Principles of Test Construction
Principles of Test Construction

For a psychological test to be acceptable it must fulfill the following three criteria:

1. Standardization
2. Reliability
3. Validity
Standardization

Standardization:
Standardizing a test involves administering the test to a representative sample of future test takers in order to establish a basis for meaningful comparison.

- AP Exams, Intelligence Tests, and SAT (your scores are ranked in average of many test takers)

Norms:
Average score for a designated group of people
Normal Curve

Standardized tests establish a normal distribution of scores on a tested population in a bell-shaped pattern called the normal curve.
Reliability

A test is *reliable* when it *yields* consistent results. To establish reliability researchers establish different procedures:

- Test scores should be same or similar each time the same person takes it.
- 1\textsuperscript{st} attempt SAT – 1800, then 2\textsuperscript{nd} attempt 1000 is this reliable?

1. **Test-Retest Reliability:** Using the same test on two occasions to measure consistency.

2. **Split-half Reliability:** Dividing the test into two equal halves and assessing how consistent the scores are.

3. **Reliability using different tests:** Using different forms of the test to measure consistency between them.
Example of Test Re-test Method

- when determining the effect of coloring on stress, a researcher may choose to evaluate the subjects' blood pressure before, during, and after coloring a set number of pages. This gives researchers a base line from which to draw later conclusions.
Split Half reliability

• one person's odd questions are compared to another person's even questions and if the scores were the same or similar the test would have a high degree of reliability.
  – The AP Psych exam is measured this way
Validity

Reliability of a test does not ensure validity. Validity of a test refers to the degree in which the test is supposed to measure or predict.

1. **Content Validity:** Refers to the extent which a test measures your definition of the construct or behavior of interest?

- Does a physical test measure your knowledge of psychology?
- Does the AP Psychology exam measure your knowledge of psychology?
- Does a physical test measure how athletic you are?
Validity

2. **Criterion-related validity**: Relationship between scores on a test and actual performance
   - Predictive Validity: Refers to the function of a test in predicting a particular behavior or trait.
     - SAT Scores to College Grade
   - Convergent Validity: The degree to which scores on a test correlate with (or are related to) scores on other tests that are designed to assess the same construct.
     - Is your SAT scores similar to ACT scores? Is there a relationship (correlation)?
Validity

3. **Construct Validity** refers to the ability of a measurement tool (e.g., a survey, test, etc) to actually measure the psychological concept being studied.

For example, if we want to know our height we would use a tape measure and not a bathroom scale because all height measurements are expressed in inches and not in pounds.
Intelligence Test
Assessing Intelligence

Psychologists define intelligence testing as a method for assessing an individual’s mental aptitudes and comparing them with others using numerical scores.
Alfred Binet and his colleague Théodore Simon practiced a more modern form of intelligence testing by developing questions that would predict children’s future progress in the Paris school system.
Alfred Binet’s goal became measuring children’s **mental age**, the level of performance typically associated with a certain age.

Identify children who will need extra need (Special Ed.)
Lewis Terman

In the US, Lewis Terman adapted Binet’s test for American school children and named the test the **Stanford-Binet Test**. The following is the formula of **Intelligence Quotient (IQ)**, introduced by William Stern:

\[ \text{IQ} = \frac{\text{mental age}}{\text{chronological age}} \times 100 \]
Stanford-Binet scale

• Stanford-Binet scale
  – Terman’s adaptation of the Binet-Simon scale
  – introduced the “I.Q.” score
  – A score of 100 is considered average
Sample Questions

If a four-year-old girl correctly answered questions on an intelligence exam similar to a five-year-old girl, she would be said to have a mental age of five. In this case her intelligence quotient (IQ) would be:

\[ IQ = \frac{\text{mental age}}{\text{chronological age}} \times 100 \]
Sample Question

- If a five-year-old girl correctly answered questions on an intelligence exam similar to a 6-year-old girl, she would be said to have a mental age of five. In this case her intelligence quotient (IQ) would be:
Sample Question

- Five-year-old Benjy has an IQ of 120 on the original version of the Stanford-Binet. His mental age is
eight-year-old John Cena has an IQ of 80 on the original version of the Stanford-Binet. His mental age is
Aptitude and Achievement Tests

Aptitude tests are intended to *predict* your ability to learn a new skill.

Achievement tests are intended to *reflect* what you have already learned.
Examples

Where would these exams fall under?
1. Pop Quizzes
2. Unit Exams
3. SATs (Enter College)
4. ACTs (Enter College)
5. ASVAB (Enter Military)
6. AP Exams
Examples

• Achievement Tests
  – Unit Exams and AP Exams (Assess what you have learned)

• Aptitude Tests
  – SATs and ACTs (predicts how well you will do in your first year in college)
  – ASVAB (What kind of job would be appropriate for you in the military)
David Wechsler

Wechsler developed the Wechsler Adult Intelligence Scale (WAIS) and later the Wechsler Intelligence Scale for Children (WISC), an intelligence test for preschoolers.
WAIS measures overall intelligence and 11 other aspects related to intelligence that are designed to assess clinical and educational problems.

**VERBAL**

General Information
What day of the year is Independence Day?

Similarities
In what way are wool and cotton alike?

Arithmetic Reasoning
If eggs cost 60 cents a dozen, what does 1 egg cost?

Vocabulary
Tell me the meaning of corrupt.

Comprehension
Why do people buy fire insurance?

Digit Span
Listen carefully, and when I am through, say the numbers right after me.

7 3 4 1 8 6

Now I am going to say some more numbers, but I want you to say them backward.

3 8 4 1 6

**PERFORMANCE**

Picture Completion
I am going to show you a picture with an important part missing. Tell me what is missing.

Picture Arrangement
The pictures below tell a story. Put them in the right order to tell the story.

Block Design
Using the four blocks, make one just like this.

Object Assembly
If these pieces are put together correctly, they will make something. Go ahead and put them together as quickly as you can.

Digit-Symbol Substitution

<table>
<thead>
<tr>
<th>Code</th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>1 5 4 2 1 3 5 4 1 5</td>
</tr>
</tbody>
</table>
Intelligence Tests

• The Wechsler Intelligence Scales
  – The Wechsler Adult Intelligence Scale – Third Edition is the most commonly used test of intelligence for adults
  – WAIS-III is divided into parts, one that focuses on verbal abilities and one that focuses on performance skills
  – Also a version for children, Wechsler Intelligence Scale for Children – Third Edition
Intelligence Tests

• Individual Tests
  – Intelligence tests that are taken by one individual
  – IQ Tests and Personality Tests

• Group Tests
  – Intelligence tests that can be given to large groups
  – Advantages
    • Quick scoring
    • No examiner bias
    • Easier to establish norms
  – Disadvantages
    • Less likely to detect someone who is ill or confused
    • Might make people nervous
    • Learning disabled children often perform worse
Intelligence Tests

• Speed Tests
  – Large number of questions are given in a short period time
  – How quickly can you solve the problems?

• Power Tests
  – Consists of items that get progressively more difficult
  – To what degree of difficulty can you solve the problems?
Intelligence Tests

• Performance tests
  – Tests that minimize the use of language
  – Used to test very young children or people with retardation
  – Also can be used to test those unfamiliar with English

• Culture-fair tests
  – Tests designed to reduce cultural bias
  – Minimize skills and values that vary from one culture to another
Approximate Distribution of IQ Scores in the Population
Flynn Effect

In the past 60 years, intelligence scores have risen steadily by an average of 27 points. This phenomenon is known as the Flynn effect.
Extremes of Intelligence
1. What are some negative terms for “below” average intelligence?

2. In an IQ test what is the value of Mean? What is the value of a Standard Deviation?

3. If you scored 2 standard deviation above the mean on your IQ test, what was your score?

4. How many percentage of people would fall under +1 and -1 SD?
Review

About ninety-five percent of all people fall within 30 points of 100.

Sixty-eight percent of people score within 15 points above or below 100.

Wechsler intelligence score

Number of scores

55 70 85 100 115 130 145

0.1% 2% 13.5% 34% 34% 13.5% 2% 0.1%
The Dynamics of Intelligence

Does intelligence remain stable over a lifetime or does it change? Are individuals on the two extremes of the intelligence scale really that different?
Stability or Change?

Intelligence scores become stable after about seven years of age. In numerous studies, stability of intelligence scores have been determined (Angoff, 1988; Deary et al., 2004).
Normal Curve

- Sixty-eight percent of people score within 15 points above or below 100.
- About ninety-five percent of all people fall within 30 points of 100.
- 68% of scores are within 15 points of 100.
- 95% of scores are within 30 points of 100.
OA
1. % of people who receive 0-100 on IQ test?
2. % of people who receive 85-100 on IQ test?
3. % of people who receive 85-115 on IQ test?
4. % of people who receive 70-100 on IQ test?
5. % of people who receive 70-130 on IQ test?
6. % of people who receive 130- up on IQ test?
Extremes of Intelligence

A valid intelligence test divides two groups of people into two extremes: the mentally retarded (IQ 70) and individuals with high intelligence (IQ 135). These two groups are significantly different.
Mental Retardation

- APA “significantly sub-average general intellectual functioning . . . that is accompanied by significant limitations in adaptive functioning.”
- Must appear before the individual is 21 years old.
- Implies an inability to perform at least some of the ordinary tasks
fun facts

• *morons* (mental age of 7-10) IQ of 51–70
• *imbeciles* (mental age of 3-7) IQ of 26–50
• *idiots* (mental age of below 3) IQ of 25-0

**you do not need to know this for AP exam**
Mental Retardation

Mentally retarded individuals required constant supervision a few decades ago, but with a supportive family environment and special education they can now care for themselves.

<table>
<thead>
<tr>
<th>DEGREES OF MENTAL RETARDATION</th>
<th>Approximate Intelligence Scores</th>
<th>Percentage of Persons with Retardation</th>
<th>Adaptation to Demands of Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild</td>
<td>50–70</td>
<td>85%</td>
<td>May learn academic skills up to sixth-grade level. Adults may, with assistance, achieve self-supporting social and vocational skills.</td>
</tr>
<tr>
<td>Moderate</td>
<td>35–50</td>
<td>10%</td>
<td>May progress to second-grade level academically. Adults may contribute to their own support by laboring in sheltered workshops.</td>
</tr>
<tr>
<td>Severe</td>
<td>20–35</td>
<td>3–4%</td>
<td>May learn to talk and to perform simple work tasks under close supervision but are generally unable to profit from vocational training.</td>
</tr>
<tr>
<td>Profound</td>
<td>Below 20</td>
<td>1–2%</td>
<td>Require constant aid and supervision.</td>
</tr>
</tbody>
</table>

Savant Syndrome

• juxtapositions of severe mental handicap and prodigious mental ability
• Savant performances include
  – Mentally calculating large numbers almost instantly
  – Determining the day of the week for any date centuries away
  – Playing long musical compositions after only hearing it once.
Savant Clips

Alonso Clemens

Daniel Tammet

Steven Wilcher
Causes

• Not much is known for reasons behind mild retardation (90% of all diagnosed)
• PKU (phenylketonuria) – liver fails to produce an enzyme necessary for early brain development
• Down Syndrome – is a genetic disorder caused by the presence of all or part of an extra 21st chromosome.
• Fragile X syndrome – boys harder hit due to only one X chromosome
Giftedness

• Terman – (1925) “academic talent and measured by an IQ in the top two percent of the population.”

• However, highly gifted individuals might possess socially maladaptive behaviors

• Renzulli – (1978) above average general ability, exceptional creativity, and high levels of commitment
Discussion Questions

• Should public schools practice inclusion?
• Should funds be public funds be used for gifted classes?
• Is singling out certain students as gifted elitist?
• Are AP classes for gifted students?