Lecture 1: Item Response Theory (IRT) vs. Classical Test Theory (CTT)

(Baker & Kim (2004), Ch. 1 & Embretson & Reise (2000), Chs. 1-2)

A brief historical overview

- Initialized from the work of Spearman's (1904) – development of the common factor model, concepts of true score, reliability and validity. Gulliksen (1950) contained a comprehensive presentation of CTT within a Pearsonian framework.

- Binet & Simon (1916) and Terman & Merrill (1937) used a tabular representation of the functional relation between the proportion of correct response to an item and chronological age to place items within their intelligence test.

- Lazarsfeld (1950s) introduced *latent structure analysis,* which attempts to deal with "two-valued" or binary items. These types of items present unique problems when analyzed in factor analysis.

- Lawley (1940s) provided a rigorous foundation for the statistical treatment of factor analysis – estimation and hypothesis testing. But implementation of these ideas did not take place until computers became available in the 1950s.

- Lord (1952) and Birnbaum (1968) developed item response theory; its use of normality assumptions originated in the *phi-gamma law*, which was used in psychophysics to relate the probability of detection of a physical stimulus to its intensity.

IRT contrasted with classical test theory (CTT)

Limitations of CTT:

- observed scores and true scores are test dependent
- measures of item difficulty and discrimination are group dependent
- equal errors of measurement for all examinees
- relies heavily on concept of strictly parallel tests, which are generally implausible

Advantages of IRT

- item descriptors are not dependent on the sample of examinees took the test
- examinee ability estimates are defined in relation to the pool of items from which the test is drawn, not the specific sample of items administered each examinee
- can devise tests that are matched to an examinee's ability level
- not based on implausible assumptions (e.g., the idea of strictly parallel tests)
- applications to practical testing problems: differential item functioning, test equating, computerized adaptive testing

Example: CTT vs. IRT in the Golden Rule Insurance Company case