Research Note

Checking Your Analytic Performance Rubrics for a Halo Effect

Beth Sheppard, University of Oregon

Abstract

In this research note, the author checks for correlations between different dimensions in an analytic rubric used for scoring discussion performance. Highly correlated dimensions can be cause for concern that the different aspects of performance are not well defined or not adequately observed. The author's analysis showed some weak to moderate correlations, leading to adjustments in how the rubric will be applied. Meanwhile, the author describes how to check for correlations between rubric dimensions and suggests this method as a successful activity for teachers and assessment committees.

Key words: *performance rubrics, halo effect, assessment*

Background

The halo effect is defined by Blum and Naylor (1968; cited in Darby, 2007) as the "tendency to let our assessment of an individual on one trait influence our evaluation of that person on other specific traits" (p. 47). In other words, when a student's excellent organization in an essay influences us to think their word choices were also strong, or when a student's poor pronunciation in a discussion gives us the impression that they also had poor grammar, our assessment has been affected by a halo effect. The halo effect is a common source of error in performance ratings (Rogers, 2005, p. 52); in the ESL context, this effect may appear when we use an analytic rubric to assess speaking or writing performance.

Of course, students who perform well on one dimension of a rubric may also do well on other dimensions without the influence of an erroneous halo effect. However, consistent similarities between scores on different dimensions may also indicate that too many dimensions were being assessed, or that there was too little time for the teacher to adequately distinguish between the different dimensions. If several of the dimensions in an analytic rubric fail to accurately measure different aspects of student performance, then giving scores for these dimensions has wasted teacher time and possibly

misinformed students. For these reasons, I think it is worth investing a little time to check for halo effects in our analytic scoring rubrics.

I became concerned about the halo effect in my own assessment after completing a research study on perceptions of international student speech (Sheppard, Elliott, & Baese-Berk, 2017) in which raters' assessments of different dimensions of student speech (i.e., various aspects of pronunciation, grammatical and lexical accuracy, fluency) were strongly inter-correlated, suggesting a halo effect. This led me to wonder if my classroom assessments might be similarly affected. In this research note, I will describe how I tested a set of rubrics for the halo effect and recommend that you try the same procedure on your own assessments.

Investigation

Having taught the same course for several years, my files included completed rubrics for many classes. I was able to collect scores using the same rubric for 10 academic terms. A few rubrics were missing from the 10 class sets, but none were deliberately excluded. I had 122 completed scores. The assignment was a project in which upper intermediate students in their last class before exiting the Intensive English Program to enter the university led their peers in a discussion on a semi-academic topic. The dimensions scored can be seen in Figure 1.

Figure 1 - Discussion leading rubric

Expectations	✓	Comments	Score
INTRODUCTION			
• Tells why the topic is important and interesting • Students can easily answer the first question			/10
COMPREHENSIBILITY			
• Pronunciation • Grammar • Word choice • Strategies			/10
GROUP MANAGEMENT			
Balanced Participation and Inclusion • Include everyone • Politely redirect participants who talk too much or off topic • Actively engage participants who haven't spoken much			/10
Interactive Discussion • Ask follow-up questions • Summarize/paraphrase contributions • Ask participants to respond to each other • Base later questions on participant answers			/10
TIMING & ORGANIZATION			
• Finish after 15-20 minutes • Spend the right amount of time on each Q • Close the discussion with a summary and/or conclusion			/10
Additional Comments		Total: _____ /50	

There are a variety of methods for testing for a halo effect. One is simply to examine the correlations between the dimensions of an assessment (Rogers, 2005, pp. 54-55). Thus, I used Excel to calculate Pearson's correlation coefficients for the five dimensions of my rubric, finding the degree to which scores for each dimension correlated with scores for each other dimension. The results are presented in Table 1.

Table 1 - Pearson's correlation coefficients for scores on rubric dimensions (ref. fig. 1)

	2. Comprehensibility	3. Part. & Inclusion	4. Interactive Disc.	5. Timing & Org.
1. Introd.	0.344*	-0.016	0.222*	0.072
2. Compre.		-0.017	0.248*	-0.008
3. Inclusion			0.023	0.075
4. Interact.				0.170*

*=statistically sig (p<0.05)

None of the correlations were very strong, but four were statistically significant. The strongest correlation was between the ratings for Comprehensibility and Introduction. On the basis of this correlation, I need to make sure not to consider the comprehensibility of the speaker when I assess the quality of their introductory statement. The other three significant correlations all involved the dimension Interactive Discussion. This indicates that I should have a closer look at the description of this dimension to make sure it does not overlap with skills assessed in the other dimensions. I should also reconsider my own mental model of the "Interactive Discussion" dimension to make sure I have a clear understanding of what I am looking for.

While the significant correlations between dimensions directed me to areas of concern, overall it appears that halo effects did not play a strong role in my scoring of this assessment. Different rules of thumb can indicate that correlation coefficients between 0.20 and 0.39 are weak or moderate. In either case, the correlations leading to concerns about halo effects in previous studies were much stronger, ranging from 0.30 to 0.96 in my own study (Sheppard, Elliott, & Baese-Berk, 2017) and from 0.59-0.84 in Darby's

(2005) dissertation on halo effects. Therefore, I concluded that my use of this rubric was generally acceptable, especially given the adjustments arising from this exploration.

Recommendation

It is not difficult to check the intercorrelation of scores given for different dimensions of an analytic rubric. If you use the same rubrics every academic term, collect them over several terms. You could also collect rubrics from several classes in one term. Technically, you can calculate correlations with any number of rubrics, but I would recommend collecting at least 50 rubrics in order to feel confident in your results.

Open a spreadsheet and create a column for the source of each rubric (term, class, teacher, etc.), a column for a student number (you can call them student 1, 2, 3, etc.), and a column for each dimension in your rubric. The first two columns are just to help you check your work if needed. Then, enter the scores for each dimension into the appropriate column. You can see an example in Figure 2.

Figure 2: Example Excel sheet

	A	B	C	D	E	F	G
	course	student	intro score	compre scor	inclusion sco	interaction s	timing score
1	course	student	intro score	compre scor	inclusion sco	interaction s	timing score
2	12.2	11	8.5	8	9	9	9.5
3	12.2	12	8.5	7.5	8	8	7.5
4	12.2	13	9	8.5	8	9	9.5
5	12.2	14	8	9	9.5	8.5	8
6	12.2	15	10	8	8	9	9
7	12.2	16	8	8	10	7	9
8	12.2	17	8	9	9	9	8
9	12.2	18	8	10	10	8	9
10	12.3	19	8	8	9	7	9
11	12.3	20	8	7	9	7	9
12	12.3	21	8	8	9	9	8

The next step is to create a blank correlations table for your dimensions, like the one in Table 1, above. For each pair of dimensions, you can calculate the Pearson's correlation using the CORREL function in Excel. Go to a blank cell of your spreadsheet and type "=CORREL(array1,array2)," where array1 is a column (scores for one dimension of your rubric) and array2 is another. Excel will fill in the arrays for you if you highlight them. For example, to find the correlation coefficient for Introduction and Comprehensibility, I had:

=CORREL(C2:C123,D2:D123)

This resulted in a correlation coefficient of 0.34377241.

Keep careful track of which two dimensions you are comparing, and enter the result from the CORREL function into the appropriate box in your correlations table. Unfortunately, this Excel function does not calculate statistical significance of the results, but you can calculate it yourself using Pearson's coefficient and your N (the number of rubrics you collected). You can find instructions for this calculation online, for example at https://mariherigstad.wordpress.com/2016/07/05/p-values-for-correlations-in-excel/

I hope that this research note has encouraged you to take a closer look at the dimensions you use in analytic rubrics to analyze ESL performance. After learning about halo effects, I felt a lack of confidence in my own analytic rubric scores; checking for correlations between dimensions in a set of scores set my mind at ease and also helped me refine my approach to the scoring dimensions I had been using. I believe this is a valuable exercise for individual teachers, and checking for halo effects is particularly recommended for departmental assessment committees.

References

Blum, M. L., & Naylor, J. C. (1968). *Industrial psychology: Its theoretical and social foundations*. New York, NY: Harper and Row.

Darby, J. A. (2005). *An investigation of social and other factors which influence evaluations of educational courses* (Unpublished doctoral dissertation). Loughborough, U.K.: Loughborough University. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/2810

Darby, J. A. (2007). Are course evaluations subject to a halo effect? *Research in Education, 77*, 46-55.

Herigstad, M. (2016). P-values for correlations in Excel [Blog]. Retrieved from https://mariherigstad.wordpress.com/2016/07/05/p-values-for-correlations-in-excel/

Rogers, E. P. (2005). *Scale construction and halo effect in secondary student ratings of teacher performance* (Unpublished doctoral dissertation). Provo, UT: Brigham Young University. Retrieved from https://scholarsarchive.byu.edu/etd/560

Sheppard, B., Elliott, N., & Baese-Berk, M. (2017). Comprehensibility and intelligibility of international student speech: Perceptions of IEP instructors and content faculty. *Journal of English for Academic Purposes, 26*, 42-51. Retrieved from https://doi.org/10.1016/j.jeap.2017.01.006

Author

Beth Sheppard is a senior instructor at the University of Oregon. She earned her Bachelor's Degree in interdisciplinary studies from UC Berkeley, and her MA in linguistics from UO. Beth is in Mexico this year, training English teachers with a Fulbright grant. She has also taught German and Chinuk Wawa.