Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 17, Number 4, January 2012 ISSN 1531-7714

Transforming Rubrics Using Factor Analysis

Ed Baryla, Gary Shelley, and William Trainor
East Tennessee State University

Student learning and program effectiveness is often assessed using rubrics. While much time and effort may go into their creation, it is equally important to assess how effective and efficient the rubrics actually are in terms of measuring competencies over a number of criteria. This study demonstrates the use of common factor analysis to identify the number of criteria actually measured by an analytic or listing rubric. This information can help designers revise their rubrics so that each criterion measured corresponds to one unique competency making the rubric more effective and efficient.

Assessing student learning and program effectiveness is necessary for any program that wants to know whether they are achieving learning objectives. In addition, many accrediting bodies require assessment of learning outcomes. One of the primary ways schools assess their programs and student learning is with rubrics. The most frequently used rubrics are holistic, analytic, and listing.

An evaluator using a holistic rubric assigns a single score to the student's work. Although easy to create, the drawback of holistic rubrics is that they do not provide a sufficient level of detailed analysis for student improvement. Analytic rubrics differ from holistic rubrics in that they allow the evaluator to provide student performance scores for a number of different evaluation criteria. Riddle and Smith (2008) also point out that a rubric can take the form of a checklist, where there is no performance measure. The student either does or does not complete the designated task

Because analytic and listing rubrics are more detailed, they are generally more useful than holistic rubrics. However, this increased value has a cost in terms of determining exactly which criteria should be included. Some rubrics have very few criteria while others are quite extensive. Popham (1997) points out that lengthy rubrics often "gather dust". However, regardless of the number of criteria used, there is little information in the literature concerning how to assess the rubric ex post beyond measuring reliability across assessors, (Stemler, 2004; Stevens and Levi, 2005). To the extent more criteria are used, the designer must keep in mind whether measuring additional criteria is worth the cost in terms of time, effort, and quality of data. More importantly, are each of the criteria measured actually unique and do they provide additional information in terms of program effectiveness?

The purpose of this research paper is, for the first time in the literature, to offer evidence regarding the efficiency of the task-oriented lengthy checklist rubric. Specifically, this study applies common factor analysis to a unique data set to demonstrate how to collapse the information compiled in a task-oriented checklist rubric into a more efficient set of performance criteria with a continuum scoring system. In a way, this study provides empirical evidence on Popham's (1997) flaws one and three regarding "Task specific evaluative criteria" and "Dysfunctional detail."

Background

Popham (1997) recognizes the difficulties that may befall an institution adopting an improperly structured rubric. Institutions measuring skills or learning outcomes can sometimes fall into a trap whereby a poorly designed task-specific assessment rubric diverts focus away from the critical components that determine whether students have mastered the desired skill or learning outcome. Overly lengthy and excessively detailed rubrics exacerbate this problem by discouraging the evaluator from conducting a fair assessment, assuming that the evaluator uses the rubric at all. He suggests that an effective rubric should contain three to five evaluative criteria. "In rubrics, less is more; each criterion would represent a key attribute of the skill being assessed." Wolf and Stevens (2007) reiterate the idea that less than six criteria are best.

The task then is to identify the list of important criteria. A study by Dunbar, Brooks, and Kubica-Miller (2006) analyzes eight competencies dealing with student oral communications skills. The authors find that the students underperformed in five of the eight competencies, leading to suggestions for program improvements. In this case, it appears key criteria are correctly identified and assessed, but this is not always the case.

Tierney and Simon (2004) offer a critical evaluation of the problems that continue to plague rubrics. They cite difficulties with describing the quality levels for each of the performance criteria as a main drawback with rubrics. The authors suggest that rubric construction should focus on three main explicit areas: performance criteria, attributes, and progressive levels of evaluation for attributes. A vital point they make is that "the performance criteria represent broad learning targets, rather than features of a particular task." This seems to echo Popham's (1997) work.

It appears that one consistent theme within the literature is that a properly developed rubric provides a measure of the extent to which a student has accomplished an evaluative or performance criterion that is critical to the mastery of a particular learning competency or outcome. Further, a properly developed rubric, featuring a scoring system that is on a continuum, is an effective guide to improve instruction and student learning, (Tierney and Simon, 2004).

Alternatively, a poorly developed rubric focuses on "present/absent or right/wrong" across an array of tasks without definitive focus on whether a student has accomplished the performance criteria that are critical to skill mastery. Indeed, such rubrics often lack rudimentary consideration of performance criteria and focus instead on task accomplishment. While most authors agree that this type of lengthy performance-list rubric is not optimal, there is no empirical evidence on whether or not the task-oriented rubric offers beneficial information beyond task measurement. This study addresses this issue by applying factor analysis to a lengthy analytic/checklist type rubric.

Factor Analysis

Factor analysis is a statistical method used to reduce a large number of variables to a smaller more manageable number by identifying the number of unique underlying criteria.

For example, when tabulating the results from a particular rubric, if the value of the first criterion usually equals the value of the second criterion, then there is no gain in additional significant information. It is possible that the redundancy in values means there is only one true underlying component. In essence, factor analysis is a quantitative way to determine if multiple criteria are measuring the same thing.

Alternatively, principal component analysis (PCA) could be applied which is a special case of factor analysis. PCA also estimates how many factors are measured and usually gives very similar results to factor analysis. However, as Costello and Osborne, (2005) point out, PCA is just a data reduction method, computed without regard to any underlying structure. Specifically, PCA calculates the components based on total variance as opposed to evaluating just the shared variance. Applying this procedure to a lengthy listing rubric may identify the number of underlying competencies measured, but falls short in revealing what these competencies may be.

Thus, this study relies on factor analysis to identify the underlying competency to which an individual criterion belongs. Furthermore, since a priori there is a belief that underlying latent variables are causing the correlation among certain groups of criteria, factor analysis is more appropriate (see SAS Usage Note: Stat-53, 1995). This study finds that PCA analysis simply confirms the results obtained from factor analysis.

Applying factor analysis to a lengthy, task-oriented analytic or checklist rubric is relatively straightforward. For example, suppose that five factors explain most of the variation across 25 criteria in a rubric. In this case, it is efficient to edit the rubric so that it contains only five performance criteria, each corresponding to a unique learning competency. Concentrating on five critical performance criteria may allow the assessor to accurately measure whether the student has mastered each competency as opposed to measuring performance on 25 different tasks. This is especially relevant during time-constrained oral presentations.

As noted by Popham (1997) on evaluative criteria, "They should be the most instructionally relevant component of the rubric. They should guide the teacher in designing lessons because it is the students' mastery of the evaluative criteria that ultimately will lead to skill mastery…lengthy rubrics probably can be reduced to succinct but far more useful versions for classroom instruction…that can still capture the key evaluative criteria needed to judge student responses," (pgs 73-74).

To demonstrate the use of factor analysis in assessing rubrics, this study applies the procedure to an oral rubric used to assess student learning at the authors' university. There are 49 criteria used in this rubric so it provides an excellent case study on the use of factor analysis to increase the efficiency and effectiveness of a rubric composed of a large number of criteria.

Data and Results

The data consists of more than 1,400 individual student rubrics measured by a variety of faculty members over a multi-year period. The oral communications rubric has 49 criteria with 4 general headings (Content, Analysis, Organization, and Presentation) respectively composed of 6, 5, 17, and 21 criteria. Each student can receive a 0, 1, or 2 for each criteria based on: "does not meet", "meets", or "exceeds expectations". Thus, a priori, at least four underlying components are expected. To determine how many significant factors there are, both Kaiser's (1960) eigenvalue one criterion and the scree test are used. Following Costello and Osborne (2005), this study uses maximum likelihood with the promax oblique rotation to estimate significant factors. Costello and Osborne suggest that this type of factor analysis is optimal, at least in regards to social science issues. Other rotations showed similar results.

Table 1 shows the first ten eigenvalues. Based on Kaiser's criteria, there are at most, seven significant factors.

Table 1: First ten eigenvalues

Number	Eigen Value	Proportion Explained	Cumulative Proportion Explained
1	23.10539	0.4715	0.4715
2	4.112864	0.0839	0.5555
3	2.281430	0.0466	0.6020
4	1.617441	0.0330	0.6350
5	1.450197	0.0296	0.6646
6	1.435633	0.0293	0.6939
7	1.194475	0.0244	0.7183
8	0.986392	0.0201	0.7384
9	0.862505	0.0176	0.7560
10	0.819914	0.0167	0.7728

The scree plot of these eigenvalues, shown in Figure 1, suggests there are only four critical factors. However, if one assumes that there are only four critical factors, then the rotated factor pattern used to determine which criteria links to each factor does not separate the criteria by the major headings delineated in the rubric. In fact, a few of the criteria link to others in a way that does not make intuitive sense. The use of seven criteria does not suffer from this issue so the following analysis assumes there are seven factors.

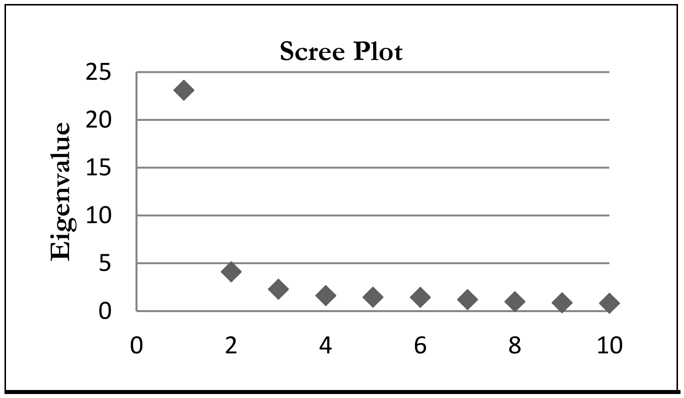

Figure 1: Scree plot of eigenvalues

The overriding conclusion is that this rubric measures too many oral communications performance criteria. For all 49 criteria, there are at most seven significant factors. In essence, using more than seven criteria is not generating any additional significant information. It is interesting to note that in the Dunbar, Brooks, and Kubicka-Miller (2006) study of student development of oral communication skills, the authors use only eight performance competencies in their rubric.

The Appendix shows the rotated factor pattern used to map the 49 criteria included in the oral rubric to the seven underlying factors. The number of criteria mapping to each factor ranges from two to fifteen. The entire analysis section (criteria 7-11) as well as three minor subheadings within the organization section (criteria 15-21), load onto Factor 1. The average loading factor for Factor 1 is 0.78 and ranges from 0.60 to 0.85. The factor analysis literature considers any loading greater than 0.5 a strongly loaded item. The fact that no question associated with this factor has a cross loading on any other factor greater than 0.18 strongly suggests that all of these criteria are related to one underlying component.

Factor 2 links criteria 29-43, which includes most of the presentation section. Factor 3 combines several organization sections together. Only two to four criteria each map to factors 4-7. The next step is to provide a description of these seven underlying factors.

As an example of this process, Table 2 shows the actual criteria that load onto Factor 1. The data suggest there is little variation in a student's assessment across these 12 criteria. They are redundant and do not supply the assessor with additional significant information. It is likely that the similarity in results across all 12 criteria is due to the fact they are assessing the same underlying latent component. Even if the developer is trying to measure more than one core competency with these criteria, the results of factor analysis suggest that the rubric needs significant modification.

Table 2: Original 12 criteria

Analyzes quality/relevance of data/source.
Is able to identify key information/data from sources to include in presentation.
Builds an adequate argument.
Student identifies audience/groups.
Student appropriately targets audience
Relevant – inclusion of key information/data.
Correct – adequate presentation of current and correct information/data.
Makes connections between ideas/facts/data to construct an effective argument.
Develops theme.
Demonstrates appropriate/logical sequence of ideas/facts/data.
Clarity of ideas/argument.
Derives logical conclusions based on information/data gathered.

A cursory examination suggests that these criteria load on the same factor because they all deal with adequately presenting a core idea. Comparing the criteria to the Association of American Colleges and Universities' (AACU) "Oral Communication Value Rubric," these criteria generally relate to the competency labeled "Central Message". Thus, a revised rubric should collapse these 12 criteria into a single performance criterion. Table 3 provides an example of such a revision.

Data quality issues also may play a role in these results. It is possible that assessors are simply checking off entire sections of the rubric based on their own idea of how the student is meeting a single underlying core competency rather than separately analyzing

Table 3: Revision of oral analysis section based on factor analysis results*.

Performance Criterion	Presentation of the central message and quality of analysis
Does not meet	Central message is not explicitly stated or inferred. Key data and its relevance are not analyzed.
Meets standards	Central message is stated or inferred. Some key data and its relevance are analyzed.
Exceeds	Central message is precisely stated, appropriately repeated, and strongly supported by analysis of all key data.

*Criteria combined with AACU's Oral Value Rubric's central message competency.

performance for each listed criterion. Enfolding these criteria into a smaller number of performance criteria that are directly linked to specific learning competencies allow the assessors to more accurately determine the student's achievement level. In addition, it is possible to assess additional important competencies in the revised rubric because the new version is more efficient and less burdensome to complete. With fewer criteria to evaluate, the quality of the data will likely improve. In this way, the use of factor analysis can help create a rubric that is more efficient, effective, and informative.

Conclusion

Assessing student performance is a difficult process. It is now common across the educational spectrum to employ rubrics to help with this process. A neglected important step in this process is to assess what the rubrics are actually measuring. Each criterion used for assessing student performance should be unique and valuable in terms of feedback to both the student and instructor for improved learning outcomes.

This study shows that factor analysis can determine the number of unique underlying competencies actually measured by a rubric. A case study examines an oral rubric containing 49 criteria. Findings suggest this rubric measures at most seven underlying factors. This information can help reduce an initially redundant instrument into a simpler one, while further increasing the value of the rubric so that each performance criterion actually corresponds to a unique learning competency.

The desired result is to attain higher quality assessment information while simultaneously reducing the redundancy of the rubric to make the assessment task less burdensome on the instructors. The importance of this result cannot be understated. Learning institutions currently using lengthy checklist rubrics that resemble multiple-choice exams should consider conducting factor analysis as part of the process of continually improving their assessment instruments. The ease of using factor analysis allows anyone to use this procedure to help evaluate an assessment program. Factor analysis can help assessors design rubrics that are more efficient. The improved rubrics then will provide higher quality information concerning achievement of learning objectives while reducing the overall burden on faculty conducting the assessment.

References

Association of American Colleges and Universities, Oral Communication Value Rubric, retrieved August, 15, 2011, from http://www.aacu.org/value/rubrics/pdf/OralCommunication.pdf.

Costello, A. B. and Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation,* 10, (7), pp. 1-9. retrieved Jan 11, 2012 from http://pareonline.net/getvn.asp?v=10&n=7.

Dunbar, N.E., Brooks, C.F. and Kubica-Miller T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31, no.2, pp. 115-128

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, pp. 141-151.

Popham, W. J. (1997). What's wrong and what's right with rubrics. *Educational Leadership*, Vol. 55 (2) 72-75

Riddle, J. and Smith, M. (2008). Developing and using rubrics in quantitative business courses. *The Coastal Business Journal*, 7, (1), pp. 82-95.

SAS Library, (1995). Usage Note: Stat-53, retrieved, Oct. 21, 2011 from http://www.ats.ucla.edu/stat/sas/library/factor_ut.htm.

Stemler, Steven E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9 (4), retrieved Jan 11, 2012 from http://pareonline.net/getvn.asp?v=9&n=4.

Stevens, D. and Levi, A. (2005). *Introduction to rubrics.* Sterling, VA, Stylus Publishing.

Tierney, R. and Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research and Evaluation*, (9) 2, retrieved Jan 11, 2012 from http://pareonline.net/getvn.asp?v=9&n=2.

Wolf, K. and Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, 7, (1), pp. 3-14.

Appendix

Rotated Factor Pattern

Criteria	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
1	0.07	0.16	0.09	**0.51***	0.01	0.08	0.01
2	0.05	0.21	0.08	**0.51***	-0.06	0.07	0.17
3	0.04	0.03	0.14	**0.84***	0.02	-0.03	0.04
4	0.05	-0.04	0.10	**0.85***	0.10	-0.01	0.03
5	0.02	-0.01	0.02	0.16	0.00	0.00	**0.82***
6	0.06	0.03	-0.03	0.11	-0.01	-0.04	**0.88***
7	**0.82***	0.08	-0.16	0.09	-0.12	0.02	0.10
8	**0.80***	0.03	-0.17	0.17	-0.05	0.06	0.04
9	**0.85***	0.06	-0.19	0.08	-0.10	0.03	0.10
10	**0.81***	-0.03	-0.10	0.15	0.10	-0.01	0.01
11	**0.78***	-0.04	-0.06	0.18	0.09	0.01	-0.03
12	-0.05	0.16	**0.42***	0.09	0.01	0.09	-0.04
13	0.04	0.18	**0.40***	0.15	0.04	0.08	-0.07
14	-0.02	0.19	**0.60***	0.04	-0.21	0.12	0.05
15	**0.61***	0.04	0.11	0.03	-0.05	0.14	0.01
16	**0.60***	0.01	0.12	0.00	-0.03	0.11	-0.01
17	**0.82***	-0.01	0.15	-0.08	0.07	-0.09	0.00
18	**0.85***	0.00	0.12	-0.07	0.07	-0.04	-0.01
19	**0.83***	0.03	0.16	-0.13	0.10	-0.08	0.02
20	**0.79***	-0.02	0.17	-0.11	0.05	-0.01	-0.01
21	**0.80***	-0.06	0.08	0.00	0.03	0.04	-0.09
22	0.44	0.13	**0.49***	-0.07	-0.11	0.00	-0.06
23	0.05	0.22	**0.63***	0.18	-0.01	-0.16	-0.06
24	-0.02	0.17	**0.70***	0.14	-0.08	-0.09	-0.06
25	0.07	0.15	**0.27***	0.15	0.19	0.05	0.08
26	0.00	0.03	**0.32***	-0.02	0.18	0.15	0.18
27	0.12	0.16	**0.42***	-0.08	0.23	-0.03	0.15
28	0.09	0.24	**0.27***	-0.02	0.31	-0.10	0.17
29	-0.07	**0.50***	0.19	-0.07	0.26	0.02	0.05
30	-0.02	**0.49***	0.04	-0.01	0.19	0.06	0.07
31	-0.04	**0.47***	0.10	0.02	0.19	0.08	0.05
32	-0.04	**0.47***	0.20	-0.03	0.13	0.09	0.04
33	0.00	**0.55***	0.19	0.01	0.01	0.04	0.01
34	-0.13	**0.46***	0.27	0.08	-0.02	0.10	0.00
35	0.08	**0.59***	0.14	0.04	0.15	-0.03	-0.03
36	0.04	**0.62***	0.07	0.14	0.22	-0.07	-0.06
37	0.08	**0.66***	0.02	0.04	0.18	-0.06	-0.05
38	0.06	**0.56***	0.19	-0.05	-0.05	0.06	0.01
39	0.04	**0.92***	-0.10	0.03	0.03	-0.07	-0.04
40	0.04	**0.91***	-0.07	0.04	-0.04	-0.03	0.00
41	0.06	**0.86***	0.04	-0.06	-0.14	-0.04	0.08
42	0.05	**0.74***	0.08	-0.03	0.03	-0.04	0.00
43	0.34	**0.48***	0.04	-0.08	-0.12	0.22	-0.13
44	0.23	0.04	-0.06	0.05	0.03	**0.71***	-0.01
45	0.09	-0.02	0.04	0.04	-0.01	**0.87***	-0.02
46	0.20	0.04	0.12	-0.14	0.28	**0.45***	0.02
47	0.02	0.18	-0.05	0.10	**0.82***	0.00	-0.05
48	0.05	0.25	-0.10	-0.01	**0.77***	0.01	0.01
49	-0.01	0.24	-0.09	0.05	**0.81***	0.05	-0.05

*Maps to the factor in that column.

Citation:

Baryla, Ed, Shelley, Gary and William Trainor (2012). Transforming Rubrics Using Factor Analysis. *Practical Assessment, Research & Evaluation*, 17(4). Available online: http://pareonline.net/getvn.asp?v=17&n=4

Authors:

Ed Baryla
Associate Professor of Finance
East Tennessee State University
Department of Economics and Finance
PO Box 70686
Johnson City, TN 37614

Gary Shelley
Assistant Professor of Economics
East Tennessee State University
Department of Economics and Finance
PO Box 70686
Johnson City, TN 37614

William Trainor, Corresponding Author
Associate Professor of Finance
East Tennessee State University
Department of Economics and Finance
PO Box 70686
Johnson City, TN 37614
TRAINOR [at] mail.etsu.edu