Psychometric Theory is the discipline that addresses the measurement and quantification of psychological phenomena (often refereed to as “latent traits”). Strictly speaking, psychological phenomena are not directly observable. Typically, they must be inferred from observations taken on some behavior that may be observed and is assumed to operationally define the unobservable characteristic (or “variable”) that is of interest. An operational definition is most useful when it delineates the kinds of items or tasks that represent the high/hard and low/easy boundaries of the variable (whether it addresses behaviors, skills, attitudes, etc) of interest and differential points between those boundaries. After this theory driven specification of the variable has occurred, a "scale" comprised of independent items is developed to measure the hypothesized unidimensional variable. Data are then gathered, and various statistical models are then employed, to determine the extent to which the scale (or “measurement instrument”) functioned as intended.

Instructor: Prof. Larry H. Ludlow
Campion Hall 336C 617-552-4221 Ludlow@bc.edu

Theme Quotes:

1. “The Reader may here observe the Force of Numbers, which can be successfully applied even to those things, which one would imagine are subject to no Rules. There are very few things which we know, which are not capable of being reduc’d to a Mathematical Reasoning; and when they cannot, it’s a sign our Knowledge of them is very small and confus’d; and where a mathematical reasoning can be had, it’s a great folly to make use of any other, as to grope for a thing in the dark, when you have a Candle standing by you.” John Arbuthnot, 1692. In I. Todhunter, A History of Mathematical Theory of Probability. (Macmillan,p48-51, 1865).

2. “Psychometry, it is hardly necessary to say, means the art of imposing measurement and number upon operations of the mind...” F. Galton, Psychometric Experiments. Brain, II, 149-162, 1879.

3. “...that until the phenomena of any branch of knowledge have been subjected to measurement and number, it cannot assume the status and dignity of a science.” Galton.

4. "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.” Sir William Thomson, Lord Kelvin. Electrical Units of Measurement. Popular Lectures and Addresses, Vol 1 of 3. (London: Macmillan, 1889, p. 73-74)
5. "The grand, and indeed only, character of truth is its capability of enduring the test of universal experience, and coming unchanged out of every possible form of fair discussion". Sir John Herschel.


**Ludlow's Challenge:**

If it exists, it can be measured;  
If it can't be measured, it doesn't exist.  
Ludlow, L.H. Psychometrics Lectures, Boston College, February 1996

**Course Objectives**

A) Introduce you to Classical Test Theory (or “True Score Theory”) and Item Response Theory (the Rasch model, in particular); and

B) Provide an opportunity for you to develop competent, practical data analysis and statistical/psychometric consulting skills.

You will spend considerable time in the library and on the computer. [It is assumed that you will exert individual initiative in solving computing/analysis problems as they arise.]

**COURSE ASSESSMENTS**

You will be evaluated on the following components:

a) data analyses (CTT, IRT)  
b) literature reactions  
c) measurement essay  
d) Rasch presentation (either in class or at the NEERO conference)  
e) class participation

**Literature Reactions**

The literature reactions (theory critiques, reviews, reaction papers) will take the form of at least 1-2 pages, (greater length is acceptable but is not encouraged) typed and double-spaced. They will be handed in the first six class meetings after the initial lecture. Their purpose is to introduce the literature to you and, in turn, your interests to me.

1) Begin the main body of your discussion with a direct quote from the article and its page number. Following the quote, write an analysis of its meaning to you. Your analysis should not be a paraphrased rendition of the quote but illustrative of your independent thinking on an interesting idea. For example, identify what may be wrong with the author's thinking on a question and suggest how the approach could be improved. Or, when your quote captures the brilliance of someone's thinking suggest ways its application may be broadened. Or, how can what we typically accept as "standard procedure" be improved by an obviously better way? Or, when you have encountered a particularly interesting topic, discuss its research potential.
for you or its potential for incorporation into your current employment or dissertation interests. Or, you may wish to challenge “Ludlow’s Challenge.”

2) Organize the reaction papers and reviews according to the format shown below. In this form, your name and date are in the upper right hand corner and the full literature citation is in the upper left hand corner of the document.

Pearson, K. The Grammar of Science. London: Adam and Charles Black, 1900. Your Name Date

"The classification of facts and the formation of absolute judgments upon the basis of this classification-judgments independent of the idiosyncrasies of the individual mind-essentially sum up the aim and method of modern science."

Now would follow your reaction to the quote.

3) Your first Reaction Paper is to answer the question "What is Measurement?" You may consult any of the materials in this syllabus. BUT, I want to know what you in your own words think constitutes measurement. Your remaining Reaction Papers will be of the form presented in steps (1) and (2) above.

4) No papers are due for the evening in which analyses are submitted.

**Data Analyses**

The data analyses will consist of your interpretation of the output from the psychometric and statistical software programs (including the critical output). You may supply your own data or you may solicit Lynch School faculty for data. A reasonable way to satisfy this course component is to analyze the same data set for each psychometric model. The report should describe the sample, the variable being measured, items of the instrument (including their number and scoring format), the psychometric model and its psychometric properties, the interpretation of whether or not the data fit the model, and what modifications (if any) would improve the instrument.

**Measurement Essay**

The measurement essay will integrate your literature reactions and your understanding of class discussions. This is an opportunity for you to formally summarize your understanding of the essentials of measurement. One reasonable way in which to satisfy this component is to take a single topic and focus each reaction paper on some aspect of that topic. The measurement essay would then trace the development of your research. This essay should be 5-10 pages in length (potentially longer), typed, double-spaced, and fully referenced. In your essay you may include a discussion of topics that remain confusing, or appear as potentially researchable. Potential topics might include: authentic assessment, item banking, tailored testing, computer adaptive testing, Rasch applications, standard setting, one-parameter versus three-parameter models, differential item functioning (DIF), comparisons of estimation algorithms, goodness of fit tests, etc. You might
even address how, if any, your interpretation of the first reaction paper “What is measurement?” has shifted/clarified/been re-defined over the course of the semester.

**Rasch Presentation**

Your last data analysis will close with the Rasch model. You will provide a brief (15-20 minute) class presentation of your results.

**Required Texts**


**Prerequisite courses**

ED462: Assessment and Test Construction
ED469: Intermediate Statistics
ED667: General Linear Models

**PROPOSED TOPICS**

1. **History: Psychophysics to Psychometrics.**

    Principals, principles, and theoretical development.

<table>
<thead>
<tr>
<th><strong>WebCT Link Name</strong></th>
<th><strong>Required Readings:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Chapters 1 and 3 of Crocker &amp; Algina.</td>
</tr>
<tr>
<td></td>
<td>2. Preface of Bond</td>
</tr>
<tr>
<td></td>
<td>3. &quot;Forward&quot; of Wright &amp; Masters.</td>
</tr>
</tbody>
</table>


**Suggested Readings:**


**2. Classical True Score Theory:**

Theory, assumptions, applications. SPSS reliability and factor analysis computer output interpretation of TASC data.

**WebCT Link Name**  **Required Readings:**

1. Chapters 5-7, 13-14 of Crocker & Algina.


Psychometrics Syllabus: Ludlow: March 23, 2005
Lord (1980)  

CTT TASC output  
5. Ludlow, L.H. CTT TASC output.

Suggested Readings:


10. SPSS Chapter on Reliability calculations:  


3. **Guttman's Scale Theory:**
Theory, assumptions, applications. Interpretation of Hillock's Taxonomy of Reading Skills Hierarchy.

**WebCT Link Name**  
**Required Readings:**

- **Stouffer (1950)**

- **Guttman**

- **Ludlow & Hillocks (1985)**

4. **Item Response Theory:**
Basics - item and test characteristic curves, the information function, one-parameter dichotomous/rating scale/partial credit models.

**WebCT Link Name**

- **Jaeger (1987)**
  Measurement Theory
  Uncon. v. Con. Max. Like.
  PROX to UCON via N-R

- **Ludlow & Haley (1999)**

  Lord (1980)  
  Chapter 2

  Reeve: Intro to Modern Psychometrics Syllabus: Ludlow: March 23, 2005
**Required Readings:**

5. Bond Ch 1-2
7. Reeve, B.B. An Introduction to modern measurement theory. Division of cancer control and population sciences, National Cancer Institute.

**Suggested Overview Readings-Past/Present/Future:**

10. Item Response Theory web site:

**General Measurement Articles:**


**5. The Rasch Model:**

Purpose, assumptions, estimation procedures, item and person fit, residual analysis, applications. Computer output interpretation of TASC and TAMP data sets.

<table>
<thead>
<tr>
<th>WebCT Link Name</th>
<th>Ludlow &amp; Haley (1995)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>Details of Rasch Model Estimation</td>
</tr>
<tr>
<td>Wright (1980)</td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td></td>
</tr>
</tbody>
</table>
Required Readings:
1. Chapters 1-5 of Wright & Stone.


Suggested Readings:


Two related articles:


Five related articles:


General other-discipline articles:


Related Books:


Other:
Any issue of Rasch Measurement: Transactions of the Rasch Measurement Special Interest Group. (see me for their location)

Variable Development and Application Examples:


**TAMP/PEDI Projects:**


6. Operation of Psychometric computer programs:
SCALE, WINSTEPS, RUMM, PARSCALE, BILOG, QUEST/CONQUEST.

7. The Two-and Three-parameter IRT Models:
Purpose, assumptions, estimation, model fit, applications.

8. Technical Applications of IRT:
Item banking, adaptive testing, item and test bias, equating, test construction, differential item functioning (DIF), scale anchoring, cut-scores, plausible values.

**Differential Item Functioning:**

**Computerized Adaptive Testing:**

**Equating:**

**Cut-scores:**


Thurstone Bibliography


Thurstone, L.L. The unit of measurement in educational scales. *Journal of Educational Psychology*. 18, 1927, 505-524.


Thurstone, L.L. Attitudes can be measured. *American Journal of Sociology*, 33, 1928, 529-554.


**General References on Science and Measurement**


Cronbach, L.J. Further evidence on response sets and test design. *Educational and Psychological Measurement*, 1950, 10, 3-31


*Psychometrics Syllabus: Ludlow: March 23, 2005*


Thorndike, E.L. *Handwriting*. Teacher's College Record, 11, 2, 1910.

Thorndike, E.L. *Theory of Mental and Social Measurement*. NY: Teachers College, Columbia University, 1916


**Rasch Bibliography** (in my files)


Rasch, G. An Individual-Centered Approach to Item Analysis with Two Categories of Answers (undated mimeograph).


Wright Bibliography

Wright, B.D. and Panchapakesan, N.A.  A procedure for sample free item analysis.  


Wright, B.D. & Douglas, G.A.  Conditional versus unconditional procedures for 
sample-free item analysis.  _Educational and Psychological Measurement_, 37, 1977, 573-
586.

Perline, R., Wright, B.D. & Wainer, H.  The Rasch model as additive conjoint measurement. 
_Research Memorandum Number 24_, The University of Chicago, Department of Education, 
Statistical Laboratory, 1977.

Wright, B.D. & Bell, S.R.  Verifying the unconditional estimation procedure for Rasch item 
analysis with simulated data.  _Research Memorandum Number 26_, The University of 
Chicago, Department of Education, Statistical Laboratory, 1977.

Wright, B.D. & Mead, R.J.  The use of measurement models in the definition and application 

Wright, B.D. & Douglas, G.A.  Best test design and self-tailored testing.  _Research 
Memorandum Number 19_, The University of Chicago, Department of Education, Statistical 
Laboratory, 1975.

Wright, B.D.  The Rasch Model for Test Construction and Person Measurement. 
Paper prepared for Fifth Annual Conference and Exhibition on Measurement and 

Wright, B.D. & Masters, G.N.  The measurement of knowledge and attitude.  _Research 
Memorandum Number 30_, The University of Chicago, Department of Education, Statistical 
Laboratory, 1980.

Masters, G.N. & Wright, B.D.  A model for partial credit scoring.  _Research Memorandum 
Number 31_, The University of Chicago, Department of Education, Statistical Laboratory, 
1981.

Wright, B.D., Mead, R.J. & Bell, S.R.  BICAL:  Calibrating items with the Rasch model. 
_Research Memorandum Number 23_, The University of Chicago, Department of Education, 
Statistical Laboratory, 1980.

Wright, B.D. Solving Measurement problems with the Rasch model.  _Journal of Educational 

Wright, B.D. & Douglas, G.A.  Rasch item analysis by hand.  _Research Memorandum_
Psychometrics Syllabus: Ludlow: March 23, 2005


Guttman Bibliography (in my files)


Guttman, L. A Basis for Scaling Qualitative Data. *American Sociological Review*, 9, 1944, 139-150.

Relevant Dissertations (Chapters with Rasch or IRT model descriptions)


Special Edition Journals


Psychometrics Syllabus: Ludlow: March 23, 2005
(“Multidimensional IRT”)

Medical Care. Vol 42, Number 1 supplement, January 2004. (“International Conference on Objective Measurement: Applications of Rasch Analysis in Health Care”).

New Directions for Testing and Measurement. (a series of quarterly publications on various current measurement models, testing issues, and applications of techniques-published by Jossey- Bass).


**Relevant Journals**

| Applied Measurement in Education | Applied Psychological Measurement |
| Educational & Psychological Measurement | Educational Measurement: Issues and Practice |
| Educational Researcher | Journal of Abnormal & Social Psychology |
| Journal of Educational Measurement | Journal of Educational Psychology |
| Journal of Behavioral (formerly Educational) Statistics |
| Journal of Experimental Education | Journal of Mathematical Psychology |
| Journal of Outcome Measurement | Popular Measurement |
| Psychological Bulletin | Psychological Review |
| Psychometrika | 
| Rasch Transactions (AERA SIG: Rasch Measurement) | 

**Web Sites**

Finally, check these interesting IRT-related web sites:

http://www.rasch.org/
http://work.psych.uiuc.edu/irt/tutorial.asp
http://www.acer.edu.au

(this site has excellent large-scale examples of Rasch applications)

And this list of relevant sources was compiled by former ERME student Steve Stemler

There are several bibliographies referenced from [www.rasch.org/rmt/index.htm](http://www.rasch.org/rmt/index.htm) - See second panel.

For the Facets model, see [www.winsteps.com/facetman/references.htm](http://www.winsteps.com/facetman/references.htm)

There is a great list of Rasch references on the Journal of Applied Measurement web page ([http://www.jampress.org/](http://www.jampress.org/)) under their "Guidelines for Manuscripts" link.

See the bibliography in "Applying the Rasch Model" by Trevor Bond and Christine Fox: [www.erlbaum.com](http://www.erlbaum.com)
Classical True-Score Theory Assignment
(Spring 2004-100 points)

Upon your data set, use SPSS procedures to perform a classical test theory (CTT) item analysis. Provide an answer to all of the following questions. An outline format is preferable. There is no need to try to write the assignment as a mini-publication at this point.

1. Instrument and sample:
   Explain the purpose of your measurement instrument. What does the instrument purport to measure? Who developed it (wrote the items)? How many items are included? What is the scoring format? How many response options are provided? Is it a speeded test? How long does it take to answer? Is it a standardized or non-standardized instrument? Is it primarily for norm-referenced or criterion-referenced purposes?
   Where did your sample come from? Who collected the data? How many subjects are there? Are they a subset of a larger study and, if so, briefly explain why they were specifically chosen. Are there any special characteristics about them? What is the population to whom they are generalizable?

2. Measurement model:
   I. Technical material
      a). Explain the statistical form of the CTT model and where it came from. Present the relevant expressions and explain them.
      b). What are the primary assumptions of CTT? Present the expressions and explain them. What are potential problems that might violate them?
      c). Why is the concept of parallel tests important for CTT—what is the theoretical problem that parallel tests solve?
      d). Show how reliability can be expressed as the ratio of two variances.
      e). What is the purpose for disattenuating a correlation—why are correlations attenuated? Show how to correct them and explain why they can lead to illogical results
      f). Show why test reliability tends to increase with test length.

   II. From your data
      Report the following statistics and show how they were computed (what equations led to the statistics):
      a). item difficulty (for dichotomous or rating scale data and what are generally accepted preferred difficulty ranges,
b) discrimination (use the corrected item-total correlation and explain why it is corrected and what are generally accepted preferred discrimination ranges),

c) reliability (split-half, KR-20, Cronbach alpha and when you might use one over the other), and

d) standard error of measurement (pick one form of reliability estimate and explain why you chose it and show how the SEM is used). Explain the various components of the equations.

e) explain the general purpose of a common factor analysis when it is applied to items of a test (how many factors do you hypothesize for your data?). Briefly explain what an eigenvalue and eigenvector are. Explain what the factor loadings are and what magnitudes we would like. What is a communality and what magnitudes would we like? What are some of the general ways to determine the number of factors in your data? What is the purpose of the scree plot? Why are solutions usually rotated and, in general, explain the difference between an orthogonal and an oblique rotation and why we might prefer one over the other. Explain the general principle of the varimax technique—are other techniques possible? Explain what the KMO, Bartlett’s sphericity, and |R| are and why they are often general procedures computed to determine the appropriateness of factoring a correlation matrix.

3. Analysis:

a. Discuss the distributional characteristics of your item difficulties and person total scores, e.g., are they as intended, are they surprising? Discuss whether your discrimination estimates are reasonable or not. Are there any particular items with statistical problems (what are the statistical problems)? What might have caused the problems, if there are any? Should any items be removed or revised? Interpret the reliability coefficients you obtained.

b. Discuss the results of your initial factor analysis (was your correlation matrix “appropriate” for factoring?) and how you subsequently decided on the number of final factors to retain. What percent of variance was extracted by those factors and what is your opinion of the magnitude of the percent that was accounted for? Was the rotated and plotted final solution interpretable (just plot the first two factors)? What verbal labels did you apply to “name” the factors (and explain why you applied those names)? Was your solution expected or surprising (did you have any idea about what might result from the factor analysis)?

c. What is the reliability of each of the final factors in your solution? How many scores for each tested person would you recommend should be reported?

d. Summarize the quality of your data and what you would do next (e.g. leave it as is or modify the instrument or get a different sample, etc).
4. Submit your write-up and output. A useful way to write your analysis is to cut and paste into it the appropriate tables/graphs/figures that are output by SPSS rather than referring the reader to the pages of your output. (NOTE: pay attention to typo’s and notation errors.)
Item Response Theory Assignment
(Spring 2004: 100 points)

Upon your data set, use SCALE/ WINSTEPS/ RUMM/ PARSCALE to perform an item response theory analysis.

1. Instrument and sample:
   Explain the purpose of your measurement instrument. Who developed it? How many items are included? What is the scoring format? How many response options are provided? Is it a speeded test? How long does it take to answer? Where did your sample come from? How many subjects are there? Are there any special characteristics about them? Basically, I want you to remind me of the characteristics of the data used for the classical analysis. For your data, what is the “variable” that is being measured? That is, what is the hypothesized structure that is to be tested by the Rasch model?

2. Measurement model details:
   Explain the statistical components of the Rasch model. Why is it called a one-parameter model when clearly there is a parameter for both persons and items? What are the primary assumptions of the model?
   Explain how the initial PROX person ability estimates and item difficulty estimates are computed. Why are persons and items with perfect correct or zero scores removed from analysis? What does the term “sufficient statistic” refer to? Explain (in your own words) what person and item “logits” are. How is the “expected” value for a person on any item computed?
   Explain the general difference between a conditional and an unconditional estimation procedure. What does the term “maximum likelihood” refer to? Explain, in general basic terms, how the Newton-Raphson algorithm operates. What is its function?
   How are person and item weighted fit statistics computed? Explain how person and item positive and negative fit statistics may be interpreted. What might be done if an item or person is considered to misfit the model?
3. **Analysis:**

Discuss the initial distributional characteristics of your item difficulties (does it appear to be a relatively easy or hard instrument) and person abilities (do they appear relatively capable or not). Are these findings as intended? For your data, what do “difficulty” and “ability” translate into?

Are there any particular persons with statistical problems? What might have caused them if there are? Are there any particular items with statistical problems? What might have caused them if there are? (How are you defining a “problem” and what have you done to try to locate their source?)

Explain what the “variable map” is and what it reveals about your data. Was your solution expected or surprising? What modifications, if any, would you suggest if the instrument were to be revised and re-administered?

Finally, compare and contrast the Rasch results to your previous classical analysis results. For example, is there any additional insight you have gained about your data? In addition, how does the standard error of measurement associated with a person’s performance differ between the two models?

4. **Submit** your write-up and output. A useful way to write your analysis is to cut and paste into it the appropriate tables/graphs/figures that are output by the software rather than referring the reader to the pages of your output. (NOTE: pay attention to typo’s and notation errors.)