Administering & Grading Classroom Tests

Grade Inflation

Princeton becomes first to formally combat grade inflation
PRINCETON, N.J. (AP) — Princeton University faculty approved a plan Monday to combat rising grades by limiting the number of A's it awards to undergraduates.

The faculty voted 156 to 84 to implement the plan, making Princeton the first college or university to formally curb grade inflation by rationing A's, said Dean of the College Nancy Weiss Malkiel, who proposed the plan. (Related item: Harvard may refocus undergraduate studies)

Under the guidelines, which go into effect in the fall for Princeton's 4,600 undergraduates, faculty are expected to restrict the number of A's to 35% in undergraduate courses. For junior and senior independent work, the percentage receiving A's will be capped at 55%.

A's have been awarded 46% of the time in recent years at Princeton, up from 31% in the mid-1970s. Since 1998, the New Jersey school has encouraged its faculty to crack down, but grades continued to rise. Finally, Princeton administrators decided that the only solution was to ration top grades.

At other Ivy League schools, the percentages of A grades in undergraduates courses ranges from 44% to 55%, according to Princeton's Web site. At Harvard University, 91% of seniors graduated with some kind of honors in 2001.

Suggestions for assembling items

- **Clean up errors -- proofread!**
- **Keep items free from racial, ethnic, and gender bias**
- **Include clear directions -- don't assume students know these**
- **Group items according to item format**

Suggestions for administering items

- **Make students aware of any time limits but assure them adequate time is provided**
- **Make sure students know how to record answers**
- **Explain guessing -- typically have students answer all questions**
- **Don't hold students on the starting line too long**
- **Avoid interruptions**
- **Avoid giving hints that might advantage some students**
- **Discourage cheating**

Purpose of Grading

Grading is often used to serve multiple purposes:
- **Administrative purpose of evaluating students and ranking**
- **Educational purpose of assessing learning and progress toward class objectives**
- **Some people also view grades as motivators**

Shortcomings of Letter Grades:

- **Are typically a combination of achievement, effort, work habits, and good behavior**
- **The proportion of students assigned each letter grade varies from teacher to teacher**
- **They do not indicate a student's specific strengths and weaknesses in learning**

Reliability of classroom measurement

Research shows classroom measurement to be less than perfectly reliable
- **Different teachers asked to grade the same essays often come up grades that differ by several letter grades**
- **Numerical grades are notoriously unreliable**

Norm vs. criterion-referenced tests

- **Norm-referenced testing's primary purpose is administrative– ranking students for selection into programs, tracks, etc.**
- **Criterion-referenced testing is typically the most useful way to assess learning/progress toward educational goals**

Let's think about curving for a minute

Aside from all the other issues around norm-referenced testing, curving should be avoided in education. Consider the following examples:

- **In Oklahoma they created a state-wide exit exam for high school students, and set a criterion of 65% to pass and graduate. Only 30% passed. They "curved" and moved the criterion lower.**
- **In some college classes the average test score might be 40% correct. An "A" might be given for getting half of the items correct.**
- **Now curving is happening with whole states under No Child Left Behind.**
- **What are the messages here?**

3 Options for Grading Systems

- **Relative grading** (norm-referenced) -- performance in relation to other group members
- **Absolute grading** (criterion-referenced/mastery) -- performance in relation to specified standards
- **Improvement/ability grading** -- performance in relation to some determined baseline or starting point

Absolute Grading

A = Outstanding. Student has mastered all the course's major and minor instructional goals

B = Very good. Student has mastered all the course's major instructional goals and most of the minor ones.

Etc.

A = 95% to 100% correct

B = 85% to 94% correct

C = 75% to 84% correct

D = 65% to 74% correct

F = below 65% correct

Must Match!

Assigning grades

Numerical (%) grades are common in education, both today and historically. In order to do this sort of grading:

- **We must assume that there are REAL differences between grades (i.e., that an 82% is different than an 83%)**
- **We must assume that the number represents something REAL**
- **We must assume that our measurement is RELIABLE**

One-chance testing

One-chance testing is the norm throughout preK-postsecondary education. In one-chance testing the students take the test, receive grades, and move on.

- **What is the message here? Ever hear a student, when asked about material studied previously, say something like "Oh, we're done with that"?**
- **Teachers often assume that students will review what they missed and learn the material. Do they?**
- **Is one-chance testing serving an educational or evaluative (administrative) purpose?**

In other areas of "life" involving skill-acquisition is it one-shot evaluation?

Driving test?
Athletics?
Basic (military) training?
Learning to use a computer?
Pottery making?

Learning a language?
Learning to walk?
Learning to play a musical instrument?
Learning to cook/bake?
Learning a trade?

One of the most important abilities to develop in schools is metacognitive ability--the ability to accurately reflect on your learning and employing increasingly effective strategies.

Suggestions for Grading:

- **Assess frequently**
- **Use true FEEDBACK to facilitate learning, base feedback on frequent or constant observations (homework, one-minute quiz, behavioral observation, peer mentoring, etc.– think coaching/mentoring here)**

Self-Report Tests

- **Keep in mind that self-report inventories assumes that individuals are WILLING AND ABLE to report accurately--this may be a big assumption!**
- **Try to also collect some observable measure of the trait of interest**

Interest Inventories

- **Strong-Campbell Interest Inventory** -- suggests particular occupations based upon responses
- **Self-Directed Search (SDS) Career Explorer** -- classifies individual according to occupational themes (realistic, investigative, artistic, social, enterprising, conventional)

Personality

Personality refers to a person's unique and relatively stable pattern of thoughts, feelings, and actions

- Personality is an interaction between biology and environment
  - Genetic studies suggest heritability of personality
  - Other studies suggest learned components of personality

Measures of Personality

Personality refers to a person's unique and relatively stable pattern of thoughts, feelings, and actions (traits)

- **Interviews**
  - **Unstructured:** "Tell me about yourself…"
  - **Structured: Set list of questions**
- **Observation: Psychologist learns about personality by observing the person**
- **Objective tests: self-inventories that involve paper and pencil tests**
- **Projective tests: subjects reveal aspects of their personality when they talk about ambiguous stimuli**

Personality Measurement Issues

Self-report personality tests can be criticized on the basis of

- **<u>Deliberate deception and social desirability bias</u>**
  - **Can the test detect deception and attempts to enhance social desirability?**
- **<u>Inappropriate use</u>: when tests are used for purposes other than their designed use**
  - **Use of a personality test to decide a presidential election**

Projective Tests

- **Projection is an idea developed by Freud in which people are thought to reveal their true feelings and thoughts when describing ambiguous stimuli**
- **A projective test presents a series of ambiguous stimuli and asks that a subject describe each stimulus**
  - **The idea is that their verbal descriptions will reveal key aspects of their personality**
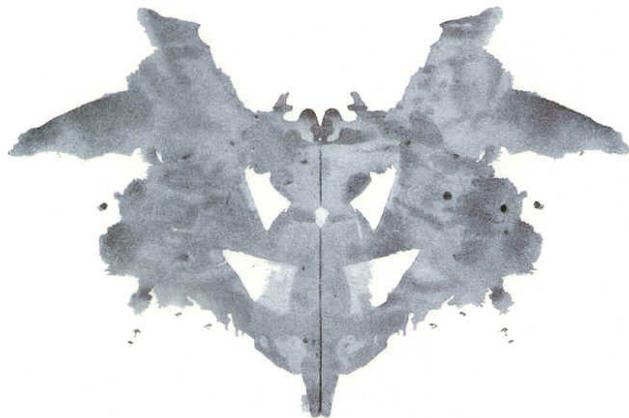
Specific Projective Tests

Rorschach test
- **Consists of 10 inkblots**
- **Reliability and validity of this test is low**

Thematic Apperception Test (TAT)
- **TAT also consists of a series of ambiguous figures**
- **Was devised to measure achievement motivation by Henry Murray in 1938**

Rorschach Inkblot Test

TAT

- *That makes me think of the garden.*
- *It is the city in the country, very much so.*
- *It looks like New York, with the Empire State Building right there.*
- *Calming, relaxing. There's a tree there so you can see the country-side and you've got the background with the city and the buildings, so it's a regional focus.*

The "Big 5"

Modern personality research argues for 5 basic personality traits (OCEAN)

- **<u>Openness</u>**: whether a person is open to new experiences
- **<u>Conscientiousness</u>**: whether a person is disciplined and responsible
- **<u>Extroversion</u>**: whether a person is sociable, outgoing and affectionate
- **<u>Agreeableness</u>**: whether a person is cooperative, trusting, and helpful
- **<u>Neuroticism</u>**: whether a person is unstable and prone to insecurity

Overview of the Big "5"

		Low Scorers	High Scorers
1	Openness	Down-to-earth Uncreative Conventional Uncurious	Imaginative Creative Original Curious
2	Conscientiousness	Negligent Lazy Disorganized Late	Conscientious Hard-working Well-organized Punctual
3	Extroversion	Loner Quiet Passive Reserved	Joiner Talkative Active Affectionate
4	Agreeableness	Suspicious Critical Ruthless Irritable	Trusting Lenient Soft-hearted Good-natured
5	Neuroticism	Calm Even-tempered Comfortable Unemotional	Worried Temperamental Self-conscious Emotional