

# How the New York State Regents Tests Are Constructed

Susan Thomas, Ph. D.  
IBM Global Services



# What makes a test “good”?

## Validity

**Does the test measure what it’s supposed to measure?**

**Evidence: Each item is linked to the test specifications**



## Reliability

**Does the test measure consistently?**

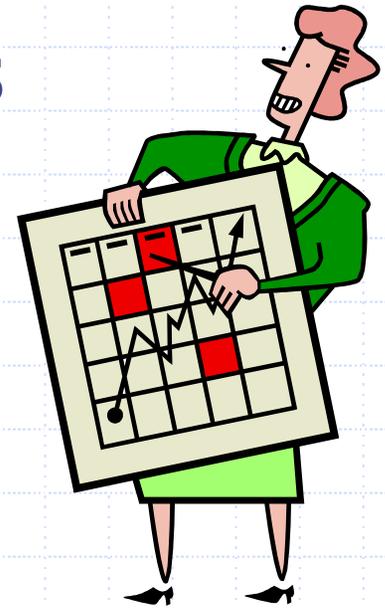
**Evidence: Statistical analysis of test-retest or internal consistency**

***NY SED goal: Provide a valid score for each student***



# How does NY SED create a “good” test?

- ◆ They build in quality
  - Wide participation in the process
  - Training for all participants
- ◆ They make data-based decisions
  - Extensive pretests and field tests
- ◆ They provide open and frequent communication



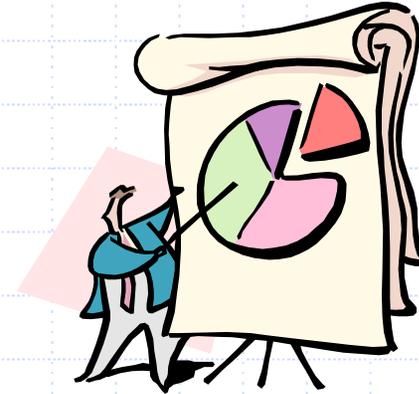
# Let's look at the process NY SED uses...



*Careful planning*



*Great communication*



*Statistical Analyses*



*Trained committees*



*Pretesting and Field testing*



# Creating a high quality Regents Test requires input from many sources

- ◆ Teachers, administrators and other school personnel review syllabus
- ◆ Statewide Committee reviews standards
- ◆ Teachers and other educators write and review test items
- ◆ NY SED staff oversees and manages the process
  - Trains item writers
  - Trains committees to review items and provide input to policy issues
  - Manages all aspects of pretesting, field testing, and analysis
  - Manages all aspects of test publication, distribution, scoring, and analysis



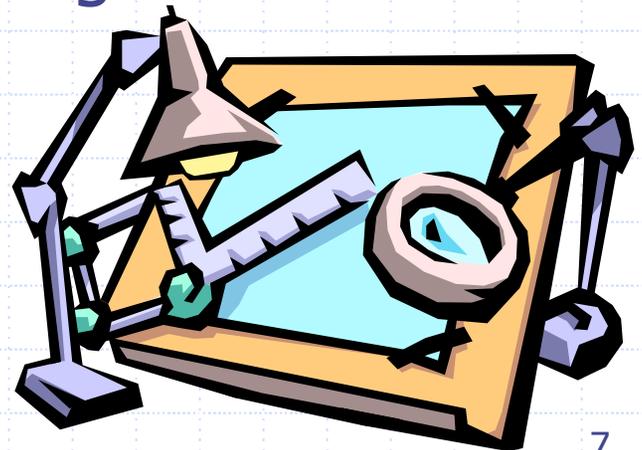
# The first step in developing a test is determining what to measure

- ◆ For existing tests, syllabi are reviewed and updated
  - Curriculum specialists and many others provide input
- ◆ For new tests, a committee creates a draft syllabus and submits it for several levels of review



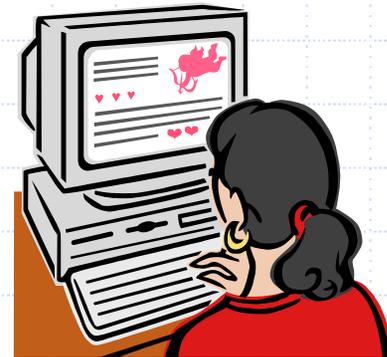
Once agreement is reached on *what* to measure, decisions must be made about *how*

- ◆ Test specifications provide the detail about what the test will look like
  - How many questions will there be?
    - ◆ Multiple choice
    - ◆ Open-ended or performance
  - What is the weighting among the various sections of the syllabus?



# When the what and how are established, it's time to create!

- ◆ NY SED staff trains teachers and other educators to write test items in a workshop setting
- ◆ Training is intensive – writing good items is a skill!
  - Item writing “rules”
  - Testwiseness
  - Sensitivity reviews

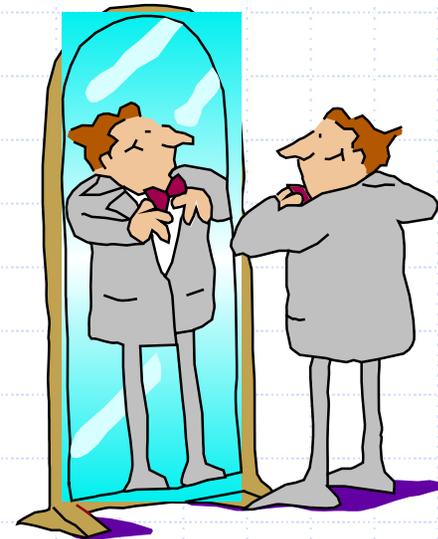


*Let's look at what it takes to write good items*



# Item writing “rules” help item writers create quality items

- ◆ Focus on must-know material, not nice to know
- ◆ Use simple, direct and clear language
- ◆ Don’t include extra information – no teaching!
- ◆ Avoid “dressing up” a stem – that is trying to make a fact-based question seem more than it is



# Some rules focus on quality from a student perspective

- ◆ Do not repeat words in the options that can go in the stem – limit the amount of reading required (unless it's a reading test)
- ◆ Make sure the options flow grammatically from the stem
- ◆ Make sure you are testing content rather than logic
- ◆ Provide enough information in the stem so the student understands the question without reading the choices!

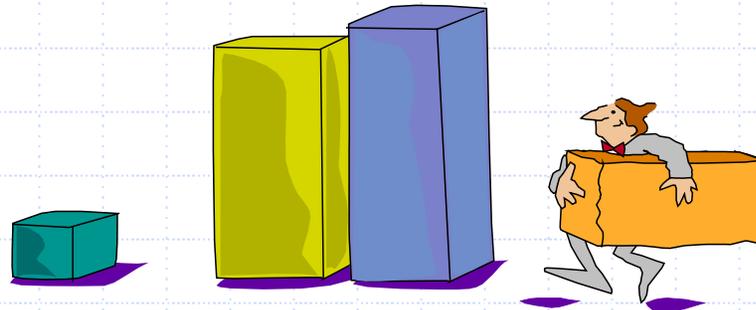


# Writing the options is the most challenging part of item writing

Have you ever taken a test... and found 2 or 3 choices that seemed correct?

The quality of the choices often determines the difficulty of an item

Item writers must carefully avoid cues to testwise students!



# Testwiseness means a student answers correctly even if they don't know the content

## ◆ Avoid ...

- words such as *all* or *never* in the choices
- choices that subsume or overlap each other
- using pairs of opposites as choices
- having one very precise or very long choice
- having key words in the stem and choices
- using items that cue each other



# “Sensitivity” in item writing means awareness and fairness

- ◆ No stereotyping based on demographic characteristics
  - Gender, socioeconomic status, ethnicity
- ◆ No use of terms or concepts that might be limited to a segment of the student population
- ◆ No content that will advantage or disadvantage any subgroup
- ◆ A sensitivity review is not “censorship”



# But writing the item is just the beginning...

- ◆ No item can appear even in a pretest until it receives extensive reviews
  - Content
  - Technical Accuracy
  - Fairness
  - Linkage to test specifications
  - Clear and concise writing



# Several independent reviews of the items helps ensure quality

- ◆ NY SED appoints committees for each Regents Test
- ◆ Committees meet and work together to review and revise items
  - Ensure item writing rules are followed
  - Ensure each item maps to test specifications



# Review committees also provide other important input

- ◆ In addition to reviewing and revising all test items, committee members also do the following
  - Advise NY SED on policy and logistics for their test
  - Advise on special populations
  - Provide input for communications about the tests



# Pretesting weeds out potentially flawed items

- ◆ Districts and schools are carefully selected to be representative of the population
  - But when there's a high refusal rate, the participants may not be as representative of the student population as desired
- ◆ Pretests take a short amount of classroom time – but yield great benefits toward creating quality tests



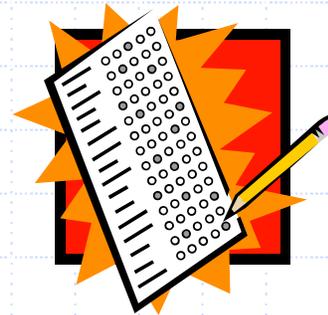
# Pretest data is carefully analyzed and items evaluated

- ◆ Multiple choice and performance items are scored
- ◆ For performance items, readers are trained, raterfinders are identified, and the pretests are scored
- ◆ Several statistical analyses are performed
  - Reliability, interrater reliability, difficulty, discrimination



# Items that meet the quality and psychometric standards are placed into field tests

- ◆ Any item that requires substantial revision based on pretest data is revised and pretested again
- ◆ Only quality items are put into the field test forms
- ◆ Field tests take only full period minutes



# Field tests focus on trying out prototype tests

- ◆ The design is carefully designed to the representative of the student population
- ◆ Test booklets are printed and distributed
- ◆ Schools administer the field tests
- ◆ Field tests are scored



# A series of statistical analyses is conducted on field test data

- ◆ Reliability/generalizability
- ◆ Item difficulty and discrimination
- ◆ Calibrate new items
- ◆ Describe performance
- ◆ Assess potential bias



Results are submitted to the Statewide Review Committee



# The final step is to set standards

- ◆ A performance standard addresses the question: How much is enough?
  - Group-based standards
  - Criterion-based standards
- ◆ Methodology is selected and implemented
  - Levels of performance are operationalized
  - Combination of professional judgements and empirical data



# Throughout the development process, communication is critical

- ◆ Prototype items and generic rubrics provide information about the test
  - Help teachers and students understand what to expect
- ◆ Syllabi and information about the tests are available
- ◆ A variety of communication vehicles are used



# The end result is a high-quality Regents Test

*Technically sound  
Evidence of validity and reliability*

*Field tested  
Data from New York state students*

*Input from many sources  
Teachers, administrators, and other  
educators serve on the review  
committees*

