A Layman’s Overview of Assessment Theories and Their Practical Applications in Designing Assessments and Analyzing Outcomes

Andrea Shomaker Bourne, ABD
Director of Academic Assessment
Georgia Military College
Discussion Points

- What kinds of in-house assessments are you doing?
  - Institutional Research
  - Academic Affairs
  - Student Affairs
  - Advancement and Alumni Relations
  - Admissions and Marketing

- What are the design and Quality Assurance processes for those assessments?

- Two approaches to assessment design – CTT and IRT

- How can assessments be better-designed to get more from the time spent developing, distributing, collecting, and analyzing? Examples from our institution.

- What kind of tools are there to accomplish more refined assessments?
What kind of in-house assessments are you doing?

- Student affairs, marketing, customer service – Surveys, Inventories of student perceptions
- Learning outcomes – diagnostics, assessment of courses, programs, departments, campuses
- Job performance and teaching efficacy – Surveys, evaluations
- Internal Reviews – Administrative department evaluations
Quality Assurance Process

- Who is involved? Do you have content/function area specialists and assessment specialists honing data gathering tools?
- What really happens? Is there a QA process?
- Who reviews findings for conclusions and assessment refinement? What is their experience and background in assessment modeling? What research supports how you are doing what you are doing?
Two approaches: CTT & IRT

CTT – Classical Test Theory
- CTT provides a model which takes a percentage correct on a test as a measure of ability, ($\theta$), aside from measurement error. From this, standardized, normed, and various other manipulated scores can be derived. In-house testing often does not have enough trait-based questions. Not group ability independent. Test or item bias is hard to pinpoint.

IRT – Item Response Theory
- IRT is based on the notion that for all possible questions or items, one overriding trait is responsible for the answer on that question. While CTT allows us to look at data in terms of test and student data, IRT allows for analysis of test, student, and item data. Ability-independent.
Limitations of CTT

• Tests, surveys, etc. measuring different traits cannot be equated.

• Raw scores on tests, surveys, etc. with varying difficulty and therefore reliability cannot be equated – ability-dependent tests have changing reliability.

• Only strictly parallel and reliable tests, surveys, etc. can be equated.

Advantages of IRT

• Trait-based. Allows for investigation of traits with ability-free mathematical model.
• Allows for assessment specification in advance.
• Good data for purging item bias via IRT DIF and DTF analysis, (Differential Item Functioning and Test Functioning).
• Allows for test and item equating.
• Can build computer adaptive tests. For a CAT, the writing of the program must be able to incorporate the algorithms that are generated based on the adjusting theta value of the examinee.
Examples of how we are using IRT

• Math 109 pre- and post-testing
• Nursing study site – equating with TEAS
• Online campus marketing research
• Establishing parity in learning outcomes among brick-and-mortar campuses and online campus
• Library services surveys – particularly for different populations
• Teacher and campus evaluations that are gain-based (scaled TCCs, ICCs) rather than absolute scores
• IRT and rubrics
• Institution-wide meta-analyses – longitudinal coding on trait-based items (e.g., information literacy, LSS vs. Non-LSS)
What do you need to add aspects of IRT to your research?

“[T]he following conditions must be achieved:

1. The purposes of the assessment are clearly defined.
2. The domain [trait] to be assessed is clearly specified.
3. Item development and selection procedures as well as administration and scoring procedures are accurately documented.
4. Alignment: there is credible evidence on the match between assessments and domain [trait] specifications.
5. Procedures for analyzing, using, and reporting data meet appropriate technical standards.
6. Validation: evidence is assembled to support the intended interpretations and uses of the assessments.”

Tools I use– not a spokesperson

• Online Rubrics, surveys, IRT-based tests: Quia - best I have found for developing and securing CTT- and IRT-based assessments. Good user experience and report-building with knowledgeable test developer.
• Moodle, Desire2Learn, other LMS – need a Learning Object Repository to effectively implement advanced CTT and IRT analyses.
• Qualtrics
• CourseEval
• GoogleDocs
• IRT software:
  • Winsteps http://www.rasch.org/software.htm
  • Bilog http://www.ssicentral.com/irt/index.html
• General online statistical support: http://statpages.org/
• Excel, SPSS, Text files, etc.
References

Web Pages

- http://www.rasch.org
- http://luna.cas.usf.edu/~mbrannic/files/pmet/irt.htm
- http://www.metheval.uni-jena.de/irt/VisualIRT.pdf
- http://ncme.org/linkservid/65E934D9-1320-5CAE-6E2B1CA07A87B6ED/showMeta/0/
- https://onlinecourses.science.psu.edu/stat504/node/113

Books

- Fundamentals of Item Response Theory, Hambleton, Swaminathan, & Rogers
- Handbook of Modern Item Response Theory, Van der Linden & Hambleton
Contact Information and Evaluation

Andrea Bourne  
Director of Academic Assessment  
Georgia Military College  
478.387.1939  
abourne@gmc.cc.ga.us

Evaluation link:  
http://www.iota.cc/iota/tamusurvey.asp