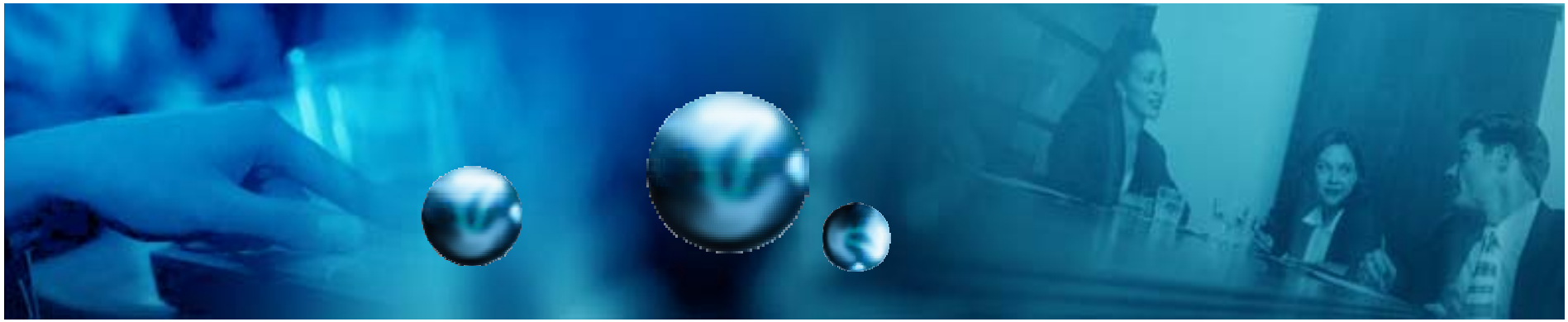
Introduction to Classical Test Theory



Ji Zeng and Adam Wyse

Psychometricians

Michigan Department of Education

Office of Educational Assessment and Accountability

Topics to Cover

- What is Test Theory?
- What is Classical Test Theory (CTT)?
- What are the common statistics used by MDE in the CTT framework?
- What are the general guidelines for the use of these statistics?

What is Test Theory?

Test theory is essentially the collection of mathematical concepts that formalize and clarify certain questions about constructing and using tests, and then provide methods for answering them (McDonald, 1999, p. 9).

What is CTT?

- The main components of Classical Test Theory (CTT) (McDonald, 1999, pp. 4-8) are:
  - --Classical true-score theory
  - --Common factor theory
  - (not discussed in detail in this presentation)

Basic Statistics

- **Sample Mean**: The arithmetic average.

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

-- Mini Example: What is the mean of the following scores?

10, 20, 30, 50, 90

Basic Statistics (Cont.)

- **Sample Variance**: One common way of measuring the spread of data.

$$S_X^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N - 1}$$

Basic Statistics (cont.)

- **Sample Standard Deviation**: Square-root of sample variance, in the same unit of measurement as the original variable.

  -- Mini Example:

  What is the sample variance and sample standard deviation of the data shown on slide 5?

Basic Statistics (cont.)

- **Sample Covariance**: Summarizes how two variables X and Y are linearly related (or vary together).

$$S_{XY} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{N-1}$$

Basic Statistics (cont.)

- **Sample Correlation**: The covariance rescaled, and is completely independent of the unit of measurement in which either X or Y is measured. It ranges from -1 to +1.

$$r_{XY} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}}$$

Common Statistics in CTT

- The four major statistics MDE examines or reports in the framework of CTT are:

  (1) item difficulty

  (2) item-test correlation

  (3) reliability coefficient

  (4) standard error of measurement (SEM)

Classical True-Score Theory

$$X = T + E$$

Where

X represents an observed score,

T represents a true score, and

E represents an error, with the population mean 0.

Reliability Coefficient

- Reliability is the precision with which the test score measures achievement.
- Higher reliability is desired. Why?
- Generally, we would like to have reliability estimates >=0.85 for high stakes tests. For classroom assessment, it should be >=0.7.

Reliability Coefficient (cont.)

There are three main recognized methods for estimating the reliability coefficient:

1. Test-retest (coefficient of stability)
2. Parallel or alternate-form (coefficient of equivalence)
3. Internal analysis (coefficient of internal consistency)

Reliability Coefficient (cont.)

- The reliability coefficient reported by MDE in the framework of CTT is Coefficient Alpha. The estimation of Coefficient Alpha is:

$$\hat{\alpha} = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{N} S_i^2}{S_X^2}\right)$$

where k is the number of items on the test

Reliability Coefficient (cont.)

- Coefficient alpha can be used as an index of internal consistency.
- Coefficient alpha can be considered as the lower bound to a theoretical reliability coefficient.
- Why is this lower bound useful? The actual reliability may be higher!

Standard Error of Measurement

- The *Standard Error of Measurement* (SEM) is a number expressed in the same units as the corresponding test score and indicates the accuracy with which a single score approximates the true score for the same examinee. In other words, SEM is the standard deviation of the error component in the true-score model shown on slide 11.

SEM (cont.)

- Mathematically, SEM can be computed using sample data as follows:

$$S_E = S_X \sqrt{1-\hat{\alpha}}$$

where S_X represents the sample standard deviation of test scores,

$\hat{\alpha}$ represents the estimated reliability coefficient.

SEM (cont.)

- Only one estimated SEM value for all examinees' scores in the given group.
- Given a fixed value of sample standard deviation of test scores, the higher the reliability of the test, the smaller the SEM.

SEM (cont.)

- Sometimes, the students' obtained score is reported on a score band, with the end of the score band computed using the value of estimated SEM.

- If the score band is composed by subtracting or adding one estimated SEM, then there is about 68% chance that the score band covers the student's true score. If we constructed the band by subtracting or adding two estimated SEM, then there is about 95% chance that the score band covers the student's true score.

Item Difficulty

- For dichotomously scored items (1 for correct answer and 0 for incorrect answer), *item difficulty* (or p-value) for item *j* is defined as

$$p_j = \frac{\text{Number of examinees with a score of 1 on item } j}{\text{Number of Examinees}}$$

-- Mini Example:

What is the item difficulty if 85 out of 100 examinees answered the item correctly?

Item Difficulty (cont.)

- Item difficulty is actually the item mean of 0/1 type of data.
- Item difficulty ranges from 0 to 1.
- The higher the value of item difficulty, the easier the item.
- Item difficulty is sample dependent.

Item Difficulty (cont.)

- Adjusted p-value for polytomously scored items (this is computed so that the result will be on the similar scale as that of the dichotomous items):

$$p_j = \frac{\text{Item mean for item } j}{\text{Difference between the possible maximum and minimum score points for item } j}$$

-- Mini Example:

What is the adjusted p-value if an item has mean of 3.5 and the possible maximum score is 5, possible minimum score is 0?

Item Difficulty (cont.)

- MDE scrutinizes MEAP items if

  (1) For MC 4 options, p-value <0.3 or >0.9

  (2) For MC 3 options, p-value <0.38 or >0.9

  (3) For CR items, p-value <0.1 or >0.9

Item-Test Correlation

- "The correlation between the item score and the total test score has been regarded as an index of *item discriminating power*" (McDonald, 1999, p. 231).
- The item-test correlation for dichotomously scored items reported by MDE is point-biserial correlation.

$$r_{pbis} = \frac{(Mean_{+} - Mean_{X})}{S_X}\sqrt{\frac{p}{1-p}}$$

Item-Test Correlation (cont.)

- Point-biserial correlation indicates the relation between students' performance on an 0/1 scored item and their performance on the total test.

Item-Test Correlation (cont.)

- For polytomously scored items, Pearson Product Moment Correlation Coefficient is used by MDE. The computation formula using sample data is shown before.

Item-Test Correlation (cont.)

- The corrected formula (each item score is correlated to the total score with the item in question removed) is (McDonald, 1999, pp. 236-237):

$$r_{i(X-i)} = \frac{S_{i(X-i)}}{S_i S_{(X-i)}}$$

where

S_i is the sample variance of item i

$S_{(X-i)}$ is the sample variance of the total score excluding item i

$S_{i(X-i)} = S_{iX} - S_i$

Item-Test Correlation (cont.)

- Higher item-test correlation is desired, which indicates that high ability examinees tend to get the item correct and low ability examinees tend to get the item incorrect.
- Obviously, a negative correlation is not desired. Why?
- MDE scrutinizes items with corrected item-test correlation less than 0.25 (e.g., MEAP).

Item-Test Correlation (cont.)

- Item-test correlation tends to be sensitive to item difficulty.
- Item discrimination indices (such as point-biserial correlation) plays an more important role in item selection than item difficulty.

Limitations of CTT and Relation between CTT and IRT

- Sample dependent
- Test dependent
- Item Response Theory is essentially a nonlinear common factor model (McDonald, 1999, p.9).

References

- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Orlando, FL: Holt, Rinehart and Winston
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Contact Information

Ji Zeng
(517)241-3105
ZengJ@Michigan.gov

Adam Wyse
(517)373-2435
WyseA@Michigan.gov

Michigan Department of Education
608 W. Allegan St.
Lansing, MI 48909