Industrial and Organizational Psychology, **1** (2008), 183–184.
Copyright © 2008 Society for Industrial and Organizational Psychology. 1754-9426/08

"When I use a word …"

GEORGE P. HOLLENBECK
Hollenbeck Associates

This comment started out to fit the "Commentary Guidelines Purpose d: a reflection on the writer's experiences in applying the issues addressed in the focal article in particular organizational settings." My career has been spent toiling in the vineyards rather than the academy, first designing performance appraisal systems and advising on how to use them as an I–O psychologist, and later, as an executive myself, appraising other executives. Very recently, I have taken part in very special sorts of performance appraisals as a member of the Society for Industrial and Organizational Psychology's (SIOP) Awards Committee and as Chair of the Fellowship Committee.

I preface my comments with those remarks because in the course of reading and rereading Murphy's piece, my purpose seems to have shifted from "d" to "a. A critical challenge to one or more aspects of the focal article …". I admit to woeful ignorance of the state-of-the-art in A-rated journals—indeed, the extensive compilation of references was a godsend for me; it sent me scurrying to locate copies of several of articles that, frankly, I cannot believe I missed. The press of business over the years has left me less than well-read, but I have been around the block a few times.

My first reading left me frustrated. If the article's purpose was to stimulate thought, it succeeded with me—I had scribbles on nearly every page, for example, "they don't really believe there is one single true score of performance, do they?"; "You sure can't measure performance if you can't define it; but you can evaluate it;" "I/O's may not take performance rating seriously, but lots of people sure do." But, I could not quite latch onto the reason for my discomfort. Was it because Murphy set up a "straw man," a valid approach certainly, but a straw man is an easy target and attacks are more annoying than enlightening.

On second reading, however, I had an "aha" that sent me reaching for the *Merriam-Webster's Collegiate Dictionary, 10th Edition* and that made me act more like William Safire than a proper I–O psychologist: *my frustration was due to confusion and misuse of words*. That frustration led me to use my Webster's to clarify what indeed these words mean.

Appraisal and *evaluation* mean: "to determine or fix the value of." *Appraisals* and *evaluations* require *judgment*: "the process of forming an opinion or evaluation by discerning and comparing." As a side bar, "subjective judgment" is oxymoronic; all judgments are "subjective" (Webster's: *subjective* [~judgments]). Performance appraisals are judgments not unlike the other judgments we have to make—which car to buy, which spouse to choose, and which movie to go to.

To *rate* means "to estimate the amount of." To *measure* is a little more difficult: "to take or make a measurement" is clearly more akin to *rate* than to *appraise* or *evaluate*. We may not (or may) consciously try to measure

Correspondence concerning this article should be addressed to George P. Hollenbeck. E-mail: geoholl@livingston.net

Address: Hollenbeck Associates, 109 Whippoorwill Drive, Livingston, TX 77351

George P. Hollenbeck, Hollenbeck Associates.

https://doi.org/10.1111/j.1754-9434.2008.00035.x Published online by Cambridge University Press

dimensions of our cars, our spouses and the way we spend our time, but evaluation—forming opinions—is not an insurmountable difficulty for most of us.

The article continually confuses *evaluation* with *measurement*. Contrary to what one might think, people *LOVE* to be observed, measured, and rated (especially executives); they *HATE* to be evaluated. 360 feedback owes its success to the fact that it is, in itself, nonevaluative, asking the observers to measure (how frequently does your boss say good morning?); in contrast, opinion surveys asking employees for judgments about how well their managers are performing have never been greeted enthusiastically by the managers in question! When 360 feedback is used for evaluation rather than description it receives the same reception as ordinary performance appraisal.

Defining *performance* is a little more difficult: "the execution of an action" and implies carrying out an action as prescribed by a set of procedures. We know, of course, that performance is relatively easy to measure in simple jobs (who among us cannot count how many envelopes are stuffed per hour?), but more elusive in complex jobs where, as Murphy points out, observable metrics are seldom sufficient. *Evaluation*, in either case, is still a *judgment*.

Arguments about different models are not very useful unless we are talking about the same things. Viewed as a measurement model, correction for attenuation seems reasonable, if not obvious. Viewed as judgment models, deciding how useful multifactor and mediated models may be requires different types of evidence and research than we have typically done.

These confusions seem to pervade the literature, as well as this article. We might get out of this fix by rethinking our thinking; perhaps it would help for thinkers to begin with examples that are close to home—like grading students or making tenure decisions: What is the metric? How do we measure it? How do we evaluate our measurement? If we truly understand how we make those judgments, we may be well positioned to move to ratings and appraisals in other settings.

I certainly would not presume to be the first person to identify the need for clarity in the distinctions among these terms; unfortunately, I do not have the command of the literature to credit those before me. I invite you to reread the article attending to these definitions to get the full weight of the problem of definitions.

https://doi.org/10.1111/j.1754-9434.2008.00035.x Published online by Cambridge University Press