

Validity, like Beauty is.....

Raoul A. Arreola, The University of Tennessee Health Science Center, Memphis

Introduction

The question before us is “Are there any valid faculty evaluation data?” The traditional, research-based answer to that question is “Yes.” Even though decades of research have demonstrated that tools used in faculty evaluation systems, especially student rating forms, can be designed so as to be both valid and reliable, the question persists. So... why does the question as to whether faculty evaluation data is valid still reverberate throughout higher education after so much research? My experience with hundreds of colleges and universities facing the task of revising or building their faculty evaluation systems has led me to conclude that there are three main reasons:

1. Many of the tools used in faculty evaluation systems are ‘home made’ and are thus of dubious validity and reliability;
2. Most academic administrators are not conversant with the finer points of psychometrics; and
3. Higher education has yet to establish a universally accepted definition as to the characteristics and skills necessary for teaching excellence.

These three conditions lead to a cascading set of circumstances that profoundly affect the general professoriate’s perception, and willingness to accept, the possibility that any faculty evaluation data is, or can be, valid. A brief examination of these conditions reveals the underlying problem.

Homemade Faculty Evaluation Tools

Faculty evaluation tools have, by and large, been dominated by the use of some form of student rating form. Student rating forms have been, and continue today, to be the one main common element of all faculty evaluation systems. In many cases student ratings constitute the only systematically gathered data used in a faculty evaluation system. Unfortunately, the great bulk of student rating forms in use across higher education in America today are ‘home-made’. That is, they have been constructed by committees comprised of various combinations of faculty, academic administrators, and students. Owing to the lack of psychometric expertise or rigor in constructing these forms, they are of dubious (and often undetermined) validity and reliability. This common situation has, over the decades, resulted in a rich and voluminous storehouse of anecdotes and stories leading to a number of ‘myths’ including the ever-popular myths that student ratings are just a popularity contest and that faculty can ‘buy’ good ratings by giving easy grades. Unfortunately, these and many other such ‘myths’ likely have a basis in fact since they may possibly be true for poorly constructed forms. The anecdotes that have produced these and other ‘myths’ about faculty evaluation tools are so common, so voluminous, and so widespread throughout the culture of higher education that they have taken on an aura of ‘common knowledge.’ Thus, in the minds of the professoriate and academic administrators this ‘common knowledge’ is so pervasive that it far overshadows the ‘truth’ concerning student ratings and other faculty evaluation tools buried in the pages of psychometric journals. Which brings us to the next point.

What the heck is psychometrics?

It is unfortunate but true that the majority of academic administrators are unfamiliar with the finer points of psychometrics. Deans and Vice Presidents for Academic Affairs are usually the ones who are faced with the task of gathering some form of evaluative information concerning their faculty’s performance for the purpose of making promotion, tenure, continuation, or similar personnel decisions. Of the many hundreds of academic administrators with which I have had the occasion to interact on the issue of faculty evaluation, I could number on one hand the ones that have been sufficiently conversant with psychometrics to understand the subtle differences between, say, content and construct validity. These individuals, biologists, musicians,

historians, physicists, physicians, nurses, pharmacists, etc., not only are generally unfamiliar with the finer points of psychometrics but rarely, if ever, read articles in such publications as *Journal of Educational Psychology*, or *Review of Educational Research*. A personal case in point:

When I first joined the University of Tennessee Health Science Center in 1983 as chairman of the Department of Education I reported directly to the Vice Chancellor for Academic Affairs – a microbiologist of some note. At our first meeting he asked me in what area I had received my doctorate. I replied “Educational Psychology – specializing in psychometrics”. His response was, first, incredulity, then contempt, and then he said “Well that must be some kind of phony degree cause I’ve never heard of that.”

This anecdote may seem to represent an extreme case, but I have encountered some form of that response among any number of academic administrators. From the perspective of many of the people in academe responsible for making significant decisions concerning the design, structure, and format of faculty evaluation systems, the entire field of research endeavor that speaks to the issues of psychological measurement is, at best, a little-known area and, at worst, a ‘phony’ or illegitimate area of study. Thus, when the issue of the validity of faculty evaluation data is raised the definition or conception of validity used is much more likely to be that of colloquial usage rather than technical precision. It is useful to look at these two definitions. Below are the dictionary definition of “valid” and the technical definition of ‘validity’ from a psychometric perspective:

Main Entry: val·id

Pronunciation: 'va-l&d

Function: adjective

Etymology: Middle French or Medieval Latin; Middle French valide, from Medieval Latin validus, from Latin, strong, from valere

1. having legal efficacy or force; especially: executed with the proper legal authority and formalities <a valid contract>
2. well-grounded or justifiable: being at once relevant and meaningful <a valid theory> b: logically correct <a valid argument> <valid inference>
3. appropriate to the end in view: EFFECTIVE <every craft has its own valid methods>
4. of a taxon: conforming to accepted principles of sound biological classification

- va·lid·i·ty /v&-li-d&-tE, va-/ noun

- val·id·ly /'va-l&d-IE/ adverb

VALID implies being supported by objective truth or generally accepted authority <a valid reason for being absent> <a valid marriage>

[REFERENCE: MIRRIAM-WEBSTER ONLINE DICTIONARY]

Validity The effectiveness of the test in representing, describing or predicting the attribute that the user is interested in.

Content validity refers to the faithfulness with which the test represents or reproduces an area of knowledge.

Construct validity refers to the accuracy with which the test describes an individual in terms of some psychological trait or construct.

Criterion-related validity, or *predictive validity* refers to the accuracy with which the test scores make it possible to predict some criterion variable of educational, job, or life performance.

[REFERENCE: Thorndike, R.L. & Hagen, E. *Measurement and Evaluation in Psychology and Education* (Third Edition), New York: John Wiley & Sons, New York, 1969, pp. 655]

Looking at the differences between these definitions, and realizing that the majority of faculty and academic administrators use the dictionary definition rather than the psychometric one, it is easy to see why the question “Are there any valid faculty evaluation data?” persists. There is a large body of stories and anecdotes that gives credence to the popular myths that faculty evaluation data are invalid and unreliable. Unfortunately, there is no similar body of anecdotes that support the position that faculty evaluation data (especially student ratings) are “*well-grounded or justifiable, being at once relevant and meaningful, logically correct, or supported by objective truth or generally accepted authority.*” Which leads us to our third point.

Excellence – I know it when I see it.

The word ‘excellence’, like the word ‘diversity’ has become so overused (and often inappropriately used) in higher education that it has lost its meaning. Once used to designate the ‘best’ or ‘superior,’ the word ‘excellence’ has come to mean a sort of a minimal expectation in virtually every faculty evaluation system. There are many problems with the use of this word in faculty evaluation – the main one being that ‘excellence’ is a term of relative position. That is, in order to be ‘excellent’ a person must be *better than* some one or some group of individuals. ‘Excellence’ is a norm-referenced term and not a criterion-referenced one. Yet, in most faculty evaluation systems the expectation persists that *all* faculty are to achieve excellence. I call this, of course, the lake woebegone model of faculty evaluation. As a profession, higher education is stuck in the silly verbal knot of expecting everyone to be what, by definition, only one or a few can be.

The vast, underlying problem in the whole of faculty evaluation is the fact that the academy has not come forward with a universally accepted definition as to what constitutes an excellent teacher. If we had some list of characteristics, some specific description of the qualities and characteristics that constitute an excellent teacher, then faculty evaluation would be relatively easy. Many faculty and administrators consider the main component of teaching excellence to be content expertise. Others hold that teaching excellence is some sort of ephemeral, immeasurable characteristic that results in some long-term (and perhaps never seen by the instructor) effect on student lives. We may never solve this particular problem to the satisfaction of all but, as an informal effort based on a career-long assimilation of research literature and professional experience, I would suggest the following as a jumping off point:

Characteristics of an Excellent Teacher

- ***Content Expertise***
 - Obviously a faculty member must be knowledgeable in the content field in order to teach it. However, content expertise is a *necessary* but *insufficient* quality for teaching excellence
- ***Affective Traits/Skills***
 - Enjoy teaching as much or more than they enjoy working in their field.
 - Model the best characteristics of an accomplished *practitioner* in the fields they are teaching.
 - Model the best characteristics of a *life-long learner*.
 - Is demanding but fair.
 - Is ethical and honest.
 - Is comfortable admitting ignorance.
- ***Performance Skills***
 - Speaks clearly.
 - Is organized when making a presentation.
 - Uses personal examples when teaching.

- Uses humor effectively.
- Creates an appropriate psychological environment for learning.
- ***Cognitive Skills***
 - *Instructional design* – develops and uses learning objectives in designing effective learning experiences;
 - *Instructional delivery* – skilled in presenting information in a variety delivery modes
 - *Instructional assessment* – skilled in the design and use of a variety of tools and procedures to assess student learning.

In Conclusion

In addressing the problem of faculty and administrators continuing to question the validity of faculty evaluation data it is necessary to take a different approach than simply conducting more studies that result in articles published in journals that appeal to a specific subset of educators. Rather, it is necessary to deal with the *perception* of the invalidity of faculty evaluation data rather than the researchable fact. This perception arises from several situations, the main ones having been described above. Therefore, I would propose the following:

1. The educational research community should undertake a concerted effort to reach the popular press with positive (but true) anecdotes and stories concerning the effectiveness and positive use of faculty evaluation data. The Chronicle of Higher Education has a fairly obvious bias against student ratings and is quick to publish any negative anecdotal stories concerning their use. A concerted effort must be undertaken to counteract this rather powerful public press force.
2. Educational researchers, especially those who have been most instrumental in producing the research literature on faculty evaluation, should make a concerted effort to produce the kinds of articles that will be published in instruments that are generally read by *all* academic administrators rather than just professionals interested in psychometrics and the psychology of teaching and learning. Such outlets as Magna Publication's *Academic Leader* or Anker Publication's *Department Chair* are a couple of examples.
3. As faculty evaluation professionals we must learn to speak the language of the non-psychometrically sophisticated academic administrator. Those professionals in the field of educational research who have gone on to take high-level academic administrative positions should take the lead in this endeavor. We must become more actively involved in professional organizations to which academic administrators belong, presenting papers and conducting sessions that are non-technical in nature and respond to the needs of administrators faced with difficult personnel decisions, but which are still founded on research findings.
4. Finally, we may wish to turn our attention, as educational researchers, to the task of developing and promulgating in the popular press, a definition of the qualities and characteristics of teaching excellence. Although there is a body of research literature on this issue, it has not yet been packaged and 'sold' to the general academic community (and society in general) in a form that could ultimately reach the level of 'common knowledge'. A first attempt in moving in this direction is presented above in the brief discussion on teaching excellence. It is important to remember the extraordinary power of the popular press in affecting the function of higher education in America. Even incomplete models of excellence, such as that used by US News & World Report to rank colleges and universities, can have an extraordinarily powerful impact on the priorities, and thus the functioning, of higher education. As professionals concerned with the valid and reliable evaluation of faculty performance, we would be remiss in ignoring this reality.