Psy 427
Cal State Northridge
Andrew Ainsworth PhD

# Test Items and Item Analysis



## Item Formats

- Dichotomous Format
  - Two alternatives
  - True/False
  - MMPI/2; MMPI/A
- Polytomous or Polychotomous Format
  - More than two alternatives
  - Multiple choice
  - Psy427 Midterm, SAT, GRE,



## Item Formats

- Distractors
  - Item Formats
    - Incorrect choices on a polychotomous test
    - Best to have three or four
  - BUT -
    - one study (Sidick, Barret, & Doverspike, 1994) found equivalent validity and reliability for a test with two distractors (three items) as one with four distractors (five items).
  - SO, best might be to have two to four (further study is needed)

## Should you guess on polytomous tests?

- Depends… Correction for guessing:

$$Corrected\ \ Score = R - \frac{W}{n-1}$$

  - R is the number correct
  - W is the number incorrect
  - n is the number of polytomous choices
- If no correction for guessing, guess away.
- If there is a correction for guessing, better to leave some blank (unless you can beat the odds)



## Other Test Items

- Likert scales
  - On a rating scale of 1-5, or 1-6, 1-7, etc. where
  - 1 = strongly disagree
  - 2 = moderately disagree
  - 3 = mildly disagree
  - 4 = mildly agree
  - 5 = moderately agree
  - 6 = strongly agree
- rate the following statements….



## Other Test Items

- Likert scales
  - Even vs. odd number of choices
    - Even numbers prevents "fence-sitting"
    - Odd numbers allows people to be neutral
  - Likert items are VERY popular measurement items in psychology.
  - Technically ordinal but are often assumed continuous if 5 or more choices
  - With that assumption we can calculate means, factor analyze, etc.

## Other Test Items

- Category format
  - Like Likert, but with MANY more categories
    - e.g., 10-point scale
  - Best if used with anchors
  - Research supports use of 7-point scales to 21-point scales



## Other Test Items

- Visual Analogue Scale

No Headache — Worst Headache

- Also used in research
  - dials, knobs
  - time sampling



## Checklists & Q-Sorts

- Both used in qualitative research as well as quantitative research
- Checklists
  - Present list of words (adjectives)
  - Have person choose to endorse each item
  - Can determine perceptions of concepts using checklists.

## Checklists & Q-Sorts

- Adjective Checklists (from http://www.encyclopedia.com/doc/1O87-AdjectiveCheckList.html)
  - In psychometrics, any list of adjectives that can be marked as applicable or not applicable
    - to oneself
    - to one's ideal self
    - to another person, OR
    - to some other entity or concept.



## Checklists & Q-Sorts

- Checklists
  - When written with initial uppercase letters (ACL), the term denotes more specifically a measure consisting of a list of 300 adjectives, from *absent-minded* to *zany*
  - Selected by the US psychologist Harrison G. Gough (born 1921) and introduced as a commercial test in 1952.
  - The test yields 24 scores, including measures of personal adjustment, self-confidence, self-control, lability, counselling readiness, some response styles, and 15 personality needs, such as achievement, dominance, and endurance.



## Checklists & Q-Sorts

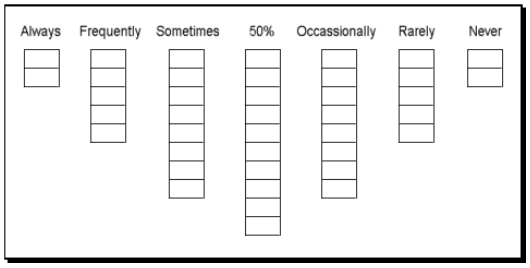
- Q-Sorts
  - Introduced by William Stephenson in 1935
    - PhD in physics 1926; PhD in psychology in 1929
    - Student of Charles Spearman
  - Goal: to get a quantitative description of a person's perceptions of a concept
  - Process: give subject a pile of numbered "cards" & have them sort them into piles
  - Piles represent graded degrees of description (most descriptive to least descriptive).

## Checklists & Q-Sorts

- Q-Sorts
  - Means of self-evaluation of client's current status
  - The Q-Sort consists of a number of cards, often as many as 40 or 50, even 100 items each consisting of a single trait, belief, or behavior.
  - The goal is to sort these cards into one of five columns ranging from statements such as, 'very much like me' to 'not at all like me.'
  - There are typically a specific number of cards allowed for each column, forcing the client to balance the cards evenly.
- Example:
  - California Q-sort , Attachment Q-sort



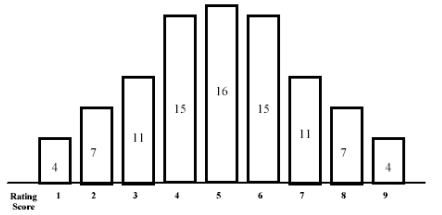
## Example Q-sort






## California Q-Sort

## Attachment Q-sort




Attachment Q-sort Distribution (number of items per pile designated)



## Item Analysis

- Methods used to evaluate test items.
- What are good items?
- Techniques
  - Item Difficulty (or easiness)
  - Discriminability
    - Extreme Group
    - Item/Total Correlation
  - Item Characteristic Curves
  - Item Response Theory
  - Criterion-Referenced Testing



## Item Difficulty

- The proportion of people who get a particular item correct or that endorse an item (if there is no "correct" response, e.g. MMPI)
- Often thought of as the item's easiness because it is based on the number correct/endorsed

## Item Difficulty

- The difficulty can be given in proportion for or it can be standardized in to a Z-value

$$Z = \frac{[\ln(1-p)] - \ln(p)}{1.7}$$



## Item Difficulty

- For example a test with the difficulty of .84

$$Z = \frac{(\ln(.16) - \ln(.84))}{1.7}$$
$$= \frac{(-1.83 + .17)}{1.7}$$
$$= -1.66/1.7$$
$$= -1.00$$
$$(-2 \rightarrow 2 \textit{ is typical range})$$



## Difficult Item (35%)

If you are taking a criterion referenced test in a social psychology course and you need to score a 92 in order to get an A, the criterion is

a) Social Psychology *
b) Scoring a 92
c) Getting an A
d) Not enough info.

## Difficult Item (35%)

$$Z = \frac{[\ln(1-p)] - \ln(p)}{1.7}$$

$$Z = \frac{(\ln(.65) - \ln(.35))}{1.7}$$

$$= \frac{(-.431 + 1.050)}{1.7}$$

$$= .619/1.7$$

$$= .364$$



## Moderate Item (51%)

The correlation between X and is .54. X has a SD of 1.2 and Y has a SD of 5.4. What is the regression coefficient (b) when Y is predicted by X?

a) .12
b) 2.43*
c) .375
d) .45



## Difficult Item (51%)

$$Z = \frac{[\ln(1-p)] - \ln(p)}{1.7}$$

$$Z = \frac{(\ln(.49) - \ln(.51))}{1.7}$$

$$= \frac{(-.713 + .673)}{1.7}$$

$$= -.004/1.7$$

$$= -.00235$$

## Easy Item (100%)

- For the following set of data [5 9 5 5 2 4 ], the mean is
  a) 4
  b) 5*
  c) 4.5
  d) 6



## Difficult Item (100%)

$$Z = \frac{[\ln(1-p)] - \ln(p)}{1.7}$$

$$Z = \frac{(\ln(0) - \ln(1))}{1.7}$$

$$= error$$



## Optimum Difficulty

- Mathematically: half-way between chance and 100%.
- Steps (assuming a 5-choice test)
  1. Find half-way between 100% and chance
     - 1 - .2 = .8, .8/2 = .4
  2. Add this value to chance alone
     - .4 + .2 = .6
- Alternately: Chance + 1.0 / 2 = optimum difficulty
- A good test will have difficulty values between .30 and .70

## Discriminability

- Can be defined in 2 ways:
  1. How well does each item distinguish (discriminate) between individuals who are scoring high and low on the test as a whole (e.g. the trait of interest).
  2. Or simply how well is each item related to the trait (e.g. loadings in factor analysis)
  - 1 and 2 are really the same the more an item is related to the trait the better it can distinguish high and low scoring individuals



## Discriminability

- Extreme Group Method
  - First
    - Identify two "extreme" groups
    - Top third vs. bottom third
  - Second
    - Compute "Difficulty" for the top group
    - Compute "Difficulty" for the bottom group
    - Compute the difference between Top Difficulty and Bottom Difficulty
    - Result = Discriminability Index



| | High Scorers (highest third) | | | Low Scorers (lowest third) | | | Dicrimination |
|---|---|---|---|---|---|---|---|
| | Total | Endorsed | % | Total | Endorsed | % | Dif % |
| " 43. I prefer to pass by people I know ..." | 29356 | 16461 | **0.561** | 28823 | 1708 | 0.059 | **0.501** |
| " 46. I am a very sociable person." | 29356 | 22806 | **0.777** | 28823 | 1378 | 0.048 | **0.729** |
| " 82. I like to go to parties and other affairs ..." | 29356 | 24537 | **0.836** | 28823 | 11289 | 0.392 | **0.444** |
| "151. It makes me uncomfortable to put on a stunt at a party..." | 29356 | 24319 | **0.828** | 28823 | 10581 | 0.367 | **0.461** |
| "160. I find it hard to make talk when I meet new people." | 29356 | 22813 | **0.777** | 28823 | 2033 | 0.071 | **0.707** |
| "178. I wish I were not so shy." | 29356 | 21216 | **0.723** | 28823 | 3531 | 0.123 | **0.600** |
| "245. In a group of people I would not be embarrassed..." | 29356 | 15981 | **0.544** | 28823 | 1985 | 0.069 | **0.476** |
| "248. I am likely not to speak to people until they speak to me." | 29356 | 20076 | **0.684** | 28823 | 1639 | 0.057 | **0.627** |
| "257. In school I found it very hard to talk in front of the class." | 29356 | 24184 | **0.824** | 28823 | 9270 | 0.322 | **0.502** |
| "262. I seem to make friends about as quickly as others do." | 29356 | 16625 | **0.566** | 28823 | 909 | 0.032 | **0.535** |
| "264. I dislike having people around me." | 29356 | 8348 | **0.284** | 28823 | 164 | 0.006 | **0.279** |
| "290. Often I cross the street in order not to meet someone I see." | 29356 | 9008 | **0.307** | 28823 | 1072 | 0.037 | **0.270** |
| "292. I like parties and socials." | 29356 | 20745 | **0.707** | 28823 | 836 | 0.029 | **0.678** |
| "301. I have no dread of going into a room by myself ..." | 29356 | 19948 | **0.680** | 28823 | 3604 | 0.125 | **0.554** |
| "304. Whenever possible I avoid being in a crowd." | 29356 | 21172 | **0.721** | 28823 | 1648 | 0.057 | **0.664** |
| "316. At parties I am more likely to sit by myself or ..." | 29356 | 25105 | **0.855** | 28823 | 1989 | 0.069 | ***0.786*** |
| "319. I love to go to dances." | 29356 | 22092 | **0.753** | 28823 | 8443 | 0.293 | **0.460** |
| "328. I am never happier than when alone." | 29356 | 7040 | **0.240** | 28823 | 606 | 0.021 | ***0.219*** |
| "331. I enjoy social gatherings just to be with people." | 29356 | 20138 | **0.686** | 28823 | 2129 | 0.074 | **0.612** |
| "335. I enjoy the excitement of a crowd." | 29356 | 21954 | **0.748** | 28823 | 4133 | 0.143 | **0.604** |
| "336. I do not mind meeting strangers." | 29356 | 15343 | **0.523** | 28823 | 595 | 0.021 | **0.502** |
| "339. My worries seem to disappear when I get into a crowd ..." | 29356 | 18255 | **0.622** | 28823 | 6973 | 0.242 | **0.380** |
| "408. Some people think it's hard to get to know me." | 29356 | 20346 | **0.693** | 28823 | 5760 | 0.200 | **0.493** |
| "410. I spend most of my spare time by myself." | 29356 | 20786 | **0.708** | 28823 | 4026 | 0.140 | **0.568** |

## Discriminability

- Item/Total Correlation
  - Let the total test score "stand in" for the trait of interest; a roughly estimated "factor" of sorts
  - Correlate each item with the total test score; items with higher item/total correlations are more discriminating
  - These correlations are like rough factor loadings



## Discriminability

- Point Biserial Method
  - If you have dichotomous scored items (e.g. MMPI) or items with a correct answer
  - Correlate the proportion of people getting each item correct with total test score.
  - One dichotomous variable (correct/incorrect) correlated with one continuous variable (total score) is a Point-Biserial correlation
  - Measures discriminability



## Discriminability

- Point Biserial Method

| Item | Item-Total Correlation |
|---|---|
| "43. I prefer to pass by people I know ..." | 0.490 |
| "46. I am a very sociable person." | 0.623 |
| "82. I like to go to parties and other affairs ..." | 0.348 |
| "151. It makes me uncomfortable to put on a stunt at a party..." | 0.350 |
| "160. I find it hard to make talk when I meet new people." | 0.589 |
| "178. I wish I were not so shy." | 0.470 |
| "245. In a group of people I would not be embarrassed..." | 0.444 |
| "248. I am likely not to speak to people until they speak to me." | 0.558 |
| "257. In school I found it very hard to talk in front of the class." | 0.379 |
| "262. I seem to make friends about as quickly as others do." | 0.542 |
| "264. I dislike having people around me." | 0.460 |
| "290. Often I cross the street in order not to meet someone I see." | 0.351 |
| "292. I like parties and socials." | 0.622 |
| "301. I have no dread of going into a room by myself ..." | 0.466 |
| "304. Whenever possible I avoid being in a crowd." | 0.585 |
| "316. At parties I am more likely to sit by myself or ..." | 0.635 |
| "319. I love to go to dances." | 0.352 |
| "328. I am never happier than when alone." | 0.326 |
| "331. I enjoy social gatherings just to be with people." | 0.562 |
| "335. I enjoy the excitement of a crowd." | 0.490 |
| "336. I do not mind meeting strangers." | 0.565 |
| "339. My worries seem to disappear when I get into a crowd ..." | 0.306 |
| "408. Some people think it's hard to get to know me." | 0.385 |
| "410. I spend most of my spare time by myself." | 0.461 |

## Discriminability

- The discimination can be standardized in to a Z-value as well

$$Z = \tfrac{1}{2}\,[\ln(1+r) - \ln(1-r)]$$



## Discriminability

- The discimination can be standardized in to a Z-value as well

| Correlation | Z-Score |
|---|---|
| 0.10 | 0.100 |
| 0.25 | 0.255 |
| 0.50 | 0.549 |
| 0.75 | 0.973 |
| 0.80 | 1.099 |
| 0.90 | 1.472 |
| 0.95 | 1.832 |
| 0.99 | 2.647 |

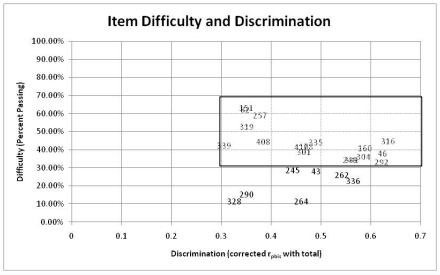


## Discriminability

| Item | Item-Total Correlation | Standardized (Z-value) |
|---|---|---|
| " 43. I prefer to pass by people I know ..." | 0.490 | 0.537 |
| " 46. I am a very sociable person." | 0.623 | 0.730 |
| " 82. I like to go to parties and other affairs ..." | 0.348 | 0.364 |
| "151. It makes me uncomfortable to put on a stunt at a party..." | 0.350 | 0.366 |
| "160. I find it hard to make talk when I meet new people." | 0.589 | 0.676 |
| "178. I wish I were not so shy." | 0.470 | 0.510 |
| "245. In a group of people I would not be embarrassed..." | 0.444 | 0.477 |
| "248. I am likely not to speak to people until they speak to me." | 0.558 | 0.630 |
| "257. In school I found it very hard to talk in front of the class." | 0.379 | 0.398 |
| "262. I seem to make friends about as quickly as others do." | 0.542 | 0.607 |
| "264. I dislike having people around me." | 0.460 | 0.498 |
| "290. Often I cross the street in order not to meet someone I see." | 0.351 | 0.367 |
| "292. I like parties and socials." | 0.622 | 0.729 |
| "301. I have no dread of going into a room by myself ..." | 0.466 | 0.505 |
| "304. Whenever possible I avoid being in a crowd." | 0.585 | 0.671 |
| "316. At parties I am more likely to sit by myself or ..." | 0.635 | 0.750 |
| "319. I love to go to dances." | 0.352 | 0.367 |
| "328. I am never happier than when alone." | 0.326 | 0.339 |
| "331. I enjoy social gatherings just to be with people." | 0.562 | 0.635 |
| "335. I enjoy the excitement of a crowd." | 0.490 | 0.535 |
| "336. I do not mind meeting strangers." | 0.565 | 0.640 |
| "339. My worries seem to disappear when I get into a crowd ..." | 0.306 | 0.316 |
| "408. Some people think it's hard to get to know me." | 0.385 | 0.406 |
| "410. I spend most of my spare time by myself." | 0.461 | 0.498 |

## Selecting Items

- Using Difficulty and Discrimination together



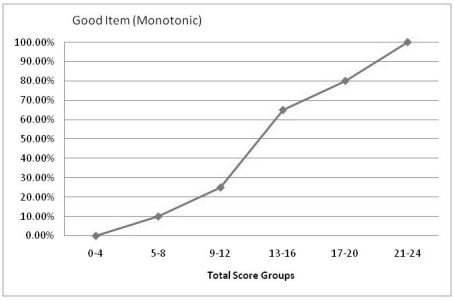



## Item Characteristic Curves

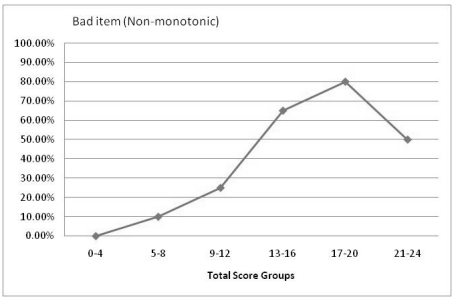
- A graph of the proportion of people getting each item correct, compared to total scores on the test.
- Ideally, lower test scores should go along with lower proportions of people getting a particular item correct.
- Ideally, higher test scores should go along with higher proportions of people getting a particular item correct.



## Item Characteristic Curves

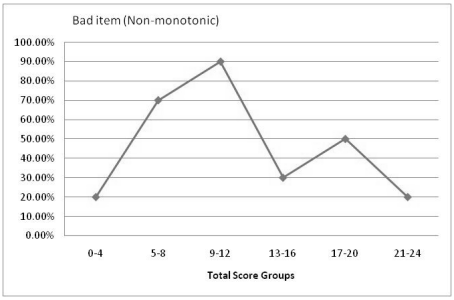


## Item Characteristic Curves



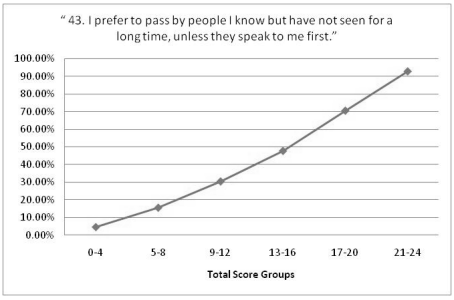



## Item Characteristic Curves

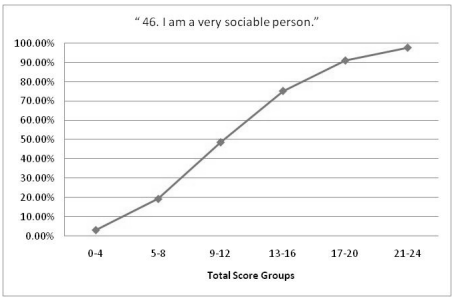





## Item Characteristic Curves

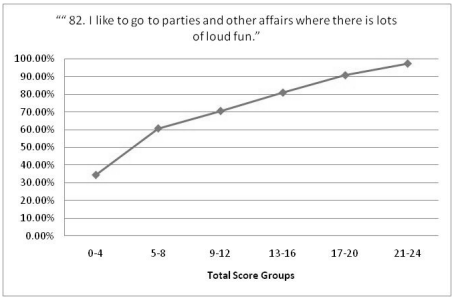


## Item Characteristic Curves



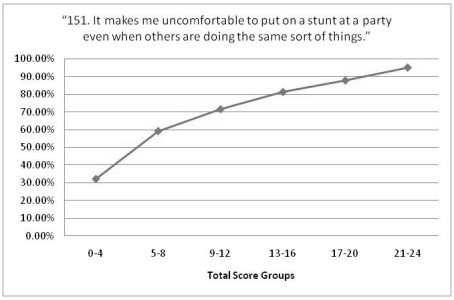



## Item Characteristic Curves

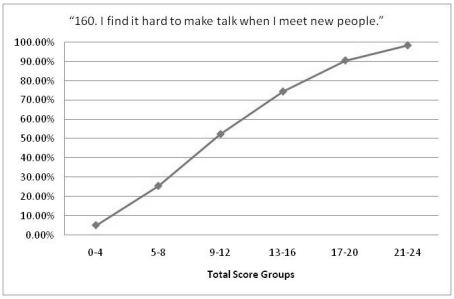





## Item Characteristic Curves

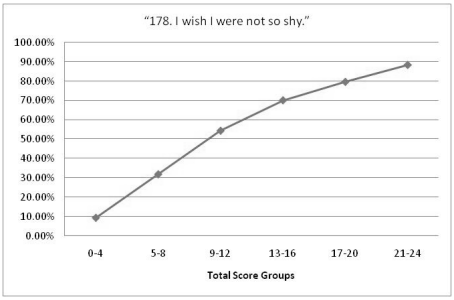


## Item Characteristic Curves



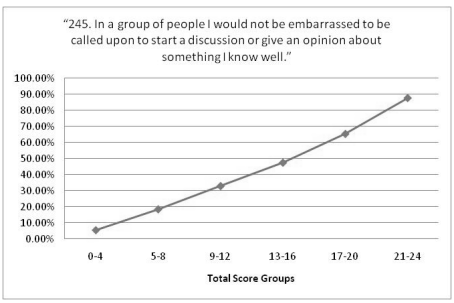



## Item Characteristic Curves

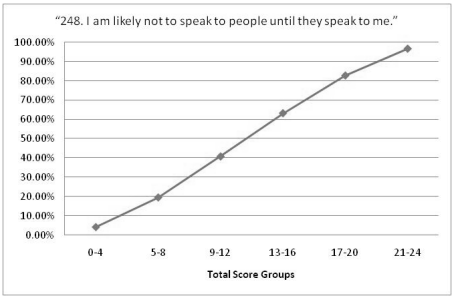





## Item Characteristic Curves

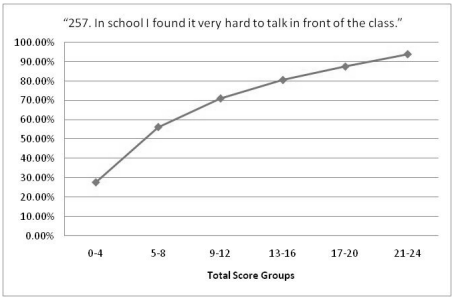


## Item Characteristic Curves



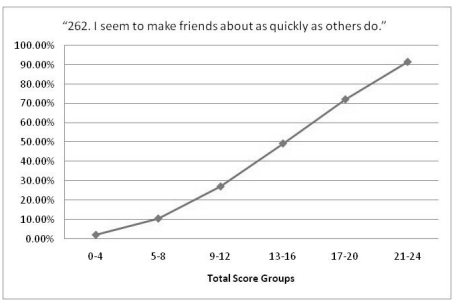



## Item Characteristic Curves

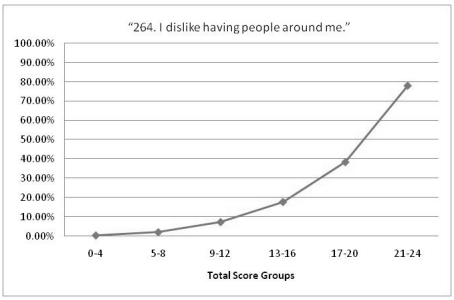





## Item Characteristic Curves

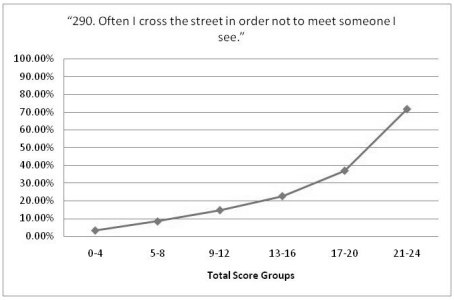


## Item Characteristic Curves



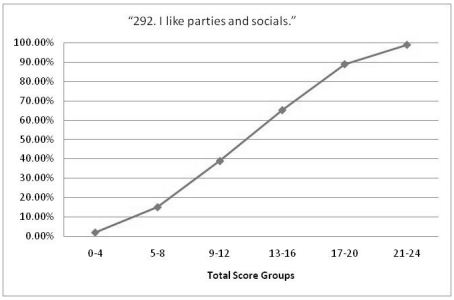



## Item Characteristic Curves

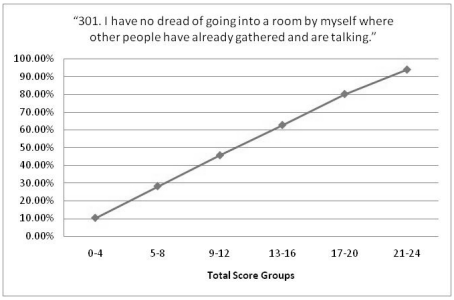





## Item Characteristic Curves

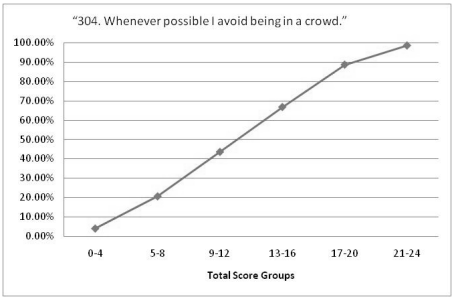


## Item Characteristic Curves



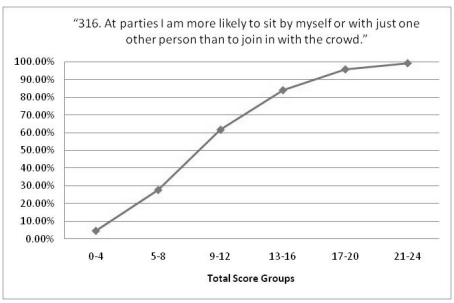



## Item Characteristic Curves

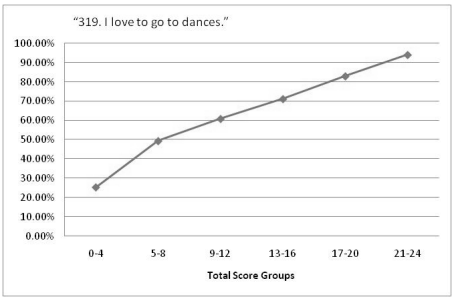





## Item Characteristic Curves

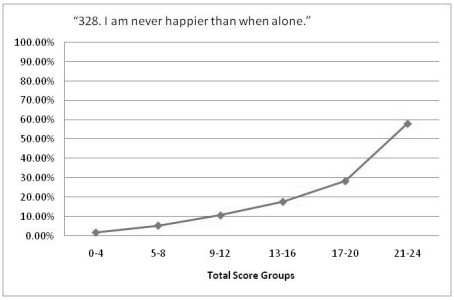


## Item Characteristic Curves



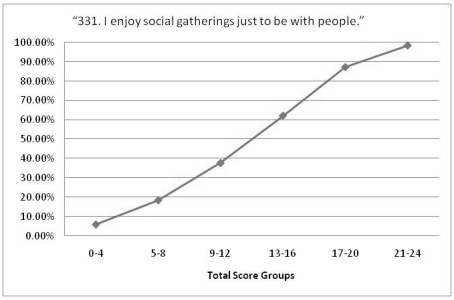



## Item Characteristic Curves

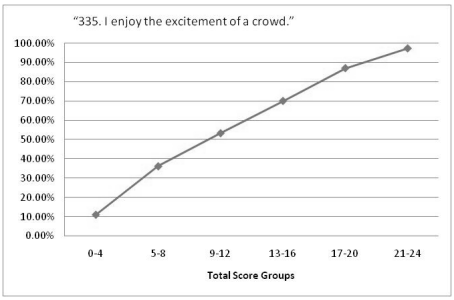





## Item Characteristic Curves

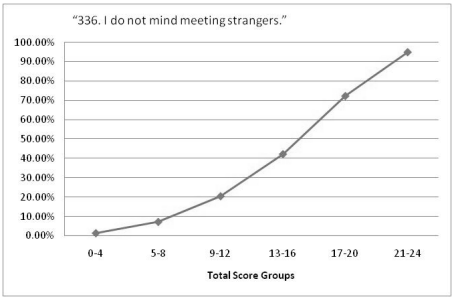


## Item Characteristic Curves



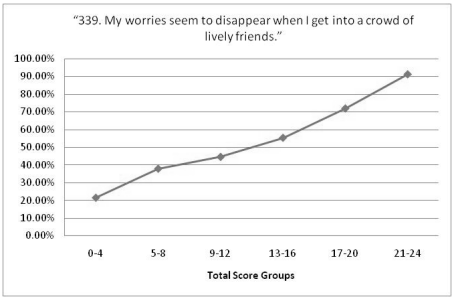



## Item Characteristic Curves

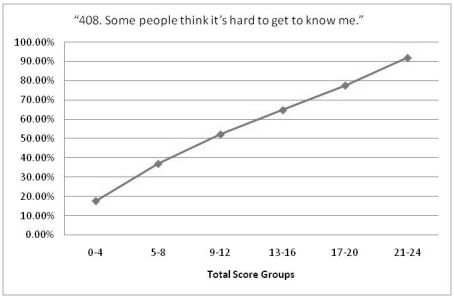





## Item Characteristic Curves

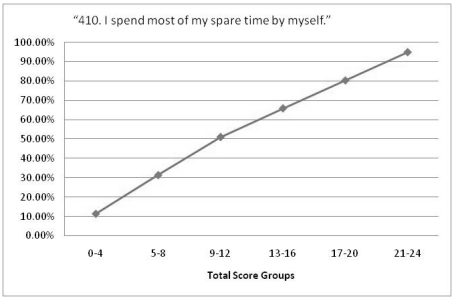


## Item Characteristic Curves






## Item Characteristic Curves






## Other Evaluation Techniques

- Item Response Theory
  - viewing item response curves at different levels of difficulty
  - Looks at standard error at different ranges of the trait you are trying to measure
  - More on this in the next topic

## Other Evaluation Techniques

- Criterion-Referenced Tests
  - Instead of comparing a score on a test or scale to other respondents' scores we can compare each individual to what they "should have scored".
  - Requires that there is a set objective in order to assess whether the objective has been met
  - E.g. In intro stats students should learn how to run an independent samples t-test a criterion referenced test could be used to test this. This needs to be demonstrated before moving on to another objective.

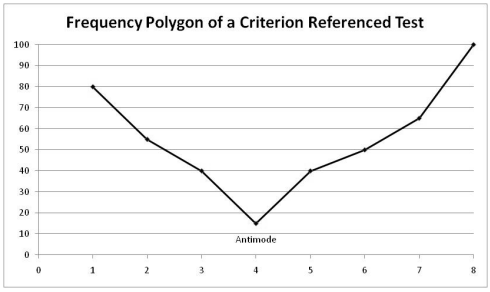


## Other Evaluation Techniques

- Criterion-Referenced Tests
  - To evaluate CRT items
    - Give the test to 2 groups one exposed to the material and one that has not seen the material
    - Distribute the scores for the test in a frequency polygon
    - The antimode (leasts frequent value) represents the cut score between those who were exposed to the material and those who weren't
    - Scores above the cut score are assumed to have mastered the material, and vice versa



## Criterion Referenced Test



## Other Evaluation Techniques

- Criterion-Referenced Tests
  - Often used with Mastery style learning
    - Once a student indicates they've "mastered" the material he/she moves on to the next "module" of material
    - If they do not pass the cut score for mastery they receive more instruction until they can master the material