EPSY 6370.01
Educational and Psychological Measurement

INSTRUCTOR: RALPH CARLSON
TERM: FIRST SUMMER SESSION 2018
PHONE: (956) 665-3487
EMAIL: ralph.carlson@utrgv.edu
OFFICE LOCATION: EEDU 1.208
OFFICE HOURS: MONDAY-FRIDAY 2:30-4:30 PM

CLASS TIME AND LOCATION: MONDAY-FRIDAY 12:20-2:20 PM MAGIC 1.208

Textbook and/or Resource Material
No textbook required.

Course Description and Prerequisites
The content of this course will include scaling, variance, scores derived through linear and nonlinear transformations; traditional item analysis and introduction to latent trait models; reliability and true score theories; partitioning total variance into true and measurement error variance and measurement error variance into its different sources; content, predictive, concurrent, and construct validity, models of unbiased assessment; new trends of measurement. These topics will be related to the construction and interpretation of norm and criterion-referenced measures and systematic observational scales.

Methods of Instruction (teaching)

A. Direct (expository) teaching
   1. Lecture method of presentation of content (topic or concept):
      gestalt → parts → gestalt →
   2. modeling
   3. demonstrations
   4. guided practice
   5. independent practice
   6. small group/cooperative learning
   7. measurement and evaluation

This syllabus subject to change in order to better meet course objectives per discretion of instructor.
B. Activities
1. homework (independent practice)
2. class discussion
3. review

C. Methods of Learning for Students
1. mnemonic system(s)
2. elaboration of meaning
3. actively thinking about one’s thinking
4. management of self and information (content or skills to be acquired)

D. Methods of Measurement and Evaluation
1. grading of homework
2. tests (there will be three tests during the semester)

*CROSSWALK COURSE*
This course satisfies: Intellectual, professional, and academic development for graduate students, masters and doctoral, and faculty.

Learning Objectives/Outcomes for the Course

**Student Learning Outcomes**

**After instruction/teaching:**

1. Students will be able to derive the psychometric properties for scales/tests/measures/instruments.
2. Students will know how to derive metrics through linear and nonlinear transformations.
3. Students will understand deriving true variance and the three primary sources of measurement error.
4. Students will understand and be able to interpret standard error of measurement and standard error of difference.
5. Students will understand and be able to derive the four types of validity.

**Course Technology**

Tools: SPSS (Statistical Package for the Social Sciences)

This syllabus subject to change in order to better meet course objectives per discretion of instructor.
Technical Knowledge Requirements
You are expected to be proficient with installing and using basic computer applications and have the ability to send and receive email attachments.

Major Requirements, Demonstration of Mastery and Evaluation
Assignments and Assessments

There will be three tests given:

1. Test one covering Chapters 1 through 4 in Hinkle
2. Test two covering Chapters 5, 9, 16, 17, 19, and 20 in Hinkle
3. Test three covering Chapters 7, 8, 10, 13, 14, and 15 in Hinkle

UTRGV University Policies

UTRGV Policy Statements

The UTRGV disability accommodation, mandatory course evaluation statement and sexual harassment statement are required on all syllabi. Additional policy statements are optional, such as those covering attendance, academic integrity, and course drop policies.

STUDENTS WITH DISABILITIES: Required on all syllabi. Do not modify.
If you have a documented disability (physical, psychological, learning, or other disability which affects your academic performance) and would like to receive academic accommodations, please inform your instructor and contact Student Accessibility Services to schedule an appointment to initiate services. It is recommended that you schedule an appointment with Student Accessibility Services before classes start. However, accommodations can be provided at any time.
Brownsville Campus: Student Accessibility Services is located in Cortez Hall Room 129 and can be contacted by phone at (956) 882-7374 (Voice) or via email at ability@utrgv.edu.
Edinburg Campus: Student Accessibility Services is located in 108 University Center and can

This syllabus subject to change in order to better meet course objectives per discretion of instructor.
be contacted by phone at (956) 665-7005 (Voice), (956) 665-3840 (Fax), or via email at ability@utrgv.edu.

MANDATORY COURSE EVALUATION PERIOD:

Students are required to complete an ONLINE evaluation of this course, accessed through your UTRGV account (http://my.utrgv.edu); you will be contacted through email with further instructions. Students who complete their evaluations will have priority access to their grades. Online evaluations will be available:

Summer 2018 Mandatory Course Evaluation Period

<table>
<thead>
<tr>
<th>June 6 - June 12</th>
<th>Summer I Module 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>July 2 - July 9</td>
<td>Summer I (standard - 5 week)</td>
</tr>
<tr>
<td>August 1 - August 7</td>
<td>Summer II Module 2</td>
</tr>
<tr>
<td>August 9 - August 16</td>
<td>Summer II (standard - 5 week)</td>
</tr>
</tbody>
</table>

ATTENDANCE:
Students are expected to attend all scheduled classes and may be dropped from the course for excessive absences. UTRGV’s attendance policy excuses students from attending class if they are participating in officially sponsored university activities, such as athletics; for observance of religious holy days; or for military service. Students should contact the instructor in advance of the excused absence and arrange to make up missed work or examinations.

SCHOLASTIC INTEGRITY: Recommended on all syllabi.
As members of a community dedicated to Honesty, Integrity and Respect, students are reminded that those who engage in scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and expulsion from the University. Scholastic dishonesty includes but is not limited to: cheating, plagiarism, and collusion; submission for credit of any work or materials that are attributable in whole or in part to another person; taking an examination for another person; any act designed to give unfair advantage to a student; or the attempt to commit such acts. Since scholastic dishonesty harms the individual, all students and the integrity of the University, policies on scholastic dishonesty will be strictly enforced (Board of Regents Rules and Regulations and UTRGV Academic Integrity Guidelines). All scholastic dishonesty incidents will be reported to the Dean of Students.

This syllabus subject to change in order to better meet course objectives per discretion of instructor.
SEXUAL HARASSMENT, DISCRIMINATION, and VIOLENCE: Required on all syllabi. Do not modify.
In accordance with UT System regulations, your instructor is a “responsible employee” for reporting purposes under Title IX regulations and so must report any instance, occurring during a student’s time in college, of sexual assault, stalking, dating violence, domestic violence, or sexual harassment about which she/he becomes aware during this course through writing, discussion, or personal disclosure. More information can be found at www.utrgv.edu/equity including confidential resources available on campus. The faculty and staff of UTRGV actively strive to provide a learning, working, and living environment that promotes personal integrity, civility, and mutual respect in an environment free from sexual misconduct and discrimination.

COURSE DROPS: Recommended on all syllabi; may be modified by the instructor as long as it is not inconsistent with UTRGV policy.
According to UTRGV policy, students may drop any class without penalty earning a grade of DR until the official drop date. Following that date, students must be assigned a letter grade and can no longer drop the class. Students considering dropping the class should be aware of the “3-peat rule” and the “6-drop” rule so they can recognize how dropped classes may affect their academic success. The 6-drop rule refers to Texas law that dictates that undergraduate students may not drop more than six courses during their undergraduate career. Courses dropped at other Texas public higher education institutions will count toward the six-course drop limit. The 3-peat rule refers to additional fees charged to students who take the same class for the third time.
### Summer I 2018 Term (June 4 – July 10)

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar. 5 (Mon.)</td>
<td>Registration Begins</td>
</tr>
<tr>
<td>May 30 (Wed.)</td>
<td>Payment Due</td>
</tr>
<tr>
<td>June 1 (Fri.)</td>
<td>Waitlist Ends</td>
</tr>
<tr>
<td>June 4 (Mon.)</td>
<td>Last day to withdraw (drop all classes) and receive a 100% refund</td>
</tr>
<tr>
<td>June 5 (Tues.)</td>
<td>Summer I classes begin</td>
</tr>
<tr>
<td>June 6 (Wed.)</td>
<td>Last day to add a class or register for Summer I classes</td>
</tr>
<tr>
<td>June 7 (Thurs.)</td>
<td>Last day to withdraw (drop all classes) and receive an 80% refund</td>
</tr>
<tr>
<td>June 11 (Mon.)</td>
<td>Census Day (last day to drop without it appearing on the transcript)</td>
</tr>
<tr>
<td>June 28 (Thurs.)</td>
<td>Last day to withdraw (drop all classes) and receive a 50% refund</td>
</tr>
<tr>
<td>July 4 (Wed.)</td>
<td>Independence Day. No classes.</td>
</tr>
<tr>
<td>July 9 (Mon.)</td>
<td>Study Day. No classes.</td>
</tr>
<tr>
<td>July 10 (Tues.)</td>
<td>Final Exams</td>
</tr>
<tr>
<td>July 11 (Wed.)</td>
<td>Grades Due at 3 p.m.</td>
</tr>
</tbody>
</table>

This syllabus subject to change in order to better meet course objectives per discretion of instructor.
Course Description
The content of this course will include scaling, variance, scores derived through linear and nonlinear transformation; traditional item analysis and introduction to latent trait models; reliability and true score theories; partitioning total variance into true and measurement error variance and measurement error variance into its different sources; content, predictive, concurrent, and construct validity; models of unbiased assessment; new trends of measurement. These topics will be related to the construction and interpretation of norm and criterion-referenced measures and systematic observational scales.

Required Texts and Other Required Readings


General Policies
Methods of Instruction (teaching)
A. Direct (expository) teaching
1. lecture method of presentation of content (topic of concept):
   gestalt → parts → gestalt
2. modeling
3. demonstrations
4. guided practice
5. independent practice
6. measurement and evaluation

B. Activities
1. homework (independent practice)
2. class discussion
3. review

C. Methods of Learning for Students
1. mnemonic system(s)
2. elaboration of meaning
3. actively thinking about one’s thinking
4. management of self and information (content of skills to be acquired)

D. Methods of Measurement and Evaluation
1. homework (10%)
2. quiz (10%)
3. midterm (30%)
4. final (50%)

Attendance is not required. The student is responsible for any information that is missed during an absence.

Course Objectives
1. Understand the basic concepts in psychological testing
2. Interpret the meaning of the test scores on various measure instruments
3. Read competently the research literature in journals and think critically and analytically about measurement
4. Understand new trends of measurement such as Rasch models and computerized adaptive testing

Specific Competencies/Standards
ExCET Competency (008) for School Counselors, Texas State Board of Examiners of Professional Counselors Licensing Requirements in the area of graduate study in Psychological Measurement, and ExCET Competency (006) for Diagnosticians.

Course Assignments
Course assignments are the problems from the Annastasi’s book. They are usually due at the beginning of every following class.

Grading Policy
A  performance greater than plus one standard deviation above the mean
B  performance between plus and minus one standard deviation
C  performance between minus one and minus two standard deviations below the mean
D  performance minus two standard deviations or more below the mean

Recommended Readings
2. Journal of Educational Measurement
3. Psychometrika
4. Psychological Bulletin
Course Schedule

Week 1. Introduction to Measurement;

Week 2. Psychological testing (Part Two):
   Variability, variance $S^2$, normal distribution $N(0,1)$

Week 3. Psychological testing (Part Three):
   Standard scores, correlation, regression, and multiple regression (optional)

Week 4. Norms: Percentile scores and standard scores (take-home assignment)

Week 5. Norm and criterion-referenced measures; item analysis

Week 6. Item analysis (Cont.); review

Week 7. Test 1

Week 8. Reliability; true-score theory

Week 9. Reliability (Cont.)

Week 10. Reliability (Cont.)

Week 11. Validity; quiz

Week 12. Validity (Cont.)

Week 13. Validity (Cont.); factor analysis (conceptual understanding)

Week 14. New trends of measurement; advanced topics

Week 15. Review

Week 16. Test 2

(This schedule is tentative and subject to change.)

Course outline is presented below:

Measurement

I. Introduction to measurement

II. History of Measurement

III. Statistical Concepts

- Variable (continuous, discrete, independent, dependent)

- Variance

- Scaling levels
  A. Nominal—property of identity; example: categorize and classify events
  B. Ordinal—properties of identity and rank (greater than or less than);
     example: age scale
  C. Interval—properties of identity, rank, and equal distance (how much
     greater than or less than); example: point scale
  D. Ratio or Absolute—properties of identity, rank, equal distance, and a
     natural zero point or reference point; example: Rasch scale
• Score distribution
  A. Univariate Normal N(0, 1)
  B. Skewness
  C. Kurtosis
  D. Depicting data
    1. Raw data
      a. ungrouped data
      b. grouped data (grouping data is anticipated)
    2. Frequency distribution or simply called distribution
      a. cumulative frequency distribution
      b. cumulative percentiles
    3. Graphs (figures)
      a. histogram
      B. frequency polygon
    4. Percentile curve of ogive
• Measures of Central Tendency
  5. Mode
  6. Median
  7. Mean
  8. Two mathematical properties of the mean
  9. Geometric mean
  10. Harmonic mean
• Measures of variability (scatter, dispersion, heterogeneity of individual differences)
  11. range
  12. variance
  13. standard deviation
  14. shape of distribution

IV. Derived scores
  A. Percentiles
  B. Grade equivalents
  C. Age scores
  D. Standard scores (interval scaling)
    1. derived through a linear transformation, i.e., Z scores, T scores, etc.
    2. derived through non-linear transformation, i.e., normal Z deviate,
       deviation IQ, normal curve equivalents (NCEs)

V. Comparing norm reference and criterion reference measures (see handout)

VI. Item analysis
  A. Item difficulty
    1. maximum differentiation and discrimination between and among individuals
    2. symmetry and asymmetry (skewness) of distribution
    3. correlation between error scores and true scores
B. Homogeneity of the measure
   1. point biserial correction analysis
   2. multivariate analysis (factor analysis and regression analysis)

VII. Correlation
A. Covariance
B. Bivariate Correlation (Pearson product-moment correlation coefficient and other bivariate statistical methods)
C. Coefficient of concordance, coefficient alpha (repeated measures analysis of variance or intra class correlation), Kuder-Richardson formula 20 and 21
D. Spearman-Brown formula

VIII. Reliability
A. True score theories
   1. classical true score theory
   2. generic true score theory (Lord & Novick)
B. Three primary sources of error variance in educational and psychological tests (every test, measure, or observational scale has some error and the amount is knowable and the source is specifiable)
   1. time sampling
   2. content sampling
   3. inter scorer or grader difference
C. Partitioning the total variance into measurement error variance and true variance; and partitioning measurement error variance into its difference sources
D. Standard error of measurement
   \[ SEM = S_\varepsilon \sqrt{1 - r_{rel}} \]
E. Standard error of difference
   \[ SE_{\text{Difference}} = \sqrt{SEM_{\text{meas.1}}^2 + SEM_{\text{meas.2}}^2} \]

IX. Validity (both B and C are criterion related validities)
A. Content validity
B. Predictive validity
   1. standard error of estimate
      \[ S.E_{\text{estimate}} = S_y \sqrt{1 - r_{xt}^2} \]
   2. correction for attenuation formulas for the predictors; criterion, and both the predictors and criterion
C. Concurrent or diagnostic validity
D. Construct or theoretical validity
   1. age differentiation
   2. correlations with other measures
   3. factor analysis
   4. internal consistency
   5. effect of experimental variable test scores
   6. convergent and discriminant validation
If you have a documented disability which will make it difficult for you to carry out the work as I have outlined and/or if you need special accommodations/assistance due to the disability, please contact the Office of Services for Persons with Disabilities (OSPD), Emilia Ramirez-Schunior Hall Room 1.101 immediately. Appropriate arrangements/accommodations can be arranged.
Psychometric Theory / Data Theory / Measurement
References

Textbooks:


Journals:

Journal of Educational Measurement
Psychometrika
Psychological Bulletin
Item Analysis
Ralph Carlson
The University of Texas Pan American

I. Item Distribution, Item Difficulty, Total Score Distribution
Item difficulty (P) is defined as the proportion of examinees who answer an item correctly. An item answered correctly by 85% of the examinees would have an item difficulty, or p-value, of .85 whereas an item answered correctly by 50% of the examinees would have a lower item difficulty, or p-value, of .50. Thus, the “easier” item actually has the higher item difficulty value, or p-value. Items with p-values of either 1.00 or .00 yield no variance or information on individual differences and thus are of no psychometric/measurement/data theory value.

When aggregating items the p-values for a set of items will effect the shape of the total (composite) score distribution. For example, if the mean difficulty level for a set of items is .85 (p = .85) the total score distribution will be negatively skewed and if the mean difficulty level for a set of items is .20 (p = .20) the total score distribution will be positively skewed. A set of items may range in difficulty, that is, some “easy” (p = .90) and some more “difficult” (p = .10); however, an approximate average difficulty level of .50 (p = .50) yields three important psychometric properties: (1) the scale will maximize the detection of individual differences; (2) the true variance will not be correlated with the measurement error variance; (3) the total score distribution will be normally distributed and thus, by definition, will have interval level scaling. Please note, including a few “easy” items in a scale provides for a sufficient basal level for the scale and a few “difficulty” items provide for sufficient ceiling for the scale. This psychometric property will

II. Item Discrimination Indices A & B
A. An index of item discrimination that can be applied to dichotomously scored items or items that can have a possible score range of 1 to 4, 1 to 5 or greater, for example, a Likert type scale. When the test item is scored dichotomously, 0 to 1, a correlation with performance on the total (composite) test score (or some other continuously distributed criterion) can be derived. This index of item discrimination is derived through a Pearson product-moment family correlation coefficient called point biserial ($r_{pb}$) or $r_{Dx}$ index).

Interpretation $r_{pb}$ discrimination index for an item:
1. $r_{pb} > .30$ item functioning quite satisfactorily and effect on total (composite) score is additive.
2. $r_{pb} < .30$ item should be modified or deleted and effects on total (composite) score is randomness (noise) [deleting these items will increase Cronbach’s coefficient alpha]
3. $r_{pb} = -.30$ or greater, item should be deleted or modified and effect on total (composite) score is subtractive. Plausible hypotheses that may account for this negative item discrimination index are:
   (A) strong negative $r_{pb}$ values indicates that either the stem (question) and/or foils (choices) are causing difficulty for the higher performing examinees
   (B) the scoring key may be incorrect
Advantages to the $r_{pb}$ item discrimination index:
1. Sampling distribution is known
2. Utilizes all variates and thus all variance (information) is included in the analysis

B. One simple discrimination parameter, sometimes called index of discrimination, can only be applied to dichotomously scored items. Presently, the examinees are separated into two groups, that is, the upper 25% and lower 25% of examinees on the total (composite) score. Then the residual ($D$) between the proportion in the upper groups ($P_{UP}$) who answered the item correctly and the lower group ($P_{LO}$) who answered the item correctly is obtained. $D = P_{UP} - P_{LO}$

Shown below are guidelines for interpretation of D-values:
1. If $D \geq .40$, the item is functioning quite satisfactorily.
2. If $.30 \leq D \leq .39$, little or no revision is required.
3. If $.20 \leq D \leq .29$, the item is marginal and needs revision.
4. If $D \leq .19$, the item should be eliminated or completely revised.

Disadvantages to the D-values as an item discrimination index:
1. Sampling distribution is unknown
2. The D-values does not utilize all variates and thus not all variance (information) in the analysis

Item Response Distribution

Items with a multiple choice format may have 4 foils (choices) with a specified correct answer (response) and distractors. If a foil (distractor) is functioning, a given proportion of examinees will respond to said foil. Item distractors are functioning effectively when they yield a uniform proportion of examinees responding to each distractor.

Bias reviews were conducted on all items for gender concerns. Item desirability information from examiners will also be considered in the selection of the items.