Chapter 7: Criterion-Referenced Measurement

Criterion-Referenced Testing

- Mastery learning
- Standard development
  - Judgmental
  - Normative
  - Empirical
  - Combination
- Guidelines for writing behavioral objectives (Mager 1962)
  - Identify the desired behavior by name
  - Define the desired behavior
  - Specify the criteria of acceptable performance
- Advantages of criterion-referenced measurement
  - Represent specific, desired performance levels linked to a criterion
  - Are independent of the proportion of the population that meets the standard
  - If not met, specific diagnostic evaluations can be made
  - Degree of performance is not important; reaching the standard is
- Limitations of criterion-referenced measurement
  - Cutoff scores always involve subjective judgment
  - Misclassifications can be severe
  - Students who meet the cutoff may no longer be motivated to do better

Setting a Cholesterol Cutoff

Statistical Analysis of CRTs

- Nominal Data
- Contingency table development
- Phi coefficient (PPM)
- Chi-square analysis

Considerations With CRT

The same as norm-referenced testing

- Reliability
  - Consistency of measurement
- Validity
  - Truthfulness of measurement
- Fitnessgram standards
- AAHPERD Physical Best standards

	Truly below criterion	**Truly above criterion**
Did not achieve standard	Correct decision	False positive
Did achieve standard	False negative	Correct decision

- Meeting criterion-referenced standards
- Possible decisions
- Test–retest reliability example
  - Criterion-referenced equivalence reliability between the 1-mile run/walk and PACER
  - A theoretical example of the divergent group method
- Examples of criterion-referenced standards
  - Cholesterol < 240 mg/dl
  - Systolic blood pressure < 140 mmHg
  - Diastolic blood pressure < 90 mmHg
  - Fitnessgram 1-mile run time for boy age 10 < 11:30
  - President's Challenge Health Fitness curl-ups for girl age 14 > 24

CRT Reliability

		Day 2	
		Fail	Pass
Day 1	Fail		
	Pass		

CRT Validity

		Criterion	
		Fail	Pass
Field Test	Fail		
	Pass		

Racquetball Example

- Can a wall volley test serve as a good criterion measure to determine who should enter intermediate racquetball?
- Example
  - Reliability study
  - Validity study

Reliability study

- Set a standard for passing the field test
  - Our standard is set at 25 hits
- You must hit the ball against the front wall at least 25 times in a trial. This meets the standard for entry into intermediate racquetball.
  - A trial consists of the total of two attempts.
- You want to see if players can achieve the standard on each trial. If you determine the consistency of their meeting the standard, this is a criterion-referenced reliability study.

Meet standard on trial 1? Meet standard on trial 2? Cross-tabulation count

		Meet standard on trial 2?		
		Did not meet standard of 25	Did meet standard of 25	*Total*
Meet standard on trial 1?	Did not meet standard of 25	37	6	43
	Did meet standard of 25	2	11	13
Total		39	17	56

Validity Study

- The standard for passing the field test is 25 hits
- We need a criterion measure of true racquetball ability
  - We used self-reported racquetball experience
    - Inexperience = novice player
    - Experienced = skilled or completed beginning racquetball class
- You want to see if experienced players are more likely to achieve the standard on the field test and the inexperienced players are less likely to meet the field test standard. This is a criterion-referenced validity study.

Meet standard on trial 1? Criterion cross-tabulation

		Criterion		
		Inexperienced	Experienced	*Total*
Meet standard on trial 1?	Did not meet standard of 25	33	10	43
	Did meet standard of 25	5	8	13
Total		38	18	56

Meet standard on trial 2? Criterion cross-tabulation

		Criterion		
		Inexperienced	Experienced	*Total*
Meet standard on trial 1?	Did not meet standard of 25	30	9	39
	Did meet standard of 25	8	9	17
Total		38	18	56

Epidemiological Statistics

- Incidence
- Prevalence

Estimates of Risk

- Absolute risk
- Relative risk
- Odds ratio
- Attributable risk
- Results of a hypothetical study relating cholesterol and heart attack mortality