tennessee at Martin

April 2021

In 2015, the University of Tennessee at Martin was placed on probation by our accrediting body, SACSCOC, for our underdeveloped system of monitoring institutional effectiveness. UT Martin had just begun collecting assessment reports from non-academic (student affairs and academic support) units. The assessment cycle was still a foreign concept, and many staff members expressed emotions ranging from uncertainty to downright fear of repercussion when preparing assessment reports. After being placed on probation, we needed to collect three cycles of assessment reports prior to submitting our next monitoring report to SACSCOC, and we needed a quick way to provide feedback to these units so that they could improve their reporting with each cycle. We settled on a "yes/no" rubric that would indicate whether or not the information requested in the assessment reports (e.g., measurable goal, assessment tool, benchmark, use of data to inform decisions) was provided. If information was missing or was presented in a confusing way, we also provided comments as guidance for preparing the next assessment report. While not optimal, the yes/no rubric met our needs at the time, and UT Martin was removed from probation in 2016 (see Figure 1 on the next page).

Our next major report for SACSCOC, the Fifth Year Monitoring Report, was due in 2019. Since we had been on probation, we were also asked to provide a Follow-Up report to demonstrate that we were continuing to meet the standards addressed in our previous Monitoring Report. In the spirit of continuous improvement, we needed to raise our expectations for our assessment reports and develop a way to conduct a meta-analysis of the reports to provide an institution-wide perspective on the institutional effectiveness of our non-academic units. That meant we needed to develop a more sophisticated rubric for providing feedback to these units. To be effective and to get buy-in, we needed these units to participate in the development of the feedback rubric. However, most of the staff in these units had never used a rubric similar to the one we needed, much less developed one.

RUBRIC FOR EVALUATING NON-ACADEMIC UNIT PLANS

Unit _____

Date _____

DOES THE UNIT HAVE:	YES	NO	COMMENTS
Measurable outcomes?			
Appropriate assessments?			
A benchmark for each assessment?			
An established timeline for administering assessments?			
Identified person(s) responsible for administering assessments?			
An established timeline for evaluating and using the assessment data?			
Identified improvements or decisions based on the evaluation of the data?			

Figure 1. Original Yes/No Rubric

In the spirit of continuous improvement, we needed to raise our expectations for our assessment reports and provide an institution-wide perspective on the institutional effectiveness of our non-academic units. That meant we needed to develop a more sophisticated rubric for providing feedback to these units.

Each January, before classes begin, we hold an Assessment Workshop. Usually the Assessment Workshop is divided into two sessions—one focused on assessment in academic units and the other focused on assessment in non-academic units. In 2018, we built our non-academic unit workshop around the concept of rubrics. The workshop was divided into three parts:

1. The basics of rubrics (What is a rubric? What does one look like? What are all of the parts of the rubric?);
2. Creating a "practice" rubric; and
3. Creating a feedback rubric for assessment reports.

In the first part of the workshop, we described a basic table-style rubric, with the rows describing the attributes or dimensions we valued or wanted to measure and the columns describing the level of "performance" or "scale" of the work we are evaluating for each of those attributes or dimensions. We discussed the idea of "growth" or "sophistication" within the scale levels for each attribute, and we gave some concrete examples of academic rubrics to illustrate such "growth."

The second part of the workshop was key in moving us toward crafting a usable, effective feedback rubric. We needed the participants to practice creating a rubric for evaluating something, and we needed that "something" to be familiar to all participants without triggering extreme emotions towards or against the topic. We chose to create a rubric for evaluating chocolate chip cookies. To prepare the participants, we provided a template for our chocolate chip cookie rubric and we provided packages of chocolate chip cookies for inspiration. We chose seven different brands of cookies and encouraged participants to sample from multiple packages. Participants were sitting at round tables with no more than 8 people per table so that natural "groups" were formed.

Chocolate Chip Cookie Rubric			
In this column, list the qualities you will use to "judge" your cookie.	*Use these columns to describe the different "levels" of quality. Be as descriptive as possible—you want to communicate to someone who is unfamiliar with your expectations what each level means.*		
	This column describes an "unacceptable" version of the attribute.	*This column describes the "acceptable" but not necessarily exemplary version of the attribute.*	*This column describes the absolute best version of the attribute.*
Attributes / Criteria	Level 1: _____	Level 2: _____	Level 3: _____

Figure 2. Chocolate Chip Cookie Rubric

Our cookie rubric template had three scale levels, with level 1 representing the lowest and level 3 representing the highest. Participants were first asked to identify words describing each scale level. We expected scale levels to be described as something like "bad – acceptable – best." After giving the groups a few minutes to describe their scale levels, we had multiple variations on this theme; some examples included "inedible – acceptable – irresistible" and "blech – OK – yummy."

We then moved to the attributes/dimensions. We asked the groups to identify characteristics they could use to evaluate the cookies. We provided five rows on the template and gave participants a few minutes to identify their characteristics. We then asked groups to share their characteristics. A few were predictable: taste, size, freshness, and number of chips were popular attributes. However, some groups gave some surprising responses: texture, visual appeal, cost, nutritional value, non-essential ingredients, and packaging were among the characteristics being evaluated.

Finally, we asked the groups to describe the differences in the levels for each characteristic. We suggested they describe the middle level first, then describe the lowest level and the highest level. Participants quickly realized that some characteristics were easier to describe than others. For example, "number of chips" and "size" were fairly easy to describe using the usual measurements

(although there were some differences among the groups concerning an acceptable number of chips or an acceptable diameter of cookie); "visual appeal" usually referred to a scale of "undercooked/chewy – not chewy but not hard/crispy – overcooked/too crispy." Characteristics such as "taste" and "freshness" were more challenging when trying to reach a consensus among the group members. Once the groups had completed their rubrics, we asked participants to share various attributes and descriptions of the levels with the whole group in order to illustrate the growth from "worst" to "best" and to point out potential flaws when the differences in levels were not well-defined.

The final portion of the workshop entailed drafting rubrics that could be used to evaluate assessment reports. We still wanted the participants to work in groups, but we wanted the groups to be "mixed" so that their final products would not use unit-specific language or criteria. As part of the planning for the workshop, we placed numbers (1-8) in the upper left corner of each chocolate chip cookie handout. Participants were instructed to find a table with their number (we had several tables with each number), and those at that table became their new "group."

Each group was provided poster paper, markers, and a blank copy of the annual unit-level assessment reporting template. The members of the group worked together to create a rubric they thought could be used to evaluate an annual assessment report. We limited them to three scale levels in their rubric, but they could have as many attributes/characteristics as they thought were necessary. In defining the continuum from lowest to highest, we encouraged them to think of the middle level as meaning "acceptable" and define that first. The highest level would be an "ideal"—what would an exceptional assessment report look like? The lowest level would be essentially a "beginner" level. We also reminded them that the descriptions at each level should be "unit neutral" and not use language that was unit-specific. Finally, we reminded the participants that the rubric should measure the quality of the assessment report, not the quality of the measurements or the data the units were using, and that the language in the rubric should be designed to provide feedback on the report itself. In other words, the rubric should be evaluating whether the unit was "telling its assessment story" in a clear manner.

Once groups had completed their draft rubrics, we had them participate in a modified gallery walk. The groups' posters were displayed around the room (see Figures 3 and 4), and participants were asked to walk around the room and read each poster. Participants were also given small stacks of sticky notes and were asked to write comments on the notes and stick them to the posters. In particular, we were interested in phrasing or in ideas that resonated with participants, with questions the participants may have had about the rubrics, or in suggestions for improving the wording on a poster. At the end of the gallery walk, we had one last group discussion about what they had observed on the posters.

The rubric should evaluate whether the unit was "telling its assessment story" in a clear manner and the quality of the assessment report, not the quality of the measurements or the data the units were using.

	Unacceptable	Acceptable	Exemplary
Measurable Outcomes	No identifiable outcomes or inappropriate outcomes	Identified Appropriate Outcomes	Identified outcomes reflecting high expectations and rigor
Appropriate Assessment	Data tool does not effectively measure No assessment	Assessment tool is appropriate and produces meaningful data	Assessment tool is appropriate and produces meaningful data to be used to affect change
Identifiable Benchmarks	Benchmarks are absent or are unrealistic	Identified realistic & achievable benchmarks	Ambitious benchmarks
Process (who, when, how, where)	One or more components of process not identified	Process criteria are identified	Process components are identified & documented
Data Collection & Use	Insufficient data Did not produce meaningful data	Meaningful data is produced + utilized	Data is meaningful + analysis has been applied to make positive program outcome changes

Figure 3. Sample Group Poster

Rating	Unacceptable (1)	Adequate (2)	Exceptional
Goals/ Outcomes	Undefined	Defined - Not Clear	Defined - Clear Method
Findings	Not reported/ unclear	Reported - Vague	Clearly reported-understandable
Changes (as needed)	No action taken	Partial Changes Made	Changes fully implemented
Data/ Assessments	One or no data/ assessments	Two to Three data Points	3+ data points that provide greater triangulation
Process	None - Undefined	Defined, but gaps exist	Defined, clear process
Timelines	No established Timeline	Established/ Unclear	Clearly Established

Figure 4. Sample Group Poster

After the workshop, we collected all of the posters and the accompanying sticky notes. We transcribed each poster, along with any sticky notes, and compiled all of the information into one document. We sorted the attributes in the document so that similar attributes appeared together and compared the language in the descriptors for similar attributes. Many of the same attributes appeared on almost every poster, and language describing the three levels for an attribute tended to be parallel across multiple posters. Perhaps more surprising was the preciseness of language prevalent across posters—in most cases, each level was clearly delineated from the other two, and very little "vague" language was noted.

CRITERIA	LEVEL 1 Unacceptable (6) Unsatisfactory Needs Improvement (post it: Like "level 1: Name") Needs Work Inadequate Rework (post it: 3-Like "Rework")	LEVEL 2 Acceptable (7) Satisfactory Meets Expectations Adequate (2)	LEVEL 3 Outstanding Exceptional (5) Exceeds Expectations Exemplary (3) Excellent
Measurable Outcomes	Subjective-difficult to measure. Population not clearly identified	Objective measures in place with population defined	• Clear, realistic timeline for outcomes • what is measured is clearly defined (post it: not sure this makes this outstanding)
Outcomes	Unmeasurable	Measurable Identified Clear	Scalable based on improvement
Measurable outcome	Not specific or measurable	Clearly stated and defined	SMART met
Measurable Outcome	Vague language; lacks specificity (post it: like wording)	clearly defined	clearly defined; specific; detailed
Measurable Outcomes	no identifiable outcomes or inappropriate outcomes	Identified appropriate outcomes	Identified outcomes reflecting high expectations and rigor (post it: thumbs up!; like)
Measurable	not acceptable/measurable	measurable (post it: Keith Carver approves this message)	quantifiable (post it: {must make sure what is being measured is clear "what is a freshman?"})
Goals/outcomes	undefined	defined-not clear (post it: should still be clear if adequate)	defined-clear method

Figure 5. Sample Transcripts of Posters

Once we had completed our analysis of the posters, we created a draft feedback rubric using the similarities we had identified in the posters. With few exceptions, the language in the draft consisted of direct quotes from the posters; little wordsmithing was needed. The draft feedback rubric and the document with the poster transcriptions was then shared with the workshop attendees for feedback. We received a handful of comments that contained a few suggestions for changes in wording, but we did not receive any negative feedback regarding the drafted rubric in general. There was also a question of whether the various characteristics should be "weighted," and feedback for that particular question was split. One staff member succinctly summarized the reason we ultimately decided not to weight the characteristics: "If you don't have acceptable goals/ outcomes, assessment tools, and benchmarks, then it is impossible to have acceptable processes, analysis, and data-driven decisions. I would argue that weighting these items is redundant and unnecessary."

RUBRIC FOR EVALUATING NON-ACADEMIC UNIT ASSESSMENT REPORTS

CRITERIA	Level 1: UNACCEPTABLE	Level 2: ACCEPTABLE	Level 3: EXCEPTIONAL	SCORE
Goals/Outcomes	• Not measurable or difficult to measure • Vague; lacks specificity or focus • Inappropriate or irrelevant for unit	• Measurable • Clearly defined • Appropriate for unit • Aligned to the unit's mission	• Measurable using multiple strategies • Clearly defined and specific • Reflect high expectations • Future-oriented • Clearly aligned to the strategic plan • Where applicable, reflect peer, industry, and/or national standards	
Assessment Tool(s)	• Not identified or not relevant • Does not measure effectively the goal or outcome • Not related to the mission/purpose of the unit	• Clearly described • Appropriate measure of the goal or outcome • Produces meaningful data	• Assessment tool is evidence-based, valid, and reliable • Provide(s) long-term comparable data • Where appropriate, industry/national assessment tool is used • Where appropriate, multiple assessments including both direct and indirect measures	
Benchmarks	• Not stated or not clearly defined • Not aligned to assessment tool • Unrealistic or trivial	• Clearly stated • Aligned to assessment tool • Appropriate for unit's mission/purpose • Where appropriate, reflect some knowledge of peer, industry, and/or national standards	• Reflect high expectations • Clearly advance the mission of the unit and/or institution • Where appropriate, aligned to peer, industry, national, or other external benchmarks • Where appropriate, reflect long-term aspirations	
Processes	• No responsible person identified • No timeline identified • Standard operating procedure (workbook) for collecting/analyzing data is not described or poorly defined	• Responsible person identified • Appropriate timeline clearly stated • Standard operating procedure (workbook) for assessment is clearly described	• Back-up assessment personnel identified • Sustainable process for collecting/storing data is in place • Results are shared with stakeholders both within and outside the unit as appropriate	
Analysis of Data	• Data summary is missing or incomplete • Basic data analysis missing or incomplete • Questionable methodology for analyzing data • Results are unclear, disorganized, or not supported by data	• Data summary is clear, concise, and informative • Basic data analysis is complete • Acceptable methodology used for analyzing data • Results are clear, organized, and supported by data • Data analysis identifies strengths and/or weaknesses related to the attainment of the goal/outcome	• Data analysis is sophisticated and reflects implications for the unit's stated goals/outcomes • Where appropriate, analysis includes comparisons to external benchmarks • Where appropriate, longitudinal analysis of data is included • Where appropriate, data from multiple measures are analyzed and compared to identify trends	
Use of Data to Inform Decisions	• Decisions based on data are missing or incomplete • Decisions provided but not relevant to the goal or outcome • Description of decisions is confusing, vague, or trivial	• Decisions are clearly stated and appropriate • Decisions reflect results of data analysis and address identified strengths and/or weaknesses as appropriate • Decisions support attainment of goal/outcome • Decisions are clearly communicated to all stakeholders • When appropriate, "closing the loop" is clearly described through examples or actions	• Decisions are future-oriented and focused on improvement(s) • Decisions describe innovations informed by unit data and (where appropriate) peer, national, industry, and/or research-based data, programs, or trends • When appropriate, "closing the loop" is clearly described through examples/actions and demonstrated through before/after comparison data	
Overall Quality of Report	• Report is confusing, vague, or nonsensical • Report is incomplete • Report is late	• Report is clearly understandable to the layperson • Report is grammatically correct with no spelling errors • All required components are included • Report is submitted on or before the due date	• Report is well-written and engaging • Report clearly reflects the mission and/or purpose of the unit • Report clearly supports the strategic plan of the institution • Where appropriate, report clearly reflects input from multiple stakeholders	
TOTAL SCORE				
COMMENTS:				

Figure 6. Final Rubric

We used the rubric for the first time in December 2018. When evaluating assessment reports using the rubric, we highlighted the portions of the rubric that applied to the report in question and then applied a score to each attribute. Providing units with this style of feedback resulted in a marked increase in responses from the non-academic units; almost all of the questions we received from these units focused on how to improve their assessment reporting, which also allowed us to talk to them about how they could improve their assessment processes. Several division leaders noted that this rubric made it easier to identify where improvements were needed. Two non-academic units set up two-hour meetings with us to discuss their results, identify ways to improve, and discuss future goals for assessment.

CRITERIA	Level 1: UNACCEPTABLE	Level 2: ACCEPTABLE	Level 3: EXCEPTIONAL	SCORE
Benchmarks	• Not stated or not clearly defined • Not aligned to assessment tool • Unrealistic or trivial	• Clearly stated • Aligned to assessment tool • Appropriate for unit's mission/purpose • Where appropriate, reflect some knowledge of peer, industry, and/or national standards	• Reflect high expectations • Clearly advance the mission of the unit and/or institution • Where appropriate, aligned to peer, industry, national, or other external benchmarks • Where appropriate, reflect long-term aspirations	2
Processes	• No responsible person identified • No timeline identified • Standard operating procedure (workbook) for collecting/analyzing data is not described or poorly defined	• Responsible person identified • Appropriate timeline clearly stated • Standard operating procedure (workbook) for assessment is clearly described	• Back-up assessment personnel identified • Sustainable process for collecting/storing data is in place • Results are shared with stakeholders both within and outside the unit as appropriate	2
Analysis of Data	• Data summary is missing or incomplete • Basic data analysis missing or incomplete • Questionable methodology for analyzing data • Results are unclear, disorganized, or not supported by data Plans for data collection and analysis are implied but not fully described.	• Data summary is clear, concise, and informative • Basic data analysis is complete • Acceptable methodology used for analyzing data • Results are clear, organized, and supported by data • Data analysis identifies strengths and/or weaknesses related to the attainment of the goal/outcome	• Data analysis is sophisticated and reflects implications for the unit's stated goals/outcomes • Where appropriate, analysis includes comparisons to external benchmarks • Where appropriate, longitudinal analysis of data is included • Where appropriate, data from multiple measures are analyzed and compared to identify trends	1

Figure 7. Sample Highlighted Rubric (Partial)

In addition to providing more specific feedback to units, using this feedback rubric has allowed us to conduct a "meta-analysis" on the data to gauge the institutional effectiveness of the university as a whole. The figure below provides the results from our first administration of the rubric. The categories highlighted in red show our three lowest-scoring areas of the rubric (see Figure 8 on the next page).

Providing this information to units also allowed units to compare their individual scores to the university as a whole (which then prompted additional questions from various non-academic units). We disseminated the meta-analysis to the entire institution through our monthly Assessment Newsletter, which also allowed us to recognize those units who scored in the top 10% on the rubric. We are now conducting a meta-analysis each year and hope that, by the time we are required to submit our next report to SACSCOC, we can show growth in each area, especially in those original low-scoring areas.

Category	All		Administrative Support Units		Student Affairs	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Goals/Outcomes	2.466	0.414	2.500	0.302	2.357	0.639
Assessment Tools	2.276	0.501	2.318	0.466	2.143	0.580
Benchmarks	2.276	0.518	2.341	0.437	2.071	0.678
Processes	1.897	0.515	1.955	0.498	1.714	0.525
Analysis of Data	1.862	0.628	1.886	0.673	1.786	0.452
Use of Data to Inform Decisions	1.707	0.689	1.659	0.713	1.857	0.580
Overall Quality of Report	2.328	0.478	2.318	0.414	2.357	0.639
Total Score	14.810	2.881	14.977	2.583	14.286	3.614

Figure 8. Table of Results

CONCLUSION

We have now adapted the feedback rubric for use with academic reports on student learning outcome assessment. The adapted rubric has been approved by department chairs and deans, and we will begin using the feedback rubric for academic units' program assessments in Fall 2020. All of our templates, rubrics, etc. are publicly available on our assessment website: www.utm.edu/ assessment; look for the "Assessment Forms" link. Once we have piloted the adapted rubric with program assessment reports, we will consider tweaking the rubric to use with our general education assessment reporting process.

Please Cite As:

About NILOA

The National Institute for Learning Outcomes Assessment (NILOA) was established in December 2008, and is co-located at the University of Illinois and Indiana University.

The NILOA website contains free assessment resources and can be found at http://www.learningoutcomesassessment.org.

The NILOA research team has scanned institutional websites, surveyed chief academic officers, and commissioned a series of occasional papers.
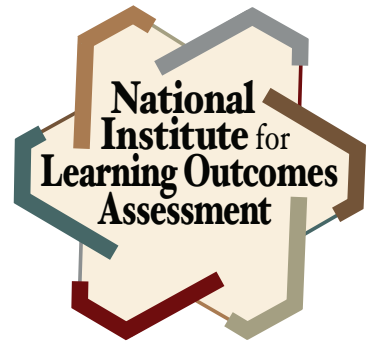
Sign up to receive our monthly NILOA Newsletter and stay up to date with our research and publications.