Psychometrics vs Academic Politics

The Impact of Administrators' Beliefs on the Design and Use of Student Rating Forms.

Raoul A. Arreola, Ph.D.
Professor Emeritus, The University of Tennessee Health Sciences Center
Paper presented at the 2010 Annual Meeting of the American Educational Research Association
May 2, 2010, Denver, Colorado

Abstract:
Academic administrators often control the development and use of various forms used in a faculty evaluation system. This paper provides a brief description of some of the beliefs that academic administrators hold that can often result in politically motivated changes to the design, development and use of student ratings that conflict with, or violate, sound psychometric principles. Also cited are two case studies that provide examples of the kinds of decisions that may be made that either weaken the reliability and validity of a student rating form or purposely hide certain information.

Research on Student Rating
Research in faculty evaluation has, for the main, focused on the design, development, and use of student rating forms. With over 80 years of literature in the field, the theory and practice of the development and use of such forms, and the data they produce, has been well documented. The psychometric principles for designing student rating forms, as well as other types of instruments intended to measure human attitudes, beliefs, perceptions, etc., have been well established. Unfortunately, however, in actual practice the decisions concerning the design and use of such forms are often made by academic administrators with little or no knowledge in psychometrics. It is the rare vice president, provost, or dean whose primary area of academic expertise is the field of psychological measurement and statistics. As a consequence, despite professional guidance and advice, these administrators may make decisions that ultimately result in the design and use of a form that may be less valid and reliable than it might have been.

Administrator Beliefs
Despite the extant literature in the field, certain common beliefs persist among academic administrators concerning the design and use of student rating forms. Lawrence Aleamoni, in his excellent 1999 article "Student Rating Myths versus research facts: 1924 – 1998" provides a detailed examination of some of the myths underlying these common beliefs. An example of these 'beliefs' that tend to cause a number of problems include:

- *Student rating forms only measure the instructor's ability to 'entertain' the students.*

- *A student rating form is just a questionnaire that needs to have a few critical questions.*

- *There is no need to have a special form printed up. You can use standard optical scanning sheets for student rating forms. Have the students simply have the students mark the appropriate "A, B, C, D, E" or "1, 2, 3, 4, 5" bubble on the optical scanning sheet.*

- *When you make up a form the response scale should have more, rather than fewer, response choices so that you can get a bigger 'spread' among the ratings.*

- *In interpreting student rating form data all you really need to look at is each item's "average" value. If more detail is needed look at the percentage responding to each choice.*

- *It is important to have an open 'Comments' section on the back of the form so that you can gather data that you might not have thought of. One or two strong comments on the back of a form can (and should) be used to override the item response data from the other students.*

- *The administrator should see all the student rating results before the instructor – and should modify them if comments on the back of the form warrant it.*

Case Study 1

A project was undertaken to develop a comprehensive faculty evaluation system for a large, multi-campus university. Part of the project involved the design and development of a new student rating form. The administrators of the institution wanted a form with a five-point scale in which the response choices were "1, 2, 3, 4, 5." No definitions for this scale were indicated. The administrators also wanted an open-ended comment area on the back of the form that would enable to students to provide any statements they wished concerning the instructor.

Design of Trial Version 1 of the Student Rating Form
Although the administrators wanted a 5-point scale, it was agreed, as simply a first trial, to usea four-point scale in the development of the draft of Version 1 of the form. The scale used for the pilot study of the form was: *Strongly Agree, Agree, Disagree, Strongly Disagree.*   The items for the form were selected from the catalog of items in Developing a Comprehensive Faculty Evaluation System 3/e (2007, R.A. Arreola, Jossey-Bass) and were intended to measure four factors: *Instructional Design, Instructional Delivery, Instructional Assessment,* and *Course Management.*

Also, as simply a trial, the open-ended comment section on the back of the form included the following three questions:
- *What aspects of this course did you find MOST effective in helping you to learn? Be specific:*
- *What aspects of this course did you find LEAST effective in helping you to learn? Be specific:*
- *If you could change anything about this course to make it MORE effective in helping you to learn what would you do? Be specific:*

Analysis of Trial Version 1 of the Student Rating Form
The form was administered to approximately 2500 students representing a broad spectrum of courses from all of the campuses in the institution. **Table 1** shows the structure resulting from the factor analysis. As can be seen, the first version of the form exhibited a factor structure that indicated that the four component areas (instructional design, instructional delivery, instructional assessment, and course management) were being measured. Specifically in Version 1 the:

- Instructional Design factor is defined by items 1, 2, 3, & 5
- Instructional Delivery factor is defined by items 6, 7, 8, 9, & 10
- Instructional Assessment factor is defined by items 11, 12, 13, 14, 15, 18, & 19
- Course Management factor is defined by items 16, 17, & 20

Table 1: Rotated Factor Matrix of Version 1 of the student rating form showing four distinct factors corresponding to the Instructional Design (F3), Instructional Delivery (F1), Instructional Assessment (F2), and Course Management (F4).

	Item Statement	F1	F2	F3	F4
			Rotated Factor Pattern		
1	The course was well organized	.50	.30	**.58**	.12
2	The course material appeared to be presented in logical content units.	.49	.34	**.58**	.10
3	I clearly understood the course objectives.	.45	.39	**.57**	.09
4	The assigned readings helped me to learn course material.	.04	.17	**.70**	.29
5	I knew what was expected of me in the course.	.38	.21	**.45**	.14
6	The instructor's communication style was easy to understand.	**.69**	.29	.35	.13
7	The instructor was enthusiastic in presenting the course material.	**.77**	.22	.09	.12
8	The instructor appeared to have a thorough knowledge of the subject.	**.78**	.21	.09	.13
9	The instructor's explanations were clear.	**.71**	.31	.32	.14
10	The instructor was receptive to students' questions.	**.73**	.31	.14	.19
11	The course objectives were reflected in the exams, homework, and other graded assignments.	.28	**.64**	.30	.14
12	My scores on the exams, homework, and other graded assignments accurately reflected how much I learned.	.13	**.77**	.20	.08
13	I clearly understood the grading system.	.27	**.75**	.18	.18
14	The exams, homework, and other graded assignments were fair.	.34	**.73**	.22	.12
15	The instructor provided meaningful feedback that helped in my learning.	.51	**.52**	.24	.27
16	Materials for in-class teaching activities were available when needed.	.18	.16	.22	**.77**
17	Materials for out-of-class learning activities were available when needed.	.05	.11	.20	**.84**
18	My scored exams, homework and other graded assignments were returned in a timely manner.	.28	**.62**	.07	.27
19	The instructor taught material that matched or supported the course objectives.	.47	**.55**	.22	.25
20	The instructor was available to students at times other than class time.	.31	.30	-.02	**.56**

Administrative Review of Trial Version 1 of the Student Rating Form

The administrators reviewed the results of the analysis of Version 1. There was some concern about the wording of certain items and alternate wordings were proposed. These concerns revolved around the belief among some of the administrators that the reading level was too high for their students and that they would not be familiar with certain terms. Thus the following changes were made in the items:

- Item 3 was changed from <u>I clearly understood the course objectives</u> to *I clearly understood what I was supposed to learn in this course.*
- Item 6 was changed from <u>The Instructor's communication style was easy to understand</u> to *The instructor presented a variety of activities that engaged me in learning.*
- Item 11 was changed from <u>The course objectives were reflected in the exams, homework, and other graded assignments</u> to *The exams, homework, and other graded assignments measured what we were supposed to learn in this course.*
- Item 12 was changed from <u>My scores on the exams, homework, and other graded assignments accurately reflected how much I learned</u> to *My scores on the exams, homework, and other graded assignments accurately reflected how much I actually learned.*
- Item 18 was changed from <u>My scored exams, homework and other graded assignments were returned in a timely manner</u> to *The instructor consistently started the class on time.*
- Item 19 was changed from <u>The instructor taught material that matched or supported the course objectives</u> to *The instructor consistently made good use of the entire class period.*

These changes resulted in Version 2 of the student rating form that was then administered to a total of approximately 2400 students. **Table 2** shows the results of the factor analysis resulting from this administration of the form.

Table 2: Rotated Factor Matrix of **Version 2** of the student rating form showing four distinct factors corresponding to the Instructional Design (F1), Instructional Delivery (F3), Instructional Assessment (F2) and Course Management (F4).

		Rotated Factor Pattern			
	Item Statement	**F1**	**F2**	**F3**	**F4**
1	The course was well organized	**.59**	.33	.32	.25
2	The course material appeared to be presented in logical content units.	**.61**	.35	.31	.20
3	*I clearly understood what I was supposed to learn in this course.*	**.65**	.36	.24	.24
4	The assigned readings helped me to learn course material.	.34	.23	.08	.32
5	I knew what was expected of me in the course.	**.66**	.37	.25	.28
6	*The instructor presented a variety of activities that engaged me in learning.*	**.49**	.18	.30	.31
7	The instructor was enthusiastic in presenting the course material.	.39	.21	**.55**	.17
8	The instructor appeared to have a thorough knowledge of the subject.	.30	.20	**.62**	.17
9	The instructor's explanations were clear.	**.55**	.28	**.48**	.20
10	The instructor was receptive to students' questions.	.39	.27	**.54**	.23
11	*The exams, homework, and other graded assignments measured what we were supposed to learn in this course.*	.38	**.59**	.29	.19
12	*My scores on the exams, homework, and other graded assignments accurately reflected how much I actually learned.*	.36	**.63**	.18	.18
13	I clearly understood the grading system.	.19	**.53**	.32	.28
14	The exams, homework, and other graded assignments were fair.	.33	**.67**	.29	.20
15	The instructor provided meaningful feedback that helped in my learning.	.31	**.57**	.31	.26
16	Materials for in-class teaching activities were available when needed.	.21	.22	.24	**.61**
17	Materials for out-of-class learning activities were available when needed.	.16	.16	.15	**.66**
18	*The instructor consistently started the class on time.*	.10	.33	**.55**	.30
19	*The instructor consistently made good use of the entire class period.*	.20	.33	**.57**	.30
20	The instructor was available to students at times other than class time.	.20	.11	.24	**.41**

Although the first version (Version 1) had a somewhat cleaner and stronger factor structure, the political decision was made to retain the wording used in the second version. Some administrators felt strongly about the students' ability to understand the wording used in the first version, despite the evidence of the analyses.

The administrators still wished to have a 5-point scale, and preferred to use "1, 2, 3, 4, 5" rather than the strongly agree-strongly disagree scale. It was recommended that if a 5-point scale was used that it be the following scale: (1) *Strongly Agree,* (2) *Agree,* (3) *Neither Agree/Nor Disagree,* (4) *Disagree,* (5) *Strongly Disagree.* At this point in the project the administrators turned over the development of the final version form to a faculty committee that was charged with using the 5-point scale the administration preferred.

Case Study 2

A health professions college commissioned development a new student rating form for a highly specialized clinical instruction environment. Since no standardized form, or specific items, existed for this type of clinical instruction setting the development of the items proceeded with the following steps:

1. All the students in the college were asked to list words or short phrases that they would use to describe or characterize an *Excellent* clinical instructor.
2. Similarly, all the students were asked to list words or phrases they would use to describe or characterize a *Poor* clinical instructor.
3. All the faculty were asked to develop the same two lists.
4. The student and faculty lists were examined to determine the words and phrases used by *both* groups.
5. Items were developed from those words and phrases used by *both* groups to develop items that represented both excellent and poor performance. This resulted in approximately 80 items.
6. The 80 items were placed into various versions of a student rating form and administered to students over a two-year period.
7. The resulting data was factor analyzed and resulted in a final form that measured 5 different factors.

The final form measured the following factors:

- *Student-Instructor Relationship*
- *Course Value*
- *Instructor Organization*
- *Teaching Method*
- *Clinical Professionalism*

Administrator Actions
The administrators reviewed the form and decided to omit the Clinical Professionalism factor from the final form. Although the Clinical Professional factor was solid and well-measured, the items that constituted the factor exposed certain instructor behaviors that would be politically damaging to the college if seen by the public.

The Clinical Professionalism factor addressed and issue that the students felt strongly about – inappropriate behavior around the live patient being used in the clinical instruction environment. The items that composed this factor included:
- *"The instructor avoids touching the patients inappropriately."*
- *"The instructor maintains a professional demeanor."*
- *"The instructor avoids using inappropriate language."*

In developing the items for the form it became apparent that there appeared to be a problem with some instructors who either became angry, frustrated, or otherwise upset during the clinical treatment/instructional process so that they would curse, call students names, and, sometimes, throw things across the room. Other times the instructor might take advantage of the 'position' of the live patient and engage in behavior that resulted in unwarranted, and unwanted, physical touching. Although these events were not common, the apparently occurred with sufficient frequency that both the faculty and the students mentioned them in their lists of words and phrases they used to characterize a poor clinical instructor.

The administrators' decision was, first, to omit those items (and the factor they measured) from the form. However, before the form was to be fully implemented, it was determined that the use of the form should be put on 'hold' for political reasons. It was never used.

References:
Aleamoni, L.M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13(2), 163-156.*
Arreola, R.A. (2007). Developing a Comprehensive Faculty Evaluation System 3/e (Jossey-Bass)