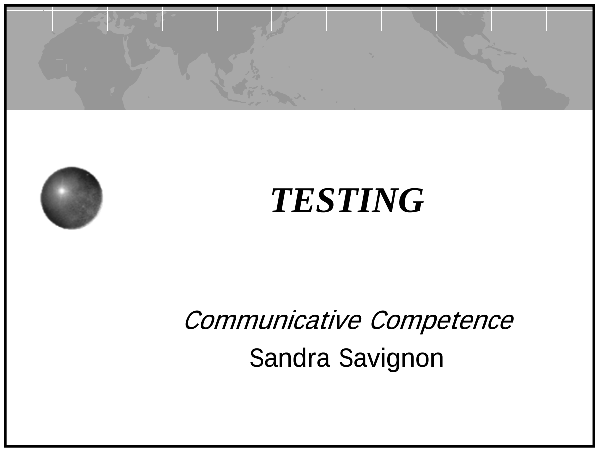
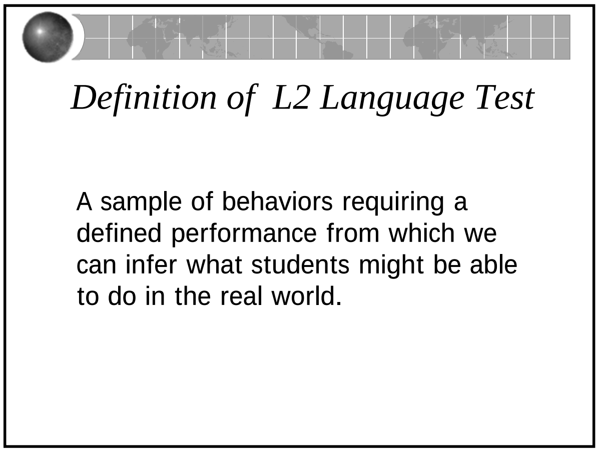


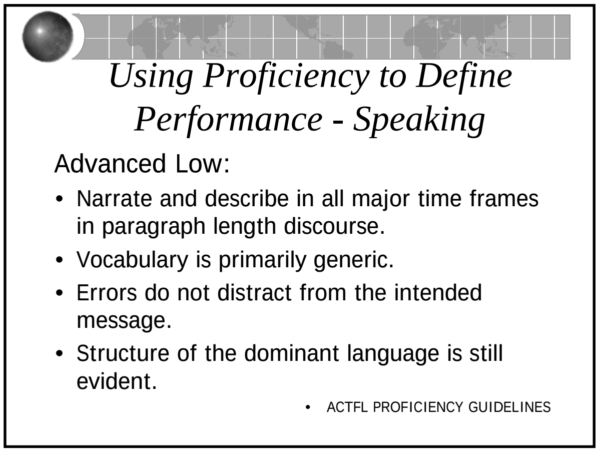
TESTING

Communicative Competence
Sandra Savignon



Definition of L2 Language Test

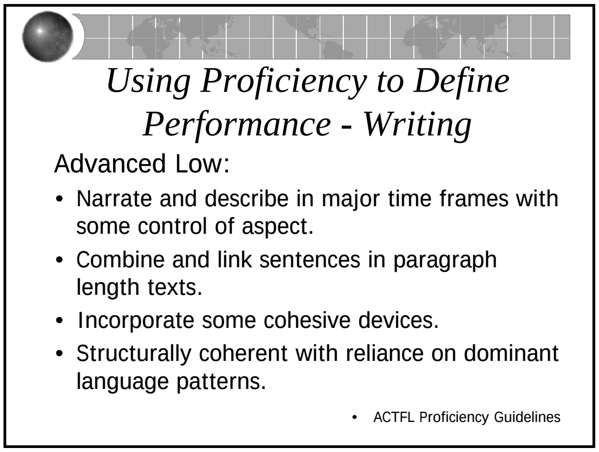
A sample of behaviors requiring a defined performance from which we can infer what students might be able to do in the real world.



Using Proficiency to Define Performance - Speaking

Advanced Low:

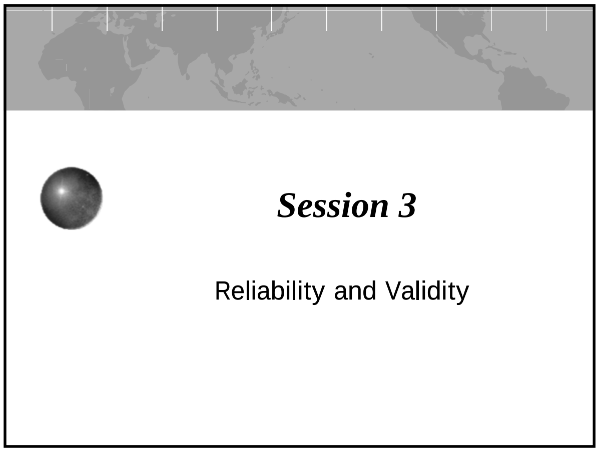
- Narrate and describe in all major time frames in paragraph length discourse.
- Vocabulary is primarily generic.
- Errors do not distract from the intended message.
- Structure of the dominant language is still evident.
  - ACTFL PROFICIENCY GUIDELINES



Using Proficiency to Define Performance - Writing

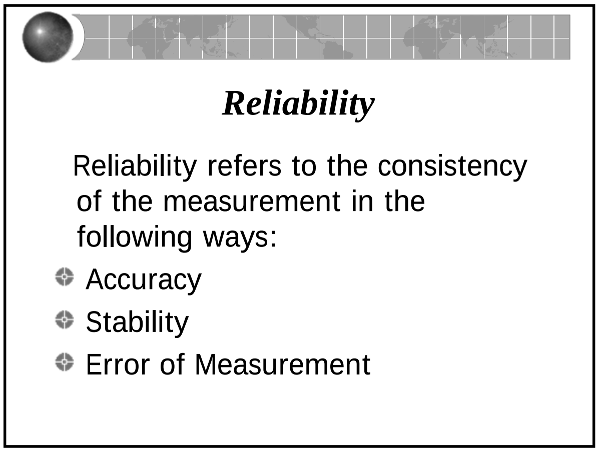
Advanced Low:

- Narrate and describe in major time frames with some control of aspect.
- Combine and link sentences in paragraph length texts.
- Incorporate some cohesive devices.
- Structurally coherent with reliance on dominant language patterns.
  - ACTFL Proficiency Guidelines



Session 3

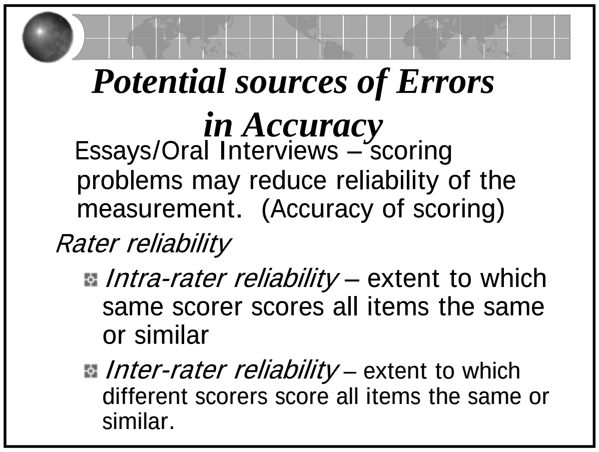
Reliability and Validity



Reliability

Reliability refers to the consistency of the measurement in the following ways:

- Accuracy
- Stability
- Error of Measurement

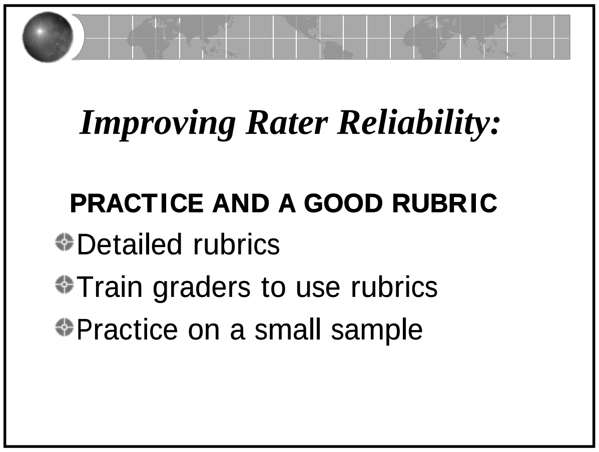


Potential sources of Errors in Accuracy

Essays/Oral Interviews – scoring problems may reduce reliability of the measurement. (Accuracy of scoring)

Rater reliability

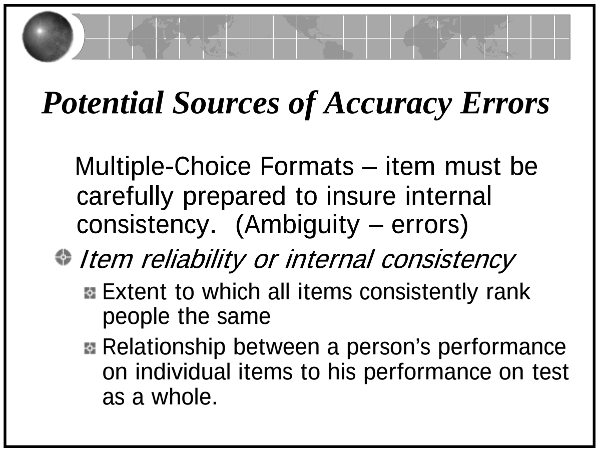
- *Intra-rater reliability* – extent to which same scorer scores all items the same or similar
- *Inter-rater reliability* – extent to which different scorers score all items the same or similar.



Improving Rater Reliability:

PRACTICE AND A GOOD RUBRIC

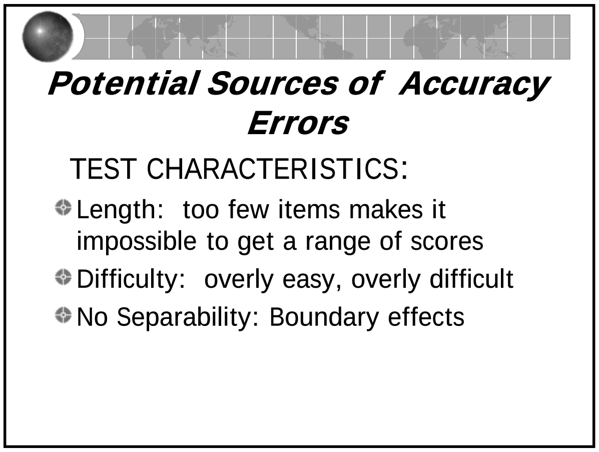
- Detailed rubrics
- Train graders to use rubrics
- Practice on a small sample



Potential Sources of Accuracy Errors

Multiple-Choice Formats – item must be carefully prepared to insure internal consistency. (Ambiguity – errors)

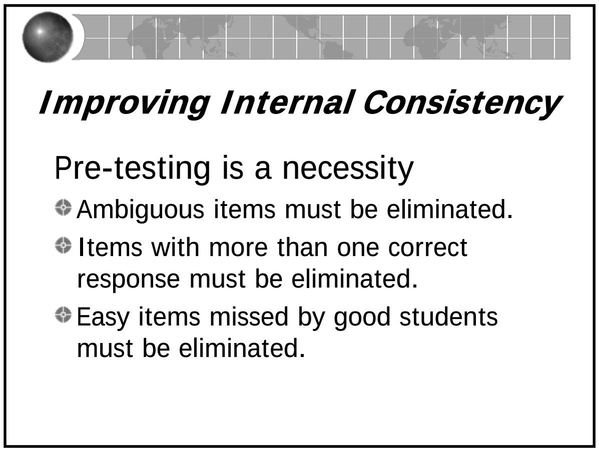
- *Item reliability or internal consistency*
  - Extent to which all items consistently rank people the same
  - Relationship between a person's performance on individual items to his performance on test as a whole.



Potential Sources of Accuracy Errors

TEST CHARACTERISTICS:

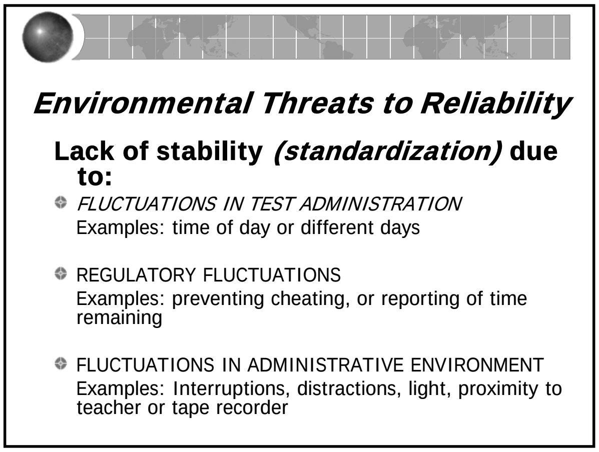
- Length: too few items makes it impossible to get a range of scores
- Difficulty: overly easy, overly difficult
- No Separability: Boundary effects



Improving Internal Consistency

Pre-testing is a necessity

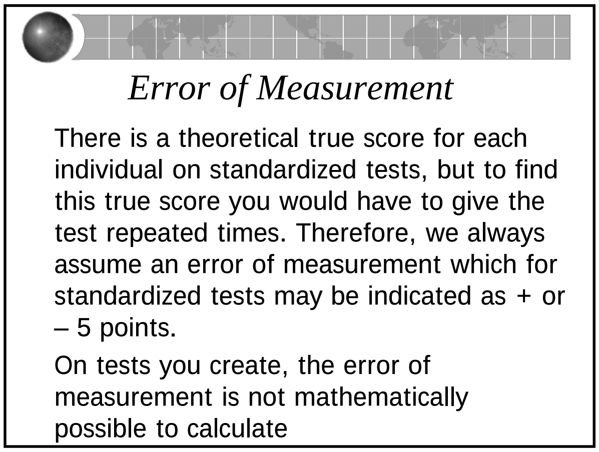
- Ambiguous items must be eliminated.
- Items with more than one correct response must be eliminated.
- Easy items missed by good students must be eliminated.



Environmental Threats to Reliability

Lack of stability *(standardization)* due to:

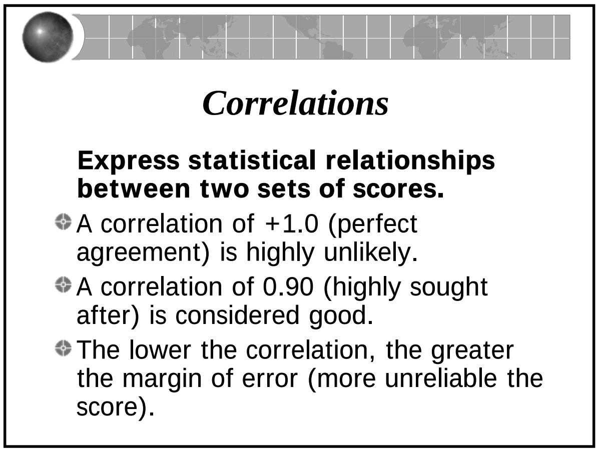
- *FLUCTUATIONS IN TEST ADMINISTRATION*
  Examples: time of day or different days
- REGULATORY FLUCTUATIONS
  Examples: preventing cheating, or reporting of time remaining
- FLUCTUATIONS IN ADMINISTRATIVE ENVIRONMENT
  Examples: Interruptions, distractions, light, proximity to teacher or tape recorder



Error of Measurement

There is a theoretical true score for each individual on standardized tests, but to find this true score you would have to give the test repeated times. Therefore, we always assume an error of measurement which for standardized tests may be indicated as + or – 5 points.

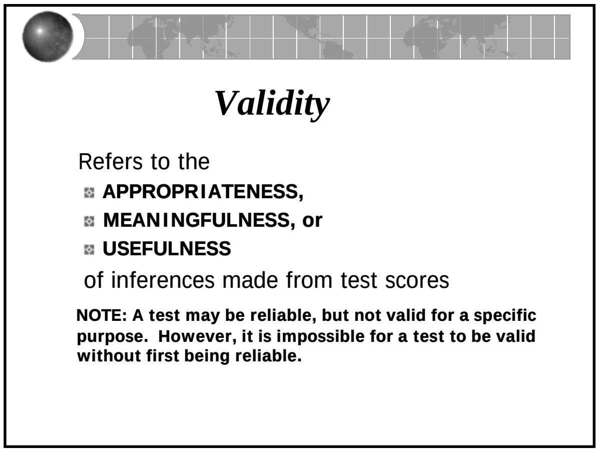
On tests you create, the error of measurement is not mathematically possible to calculate



Correlations

Express statistical relationships between two sets of scores.

- A correlation of +1.0 (perfect agreement) is highly unlikely.
- A correlation of 0.90 (highly sought after) is considered good.
- The lower the correlation, the greater the margin of error (more unreliable the score).



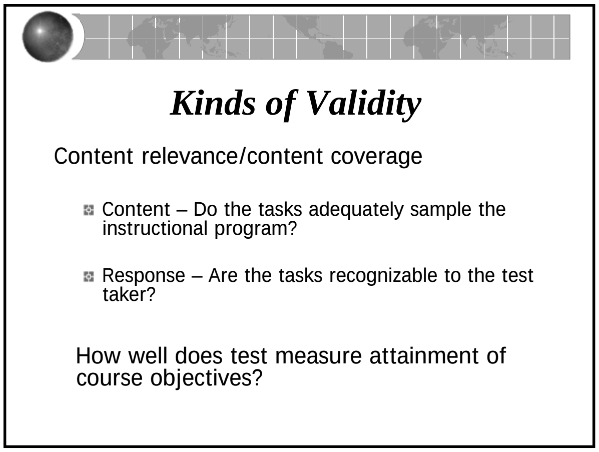
Validity

Refers to the

- **APPROPRIATENESS,**
- **MEANINGFULNESS, or**
- **USEFULNESS**

of inferences made from test scores

NOTE: A test may be reliable, but not valid for a specific purpose. However, it is impossible for a test to be valid without first being reliable.

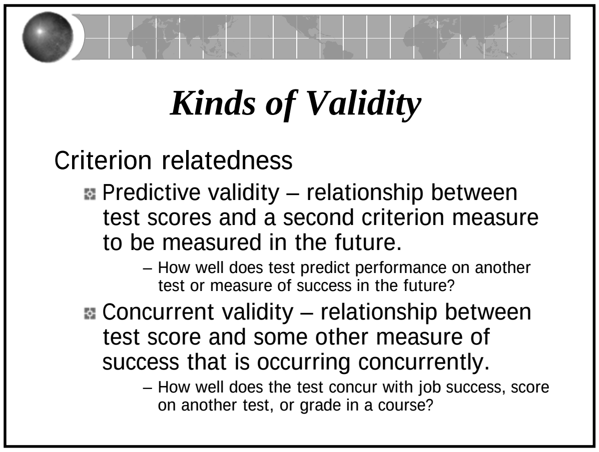


Kinds of Validity

Content relevance/content coverage

- Content – Do the tasks adequately sample the instructional program?
- Response – Are the tasks recognizable to the test taker?

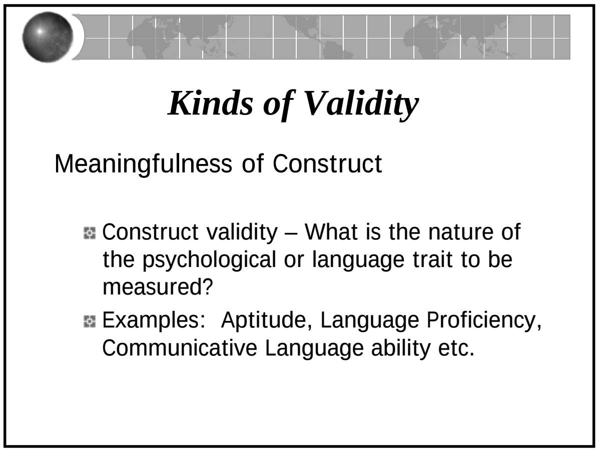
How well does test measure attainment of course objectives?



Kinds of Validity

Criterion relatedness

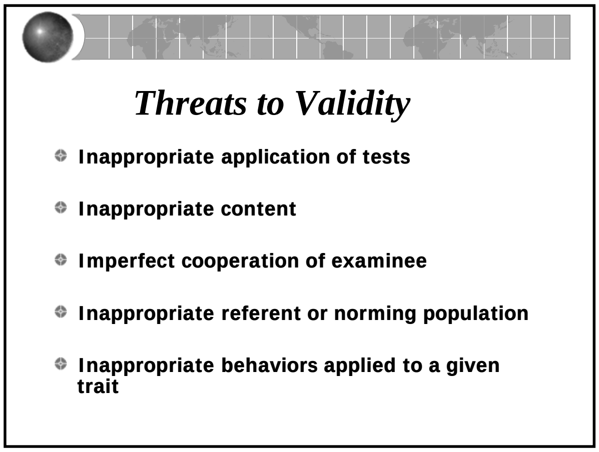
- Predictive validity – relationship between test scores and a second criterion measure to be measured in the future.
  - How well does test predict performance on another test or measure of success in the future?
- Concurrent validity – relationship between test score and some other measure of success that is occurring concurrently.
  - How well does the test concur with job success, score on another test, or grade in a course?



Kinds of Validity

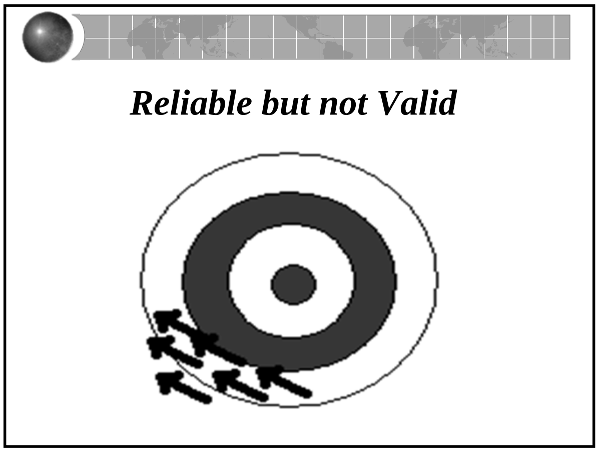
Meaningfulness of Construct

- Construct validity – What is the nature of the psychological or language trait to be measured?
- Examples: Aptitude, Language Proficiency, Communicative Language ability etc.

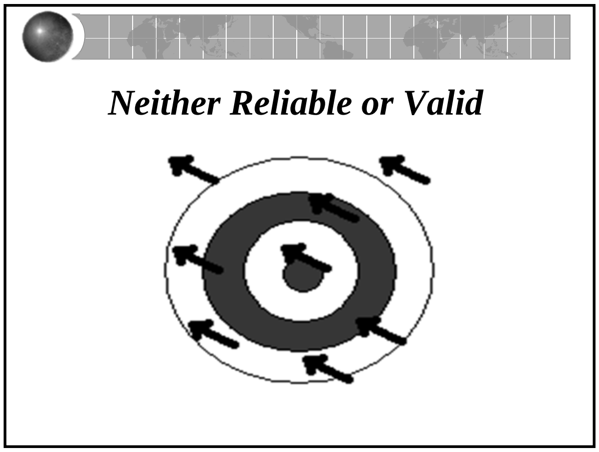


Threats to Validity

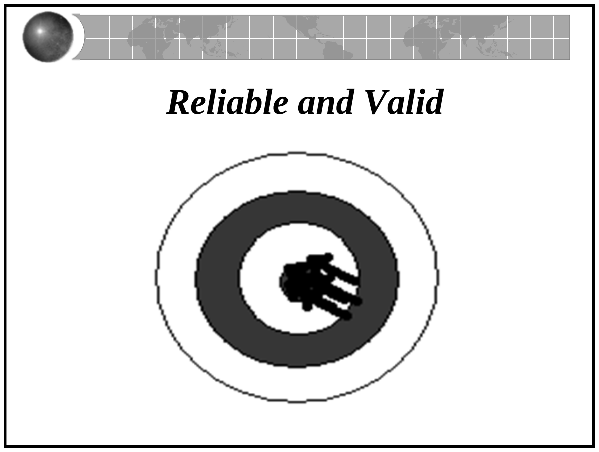
- Inappropriate application of tests
- Inappropriate content
- Imperfect cooperation of examinee
- Inappropriate referent or norming population
- Inappropriate behaviors applied to a given trait



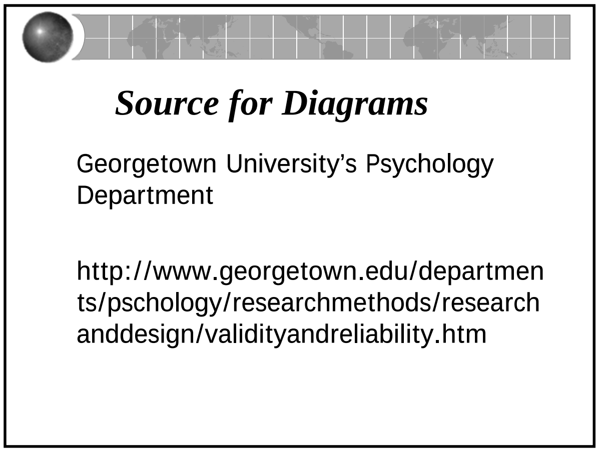
Reliable but not Valid



Neither Reliable or Valid



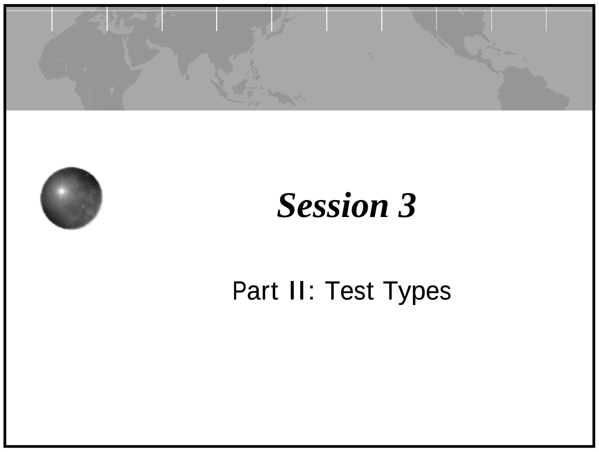
Reliable and Valid



Source for Diagrams

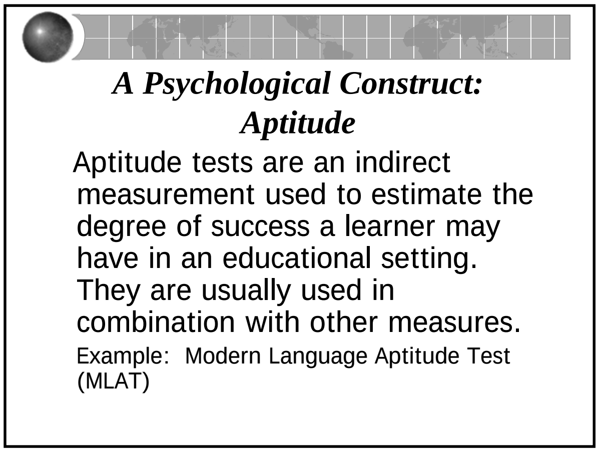
Georgetown University’s Psychology Department

http://www.georgetown.edu/departments/pschology/researchmethods/researchanddesign/validityandreliability.htm



Session 3

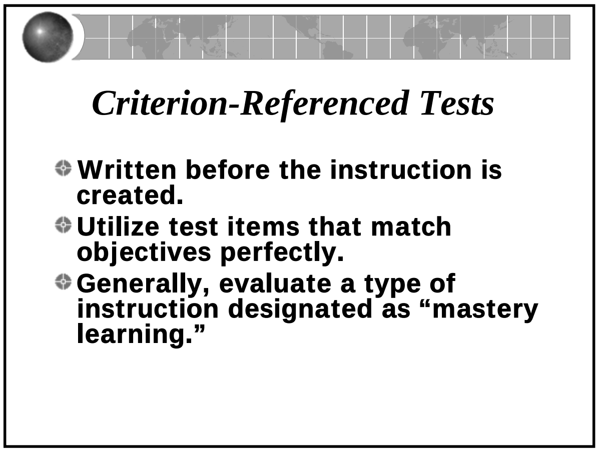
Part II: Test Types



A Psychological Construct: Aptitude

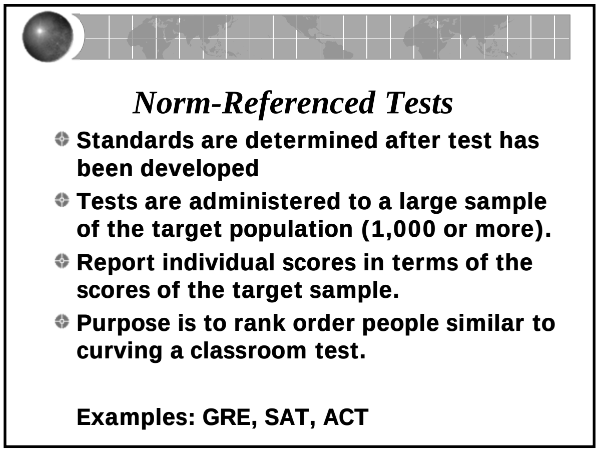
Aptitude tests are an indirect measurement used to estimate the degree of success a learner may have in an educational setting. They are usually used in combination with other measures.

Example: Modern Language Aptitude Test (MLAT)



Criterion-Referenced Tests

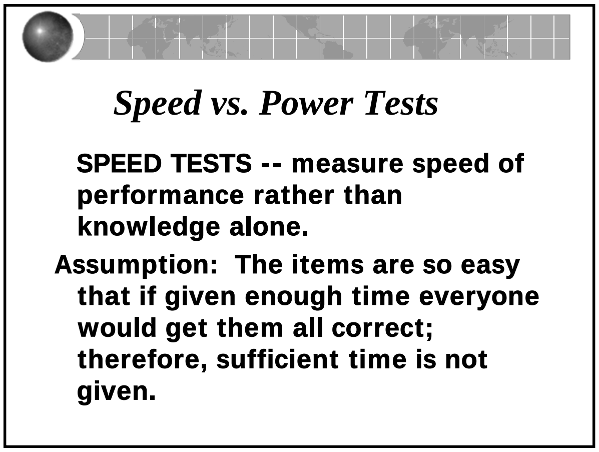
- **Written before the instruction is created.**
- **Utilize test items that match objectives perfectly.**
- **Generally, evaluate a type of instruction designated as "mastery learning."**



Norm-Referenced Tests

- **Standards are determined after test has been developed**
- **Tests are administered to a large sample of the target population (1,000 or more).**
- **Report individual scores in terms of the scores of the target sample.**
- **Purpose is to rank order people similar to curving a classroom test.**

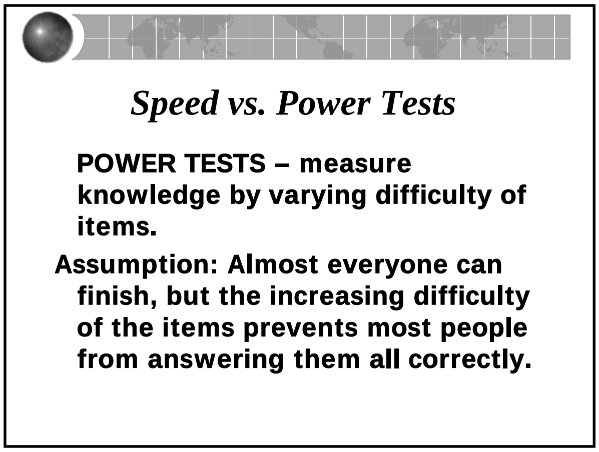
Examples: GRE, SAT, ACT



Speed vs. Power Tests

SPEED TESTS -- measure speed of performance rather than knowledge alone.

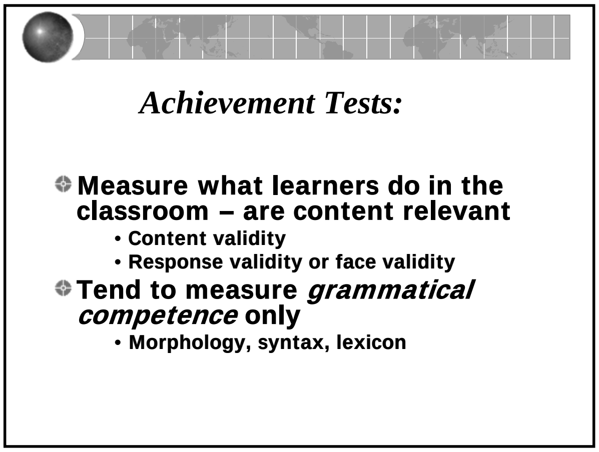
Assumption: The items are so easy that if given enough time everyone would get them all correct; therefore, sufficient time is not given.



Speed vs. Power Tests

POWER TESTS – measure knowledge by varying difficulty of items.

Assumption: Almost everyone can finish, but the increasing difficulty of the items prevents most people from answering them all correctly.



Achievement Tests:

- **Measure what learners do in the classroom – are content relevant**
  - **Content validity**
  - **Response validity or face validity**
- **Tend to measure *grammatical competence* only**
  - **Morphology, syntax, lexicon**

Proficiency Tests:

Measure an individual's general competence in L2, independently of any particular curriculum.

Appropriate for measuring general progress in acquisition of an individual's L2.

Pro-Achievement

- Achievment tests should reflect nature of proficiency or competence toward which learners are supposed to be advancing.
  - Savignon, p. 223

Importance of Classroom Tests

1. Tests function as a measurement for learner progress.
2. Tests are a powerful motivating factor.
3. Tests should tell us what learners can really do with language if communicative competence is our goal.
4. Tests of communicative competence are best assurance that we are preparing learners for the real world.

What is Communication?

"*Communication is expression, interpretation, and negotiation of meaning.*"

"Communicative competence is always context specific, requiring simultaneous, integrated use of grammatical competence, discourse competence, sociolinguistic competence and strategic competence"

Savignon, 1997, p. 225

Direct/Natural/Situational Tests

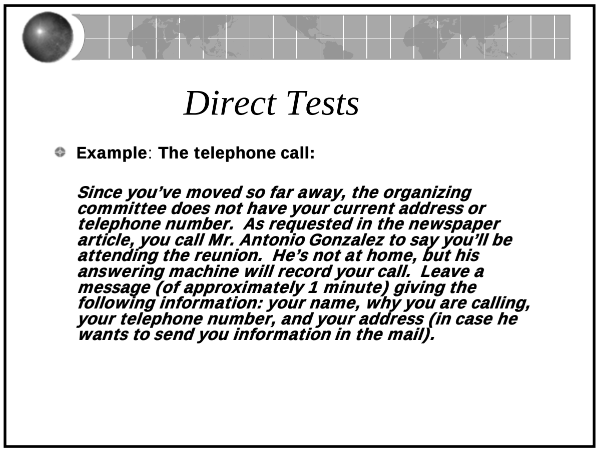
DIRECT TESTS – provide problems and solutions students will find in real life; that is, the format of the testing situation is real-world.

AUTHENTIC ASSESSMENT – rates language use in real, uncontrived communicative situations

Indirect/Unnatural/Contrived Tests

INDIRECT TESTS – represent competence by measuring knowledge and skills outside of their real-world contexts. They are distinguished by ease of grading and the contrived nature of the test.

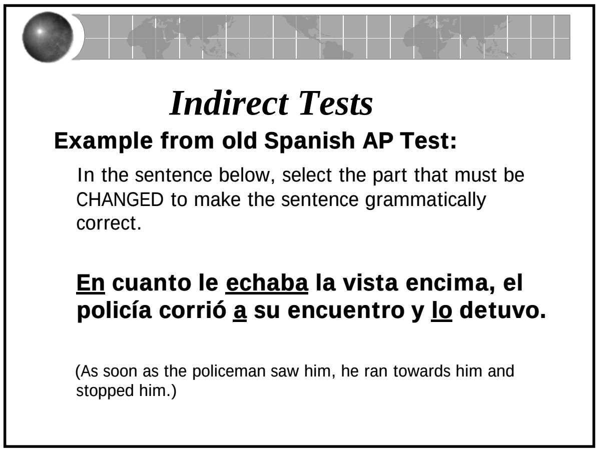
CONTRIVED TESTS – utilize multiple choice, binary, single word response measure language proficiency indirectly.



Direct Tests

- **Example**: **The telephone call:**

Since you've moved so far away, the organizing committee does not have your current address or telephone number. As requested in the newspaper article, you call Mr. Antonio Gonzalez to say you'll be attending the reunion. He's not at home, but his answering machine will record your call. Leave a message (of approximately 1 minute) giving the following information: your name, why you are calling, your telephone number, and your address (in case he wants to send you information in the mail).



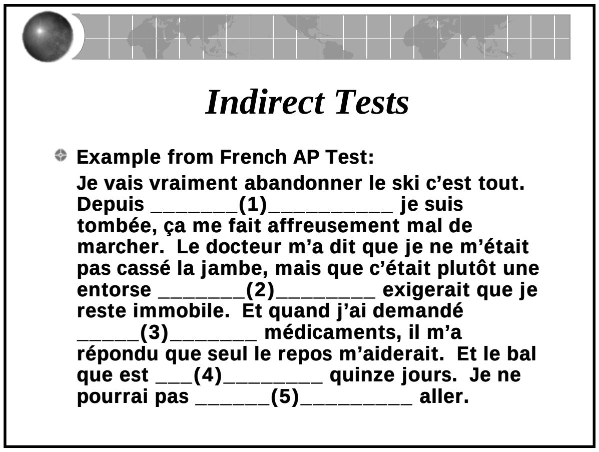
Indirect Tests

Example from old Spanish AP Test:

In the sentence below, select the part that must be CHANGED to make the sentence grammatically correct.

En cuanto le echaba la vista encima, el policía corrió a su encuentro y lo detuvo.

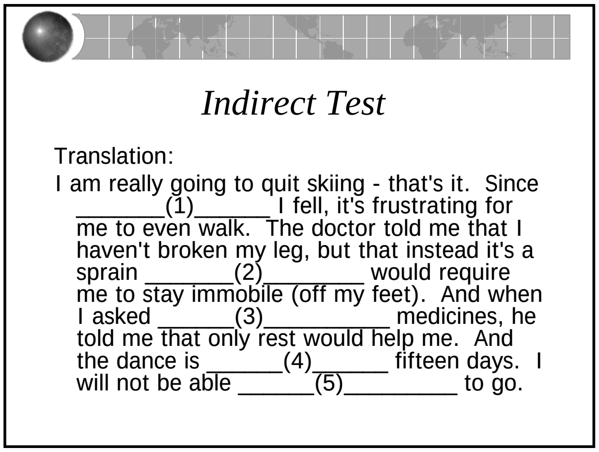
(As soon as the policeman saw him, he ran towards him and stopped him.)



Indirect Tests

- **Example from French AP Test:**

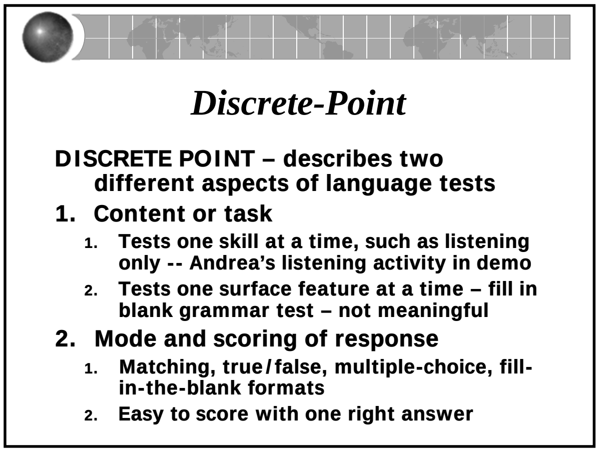
Je vais vraiment abandonner le ski c'est tout. Depuis _______(1)__________ je suis tombée, ça me fait affreusement mal de marcher. Le docteur m'a dit que je ne m'était pas cassé la jambe, mais que c'était plutôt une entorse _______(2)________ exigerait que je reste immobile. Et quand j'ai demandé _____(3)_______ médicaments, il m'a répondu que seul le repos m'aiderait. Et le bal que est ___(4)________ quinze jours. Je ne pourrai pas ______(5)_________ aller.



Indirect Test

Translation:

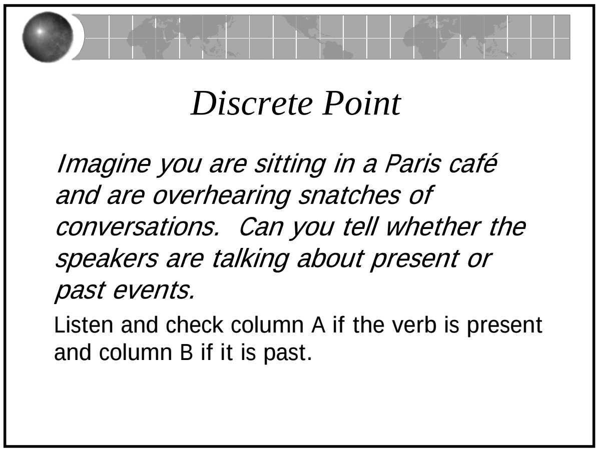
I am really going to quit skiing - that's it. Since _______(1)______ I fell, it's frustrating for me to even walk. The doctor told me that I haven't broken my leg, but that instead it's a sprain _______(2)________ would require me to stay immobile (off my feet). And when I asked ______(3)__________ medicines, he told me that only rest would help me. And the dance is ______(4)______ fifteen days. I will not be able ______(5)_________ to go.



Discrete-Point

DISCRETE POINT – describes two different aspects of language tests

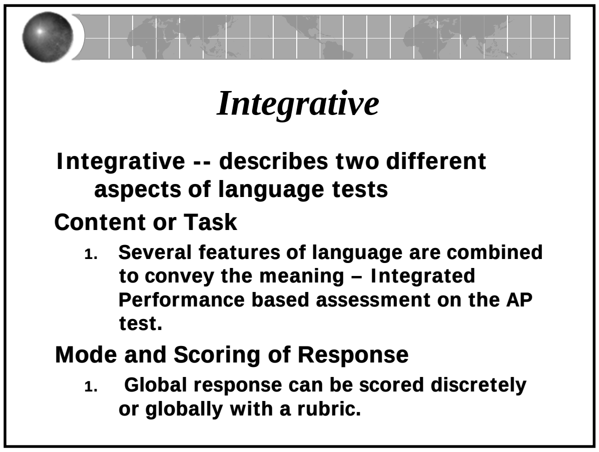
1. **Content or task**
   1. **Tests one skill at a time, such as listening only -- Andrea's listening activity in demo**
   2. **Tests one surface feature at a time – fill in blank grammar test – not meaningful**
2. **Mode and scoring of response**
   1. **Matching, true/false, multiple-choice, fill-in-the-blank formats**
   2. **Easy to score with one right answer**



Discrete Point

Imagine you are sitting in a Paris café and are overhearing snatches of conversations. Can you tell whether the speakers are talking about present or past events.

Listen and check column A if the verb is present and column B if it is past.



Integrative

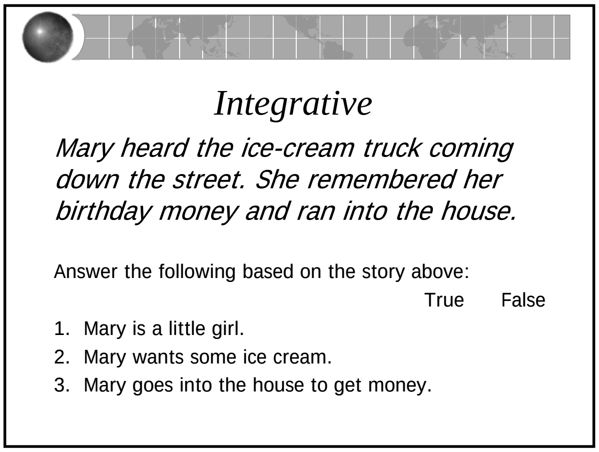
Integrative -- describes two different aspects of language tests

Content or Task

1. **Several features of language are combined to convey the meaning – Integrated Performance based assessment on the AP test.**

Mode and Scoring of Response

1. **Global response can be scored discretely or globally with a rubric.**

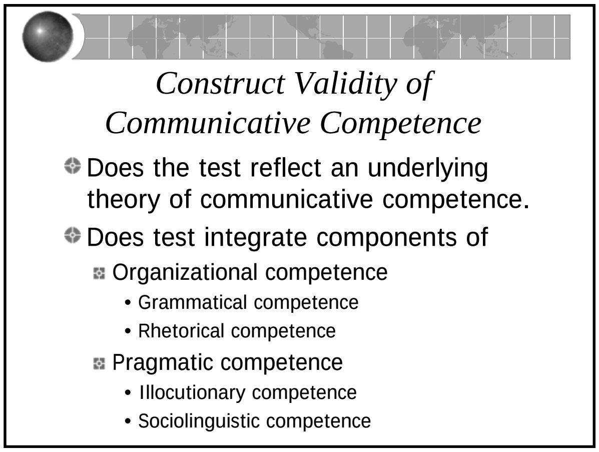


Integrative

Mary heard the ice-cream truck coming down the street. She remembered her birthday money and ran into the house.

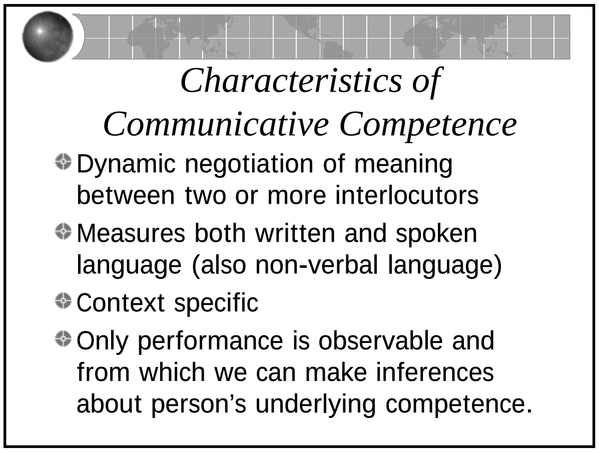
Answer the following based on the story above:

	True	False
1. Mary is a little girl.		
2. Mary wants some ice cream.		
3. Mary goes into the house to get money.		



Construct Validity of Communicative Competence

- Does the test reflect an underlying theory of communicative competence.
- Does test integrate components of
  - Organizational competence
    - Grammatical competence
    - Rhetorical competence
  - Pragmatic competence
    - Illocutionary competence
    - Sociolinguistic competence



Characteristics of Communicative Competence

- Dynamic negotiation of meaning between two or more interlocutors
- Measures both written and spoken language (also non-verbal language)
- Context specific
- Only performance is observable and from which we can make inferences about person’s underlying competence.