Available online at www.sciencedirect.com

ScienceDirect

Currents in Pharmacy Teaching and Learning 7 (2015) 662–668

Currents in Pharmacy Teaching & Learning

http://www.pharmacyteaching.com

Methodology Matters

Measuring rater judgments within learning assessments—Part 2: A mixed approach to creating rubrics

Michael J. Peeters, PharmD, MEd, FCCP, BCPS*

University of Toledo, College of Pharmacy and Pharmaceutical Sciences, Toledo, OH

Abstract

Assessments will focus learning, and hence they should be aligned with desired educational outcomes. Frequently human raters (experts) are needed to judge a variety of advanced learners' abilities. To use subjective human judgments effectively, this second part of a two-part article in the Methodology Matters section considers the use of a mixed approach to rubric creation. Prior empirical research further introduces this mixed approach while dual-processing theory helps explain it. Applications and implications are discussed. After reading this article, the readers should be able to (a) recognize the differences between holistic and analytic rubrics, (b) discuss integration with cognition and dual-processing theory, and (c) create a rubric for a learning assessment, using a mixed approach.

Keywords: Rubric; Holistic rubric; Analytic rubric; Pharmacy education; Learning assessment

Situation

Recently I was asked to judge posters at a clinical pharmacy conference using a rubric. There were two sessions of 35 posters, and 13 judges including myself. This rubric had four sections for scoring criteria with each section represented as a rubric row and each row was separated into five columns. Column 1 had a section subheading along with three specific criteria (with each criteria requiring a dichotomous yes/no response); the next four columns represented a rating scale, with the first category descriptor as "none of the three criteria were met," second category as "one of the three criteria was met," third category as "two of the three criteria were met," and fourth category as "all the three criteria were met." The categories were scored with 1–4 points, and a judge's total score for a poster was by summation of the four sections (i.e., out of 16-point maximum). As I judged, my scores were bunched between 10 and 14 points, though only one poster received 14 points. Might this rubric be improved? If a judge does not agree fully with all criteria specified or the weighting of those criteria (each rubric section was worth the same amount of points), the rubric design can "lock" that judge into scoring a certain way. If that judge is fervent in his/her opinion, he/she may deviate from the scoring structure to make his/her total score what that judge desires instead of a simple addition of the rubric sections as was intended.

Methodological literature review

As discussed in Part 1, among many experts and generalists it is commonly accepted that "assessments drive students' learning."[1–7] Students are astute and will learn what they need to succeed in a course's learning assessments.[1,3,5,7] This highlights the need for assessments to be constructively aligned with the course's objectives.[8] Within the current Standards for Educational Testing,[9] reliability (now renamed as an index of "internal structure") is a necessary evidence source.[10,11]

As related to our understanding of assessments overall, a common misconception is that objective assessments are more reliable.[12,13] "Objective assessment" simply means

* Correspondence to: Michael J. Peeters, PharmD, MEd, BCPS, FCCP, College of Pharmacy and Pharmaceutical Sciences, University of Toledo, 3000 Arlington Ave, Mail Stop 1013, Toledo, OH 43614.

E-mail: michael.peeters@utoledo.edu

http://dx.doi.org/10.1016/j.cptl.2015.06.022

that question scoring treats all learners the same; a computer scores each question response as either right or wrong.[12,13] Common "objective" written assessment formats include multiple-choice, true−false, and matching questions. These questions are created subjectively by the test author and only scored objectively. Common "subjective" formats include long-answer, open-ended, and essay questions. Examinations with open-ended questions have shown similar reliability to those of multiple-choice question assessments.[12] The question remains about whether performance-based assessments, which involve the human judgments of external raters, are reliable.

As Singh notes, "Over the past two decades or so, our understanding of assessment (and learning) has undergone a sea change… The most important shift in our understanding has been with respect to the assessment of soft learning skills [such as clinical decision-making, professionalism and inter-professional collaboration], which do not easily lend themselves to objective assessment."[13] An "objective assessment" is no better than one with "considered subjectivity"[13]—it is just different and each should have a place in a strong assessment program. In fact, Miller's Pyramid[14] (introduced and discussed within the first part of this article[15]) offers that assessment types that are farther up the pyramid require an increasing need for experts[1,3,13] and therefore we must rely more on their considered subjectivity (Fig. 1).

Versions of the widely-endorsed objective structured clinical examination (OSCE) method of assessment gives another example of objective and subjective assessments. An OSCE has a structured rubric for each station to inform the judgment process. By its very name, OSCE is noted as "objective" because in its early versions it had used objective checklists.[10,12,13] Despite its name, recent versions of OSCEs are subjectively scored as rating scales with multiple scoring options are often used. Importantly, ratings by experts have been shown to have similar or better reliability and improved validity compared with using observation-based checklists.[1,3,6,12,16] These findings demonstrate that through some standardization, a simple rubric that condenses complex, "subjective" expert judgments of a performance into simple scoring can become as or more reliable and more valid than "objective" formats.[3] Human judgments do not give inherently poor reliability, but these judgments will need to be categorized and ordered, most commonly with a rubric. The reliability of performance assessments can vary, not because of "objective" and "subjective" assessment methods but because of inconsistencies or inattention to good rubric design practices.

Rubric types

A discussion of rating scales in Part 1[15] is closely interwoven with, and should be qualified by also discussing types of rubrics. Holistic and analytic are two common types of performance-based scoring rubrics. Holistic rubrics use a single scale and summarize an entire (whole) performance. However, analytic rubrics separate the critical aspects of a performance into discrete elements for scoring; each element receives a separate score, which may then be summed into a

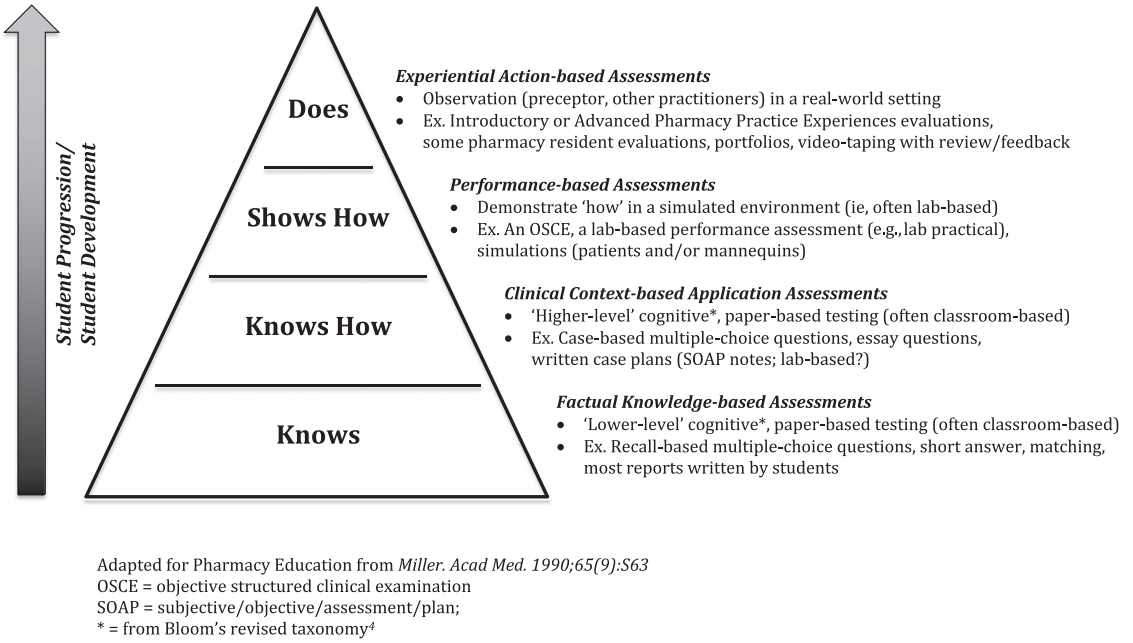

Fig. 1. Miller's Pyramid of Competence for Learning Assessments in Pharmacy Education. (Used with permission from Cor and Peeters.[27]) OSCE = objective structured clinical examination; SOAP = subjective/objective/assessment/plan; * = from Bloom's revised taxonomy.[4] (Adapted for Pharmacy Education with permission from Miller.[14])