Available online at www.sciencedirect.com

ScienceDirect

Currents in Pharmacy Teaching and Learning 7 (2015) 332–341

Currents in Pharmacy Teaching & Learning

http://www.pharmacyteaching.com

Review article

Using generalizability theory for reliable learning assessments in pharmacy education ☆

M. Ken Cor, PhD (Ed Psych)[a], Michael J. Peeters, PharmD, MEd, BCPS, FCCP[b,*]

[a] Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta, Canada
[b] College of Pharmacy & Pharmaceutical Sciences, University of Toledo, Toledo, OH

Abstract

The value of conducting good assessment of learning is becoming an increasingly large focus in pharmacy education. Having a framework to understand learning assessments and recognizing sources of error that contribute to unreliability in measurement are initial steps toward designing more reliable learning assessments in pharmacy education. In this article, we provide a primer on generalizability theory (G-theory), a widely accepted psychometric model used within higher education and present original empirical findings applying G-theory to data from classroom and laboratory pharmacy education as examples. In example 1, we illustrate how the reliability of didactic course grades is affected by the length and number of examinations (i.e., more testing occasions). Our results show that a high level of reliability can be achieved with fewer overall numbers of questions spread out over more occasions of testing. In example 2, we demonstrate how G-theory can be used to establish the reliability of a drug information task in a laboratory-based course. Results reveal that, once again, using more occasions improves reliability of performance assessments. We discuss how the results can be used to begin revising a rater-scoring instrument to improve reliability. This G-theory framework and the worked examples provide a clear path forward for pharmacy educators to consider when developing learning assessments.

Keywords: Psychometrics; Generalizability theory; Reliability; Pharmacy education; Learning assessment

Introduction

A learning assessment framework

Testing with multiple-choice questions is widely used because multiple-choice questions can be scored objectively and quickly; they are often assumed to produce higher reliability than other testing methods like open-ended and short- or long-answer tests.[1,2] Despite these advantages, concerns remain that testing with multiple-choice examinations alone does not adequately assess the full scope of students' abilities developed in pharmacy education. Miller's[3] pyramid illustrates a need to include other types of learning assessments beyond multiple-choice testing to assess higher-level learning objectives. Figure 1, with adaptation to pharmacy education, illustrates how a classroom course's multiple-choice testing can be used to assess "knows" and "knows how" of the pyramid if cased-based (i.e., application-based), while alternate methods such as a pharmacy practice laboratory's performance-based learning assessments are needed to evaluate "shows how" ability-based outcomes.

Colleges/schools of pharmacy have recognized a need for more "shows how" learning assessments in their programs and have begun to implement more performance-based learning assessments as a result. While many of the

☆*Note*: Contents of this manuscript were presented as a Special Session during the 2014 Annual Meeting of the American Association of Colleges of Pharmacy in Grapevine, TX.

* Corresponding author: Michael J. Peeters, PharmD, MEd, BCPS, FCCP, College of Pharmacy & 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.2014.12.003
1877-1297/

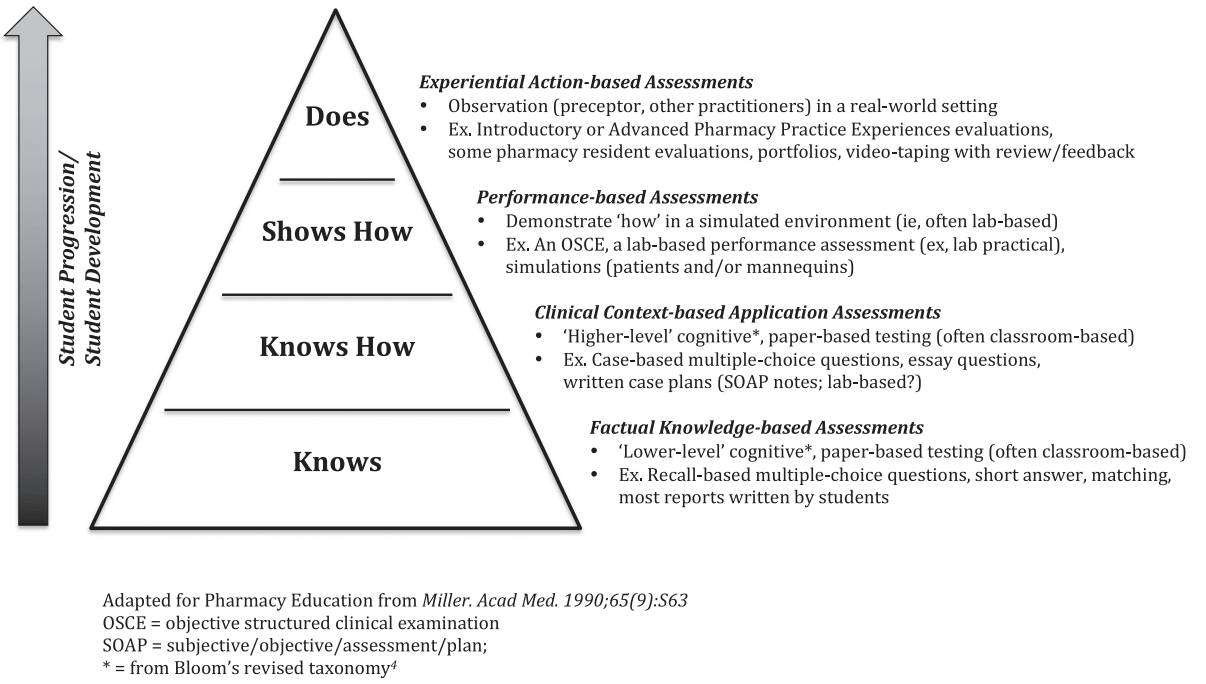

Fig. 1. Miller's pyramid for learning assessments for pharmacy education.

same principles that apply to building reliable multiple-choice-type learning assessments apply to these types of learning assessment, there are some new variables that must also be considered. This article attempts to highlight some of these using specific examples that should resonate with what many pharmacy educators experience in their own programs.

We begin by discussing some of the fundamentals of validity and reliability that apply to all types of learning assessments. While others have covered these fundamentals in more depth,[4,5] a brief summary is provided here to form a basis for some of the specific information that follows. Most importantly, validity is a unitary concept (i.e., there is only one validity), and language for validity involves "validity of conclusions" when applied to proposed or intended assessment score uses. Rather than stating that a learning assessment is simply "valid" or "invalid," we need to seek and justify our uses of test scores by generating multiple sources of validity evidence for our specific assessment context.[4] There are a number of sources of evidence that can be brought forward in order to come to a validity conclusion about the use of a score, and multiple sources are needed in making a stronger validity argument. Content and reliability are important sources of validity evidence to consider with every learning assessment[5]; aligning learning assessment content to course objectives is key to content evidence, while optimizing reliability is also essential for each learning assessment.

We do not mean to minimize or marginalize evidence sources for validity beyond reliability, but focusing on reliability is imperative for evaluative learning assessments. For summative assessments, reliability is a necessary, though insufficient evidence for validity on its own; strong reliability is required, but more validity evidence than reliability alone is needed.[5] With that said, we turn our attention to reliability and improving it within pharmacy education learning assessments. To best improve reliability, we suggest taking a perspective using a generalizability theory (G-theory) framework.

Overview of generalizability theory

Achieving a high level of reliability is a testing standard for any summative assessment of student learning[5,6] and is an ethical imperative.[7,8] As a notable advanced psychometric model, G-theory provides a conceptual framework to understand and account for multiple variables that impact reliability. To that end, G-theory is a widely accepted psychometric model used within many higher education settings that employ complex assessment methods to quantify student learning.[9–12] In particular, it has been used in health professions (e.g., medicine, nursing, midwifery, physical therapy, and pharmacy) as a basis for establishing reliability. While use of G-theory cuts across assessment types, it has primarily been used for establishing the reliability of performance-based assessments.

At its roots, G-theory can be described as a classic measurement model that assumes that a student's observed assessment score is made up of parts "truth" and "error." While the "truth" portion of the score is supposedly fixed,