# Applying Generalizability Theory to Provide Validity Evidence for a Pharmaceutics Course Grade: Combining Exam Reliability Coefficients

Michael J Peeters, PharmD, MEd*; M Kenneth Cor, PhD, MEd, BScEng#; Sai HS Boddu, PhD*; Jerry Nesamony, PhD*

*University of Toledo College of Pharmacy & Pharmaceutical Sciences, Toledo OH, USA

# University of Alberta Faculty of Pharmacy and Pharmaceutical Sciences, Edmonton AB, Canada

## Purpose

- Demonstrate use in pharmacy education of Generalizability Theory (G-theory) to combine reliability from three separate exams into an overall course-grade reliability

## Key Findings *[Implications]*

- G-theory integrated reliability from multiple exams
  - ➢ Reliability of final course-grades was okay but could be improved
  - ➢ Reliability improved if:
    - ✓ More examinations (occasions)
    - ✓ Can have fewer MCQ on each exam (AND fewer MCQ over all exams combined!)
- G-theory provides vital generalization evidence for course-grades
- By extension, reliability for grade-point-average of an entire program will likely be improved further by combining multiple courses

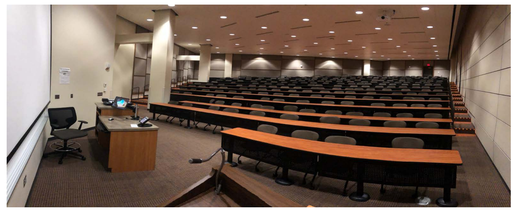
## Why did we do this study? *[Background]*

- Pharmacy colleges/schools should be generating validation evidence for assessments used to make important decisions[1]
- Kane's Framework of validation[2]
  - •Scoring ➢ Generalization ➢ Extrapolation ➢ Implications
- The vast majority of pharmacy education uses Classical Test Theory (CTT), where reliability is limited to each exam's scores[3]
  - ➢ In CTT, individual exam reliabilities cannot be combined from multiple exams/occasions[4]
- While rarely reported in pharmacy education,[3] Generalizability Theory (G-theory) can combine results from multiple exams—reliability can be estimated for more complex measures

## What did we *do*? *[Methods]*

- First-year PharmD students took a pharmaceutics course
  - 12 weeks of lectures
  - Three exams
- Kuder-Richardson Formula-20 (KR20) estimated the CTT-based reliability of each exam
- For G-theory, we used G-String-IV
  - G-theory model was: ***students*** *crossed with* ***items*** *nested within* ***occasions/exams*** (p x i:o)

## What did we find? *[Results]*



101 students took two midterms and one final-exam
- •Exam1=50MCQ (KR20=0.685)
- •Exam2=43 MCQ (KR20=0.647)
- •Exam3=67 MCQ (KR20=0.665)

Our combined G-coefficient/reliability of course-grade was **0.710**

**Our Variation Sources (%)**

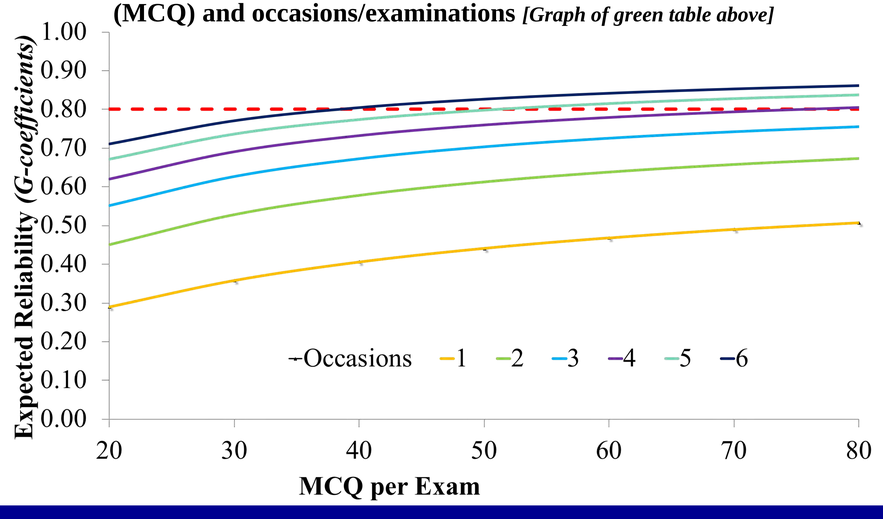
| Sources | % |
|---|---|
| student | 2 |
| occasion | 0 |
| item:occasion | 21 |
| student x occasion | 1 |
| student x item:occasion (& other error) | 76 |

**Our reliability changes (G-coefficients) by number of occasions & items per occasion**

| Occasions \ Number of Items | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|
| 1 | 0.290 | 0.359 | 0.406 | 0.442 | 0.469 | 0.490 | 0.508 |
| 2 | 0.450 | 0.528 | 0.578 | 0.613 | 0.638 | 0.658 | 0.673 |
| 3 | 0.551 | 0.626 | 0.672 | 0.703 | 0.726 | 0.743 | 0.756 |
| 4 | 0.621 | 0.691 | 0.732 | 0.760 | 0.779 | 0.794 | 0.805* |
| 5 | 0.672 | 0.736 | 0.774 | 0.798 | 0.815* | 0.828* | 0.838* |
| 6 | 0.710 | 0.770 | 0.804* | 0.826* | 0.841* | 0.852* | 0.861* |

**Threshold=0.800**

**Our expected reliability with increased multiple-choice questions (MCQ) and occasions/examinations** *[Graph of green table above]*




## References

1. Accreditation Council for Pharmacy Education. Accreditation Standards and Key Elements for the Professional Program in Pharmacy Leading to the Doctor of Pharmacy Degree ("Standards 2016"). Published February 2015. https://www.acpe-accredit.org/pdf/Standards2016FINAL.pdf.
2. Peeters MJ, Martin BA. Validation of learning assessments: a primer. *Curr Pharm Teach Learn*. 2017; 9(5):925-933.
3. Hoover MJ, Jung R, Jacobs DM, Peeters MJ. Educational testing validity and reliability in pharmacy and medical education literature. *Am J Pharm Educ*.2013; 77(10):article 213.
4. Peng SK. Classical versus Generalizability theory of measurement. *Examinations Research*. 2007; 4:009.