

Test development

A close-up photograph of a tortoise shell, showing the intricate patterns and colors of the scutes. The shell is dark brown with lighter, yellowish-brown edges. The text is overlaid on the shell, with the title 'Test development' at the top and a list of stages below it.

Scope of the presentation

- ✓ Stages in test development
- ✓ Stage 1: Determining test type and test format
- ✓ Stage 2: Planning
- ✓ Stage 3: Item writing
- ✓ Stage 4: reviewing
- ✓ Stage 5: Piloting
- ✓ Stage 6: Validating

Stage 1: Determining test type and test format

- In advance determination of function because of its influence on format and content
- Three factors in determining the function of the test
 1. The characteristics of examinees
 2. The specific purpose of the test
 3. The scope of the test

Stage 2: Planning

- ❑ Content is heavily influenced by the function and form of the test.
- ❑ Three steps in planning:
 1. Determining the instructional objectives
 2. Dividing objectives into manageable topics
 3. Preparing a table of specifications
 - a. Provision of a representative sample of the materials

An example of table of specifications

Instructional objectives	Test format	Test type	Number of items
a. the ability to pronounce words correctly			
b. the ability to know and write the phonetic symbols for words			
c. the ability to utter sentences correctly			
d. the ability to place stress on the correct syllables			
contents			
1. Stress patterns of words (primary stress)			10
a. two syllable words	True or false	recognition	3
b. three syllable words	True or false	recognition	4
c. beyond three syllable words	Multiple-choice	recognition	3
2. Intonation of sentences			8
a. direct questions	True or false	recognition	2
b. indirect questions	True or false	recognition	2
c. statements	True or false	recognition	2
d. imperatives	True or false	Recognition	2
3. Pronunciation of words			30
a. individual words	Read aloud	production	10
b. a set of words	Read aloud	production	5
c. words and phrases in sentences	Read aloud	production	15
4. Writing phonetic symbols of words			15
a. easy words	completion	production	7
b. difficult words	completion	production	8

Stage 3: Item writing

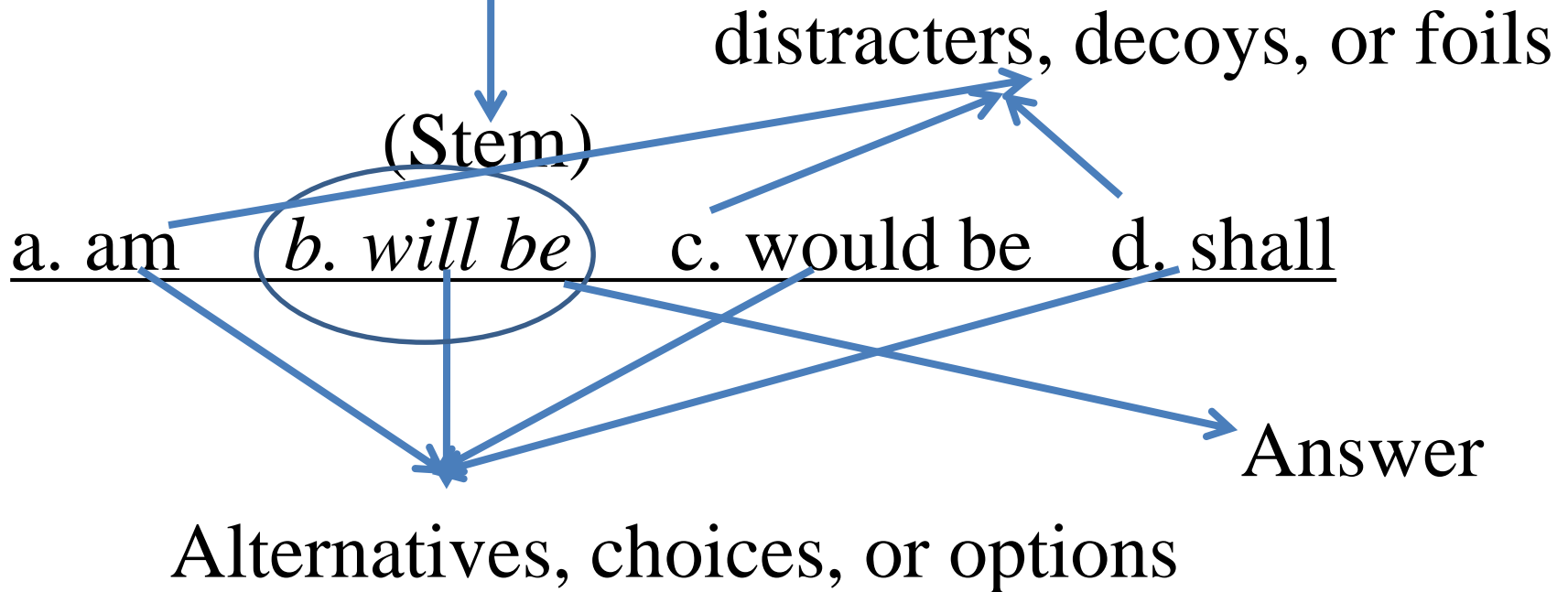
- ❖ Nature of items
- ❖ Subjective in nature
- ❖ Guidelines for writing various item types
 1. Productive response items
 - a. Short answers or short response items
 - b. Fill in items or completion items
 2. Receptive response items
 - a. Matching items
 - b. True or false items
 - c. Multiple choice items
 3. Task items

General guidelines for all item types

1. The item format should be correctly matched to the purpose and content of the items.
2. There should be only one correct answer.
3. The item should be written at the students level of proficiency.
4. Negatives and double negatives should be avoided.
5. Items should avoid clues used in answering other items.
6. All parts of the item should be on the same page.
7. Only relevant information should be tested.
8. Items should be inspected by at least one colleague.

Receptive response items: Multiple-choice items

Don't call me at 4 o'clock tomorrow afternoon. I
..... giving my students a midterm exam.



1. Multiple-choice questions can be presented in either questions or statements.

A. Which of the following is closest in meaning to “abandon”?

- a. molest b. quit c. reveal d. merge

B. The word “abandon” is closest in meaning to
.....

- a. molest b. quit c. reveal d. merge

2. The stem of the item should be meaningful by itself and present a definite problem.

Poor: The blind man

- a. Was left alone in the street
- b. Had no money with him
- c. Used his stick to walk around
- d. Slipped over the ice and fell down

Better: We helped the blind man to the street.

- a. cross
- b. catch
- c. walk
- d. stop

3. Repeated words in the choices should be included in the stem.

Poor: Don't call me at 4 o'clock tomorrow afternoon. I my students a midterm exam.

- a. Am giving b. will be giving c. would be giving d. should be giving

4. Negative words should be avoided as far as possible.

Poor: Which of the following countries has not just been affected by debt crisis in Europe?

- a. Greece b. Italy c. Cyprus d. Spain

Better: All of the following countries have already been struck by debt crisis except

- a. Greece b. Italy c. Cyprus d. Spain

5. All the alternatives should be grammatically consistent with the stem and correct by themselves.

Poor: She can not stand the party

a. Because of the noise

b. It was overcrowded

c. To keep staying in

d. Were you step in

e. Better: she can not stand the party at the moment because

f. It is very noisy

g. There were many people in

h. The guests will be leaving it

i. She may have been treated badly.

Poor: She you with your math problems next time.

a. is going to help

b. will helping

c. helping

d. will have helped

Better: She you with your math problems next time.

a. is going to be helped

b. will help

c. helping

d. will have helped

➤ There must be one correct or clearly one best answer.

Poor: I will send your regards to my parents when I home.

- a. Arrive
- b. Have arrived
- c. Will arrive
- d. Arrived

Better: I will send your regards to my parents when I Home.

- a. Arrive b. am arriving c. will arrive d. arrived

➤ All distracters should be plausible.

Poor: We had to a scheme to enter the bank.

a. devise b. device c. nervous d. secretly

Better: We had to a scheme to enter the bank.

a. devise b. struggle c. rotate d. orbit

➤ Alternatives should be of similar length and difficulty.

Poor: The taxi driver rushed out of the house

- a. for calling an ambulance
- b. To call an ambulance
- c. He was petrified by a rat
- d. Were he to desire to rescue the afflicted landlord

Better: The taxi driver rushed out of the house
because the landlady was seriously ill.

- a. To ask for help
- b. To hide himself
- c. To cry for her
- d. To stop her death

➤ The stem should be worded as simply as possible.

Poor: We were furious at him, and made up our minds not to let her in because he was for the meeting.

a. absent b. late c. sick d. ready

Better: We were angry because he was for the meeting again.

a. absent b. late c. sick d. ready

❖ All the clues in the stem should be avoided.

Poor: We noticed abreak into your house.

a. thief b. earl c. heir d. entrant

Better:

We noticed a (n)break into your house.

a. thief b. earl c. heir d. entrant

❑ There should be only one blank in each stem.

Poor: The fog with the narrow road
made driving difficult.

- a. heavy-concerned b. thick-coupled
c. dense-nested d. sparse-aligned

Better: The thick fog with the narrow road
made driving difficult.

- a. coupled b. aligned c. concerned d. nested

✓ The stem should not start with a blank.

Poor: are classified as cold-blooded animals.

- a. cows b. snakes c. sheep d. ducks

Better: Snakes are one type of animals.

- a. warm-blooded b. cold-blooded
c. amphibious d. crustacean

- The correct answer should appear in each of the alternative positions an approximately equal number of times but in random order.
- Alternatives could range from three to five.
- Make the length of blanks equal in all stems.

Avoid review options

- Special alternatives such as all of the above, none of the above, both a and b, nether c and d should be used sparingly.

Poor: Like a topic sentence, a reworded thesis statement expresses the same meaning in different words and structures, but, unlike a topic sentence, a reworded thesis statement

- a. must be a statement
- b. puts forward two sides of the same argument
- c. expresses the main idea of the whole essay
- d. both b and c

Can you identify two problems with this item?

Avoid hybrid or impure items

Poor: the guests the hostess for her last night's entertainment.

- a. Thank b. thanked c. praise d. praised

Better:

The guests the hostess for her last night's entertainment.

- a. Thanked b. thank c. have thanked d. will thank

Poor: As a vampire, he the city every two year.

- a. visits b. is visiting c. has visited d. visit

Better: As a vampire, he the city every two year.

- a. Visits b. is visiting c. has visited c. was visiting

➤ Arrangement of the alternatives should be either vertical or horizontal.

Pattern A: vertical

a.

b.

c.

d.

Pattern B: horizontal

a.

b.

c.

d.

Pattern C: a mixture of both (not recommended)

a.

b.

c.

d.

Some guidelines for selection of plausible distracters

- A. Errors students make in free writing or speaking
- B. Students' answers in previous tests
- C. Teacher's experience
- D. Contrastive analysis between native language and mother tongue
- E. The erroneous responses in a cloze passage

Advantages and disadvantages of multiple-choice items

I. Advantages

- A. Effective measurement of simple learning outcomes
- B. Easy to score
- C. Applicable to a wide variety of skills

II. Disadvantages

- A. Test of recognition knowledge
- B. The problem of guessing on test scores
- C. Restriction of what to test
- D. Difficulty of writing effective items
- E. Facilitation of cheating and possibility of harmful backwash

True or false items (alternative response items)

I. Broad general statements should be avoided.

Poor: students are always given a placement test to be grouped into various levels. (F)

Poor: students are usually given a placement test to be grouped into various levels. (T)

Usually, generally, often, and sometimes make a statement true.

Always, never, absolutely, all, and only make a statement false.

II. Trivial points should be avoided in statements.

Poor: The colour of the actress's dress in the movie was blue.

Poor: The tree had many braches.

III. Negative statements and double negatives should be avoided.

Poor: None of the poets mentioned in the passage were unknown.

Better: all the poets mentioned in the passage were known.

IV. Including two ideas in a statement should be avoided.

Poor: The stress on the word “dictate” falls on the second syllable because it is a disyllabic word.

Better:

1. The stress on the word “dictate” falls on the second syllable.
2. The word “dictate” is a disyllabic word.

V. long, complex statements should be avoided.

THE longer the euro area's debt crisis drags on, the more it resembles an instrument of economic torture. Like the medieval rack, every turn of the crisis tears Europe further apart. This week Cyprus announced it would seek a bail-out. Spain formally asked for money to recapitalise its banks. The Greek limb is close to being ripped off. How long can the Italian one hold?

Poor: The approximately endemic paralysis of the euro area has just contaminated Cyprus.

Better: Cyprus has just been affected by debt crisis.

VI. Statements should be approximately equal in length and distribution.

1. The author does not know how long the Italian economy could keep moving forward.
2. Greece is still intact.
3. The economic situation in Europe is showing signs of improvement.
4. Debt crisis is yet to affect Cyprus.

Advantages and disadvantages of true or false items

I. Advantages

- A. Efficacy of true or false item
- B. Easy to construct and to score
- C. Inclusion of many true or false items in a given test

II. Disadvantages

- A. Based on chance
- B. Measurement of simple language learning outcomes
- C. Test of recognition knowledge

Matching items

- The two columns in a matching item should be related to one central theme.

What is wrong with the following item?

1. Invented telescope
 2. Was the first Iranian woman sent to space
 3. Discovered penicillin
 4. Freed slavery in America
- A. Anushe Ansarian
 - B. Abraham Lincoln
 - C. Galileo Galilei
 - D. Alexander Fleming

- An unequal number of responses should be included and the student should be told that the options could be used only once, more than once, or not at all.

1. Easily found
 2. Well-arranged
 3. Unable to decide
 4. Capable of doing
 5. Clever at doing
- A. Available
 - B. Able
 - C. Dead
 - D. Neat
 - E. Hesitated
 - F. mature

- The list of items to be matched should be brief and the shorter options should be on the right column. (between four and seven)

What is wrong with this item?

- A. Taciturn
 - B. Daybreak
 - C. Because
 - D. Inaccessible
 - E. Jury
 - F. Review
 - G. Janitor
 - H. responsible
1. On account of
 2. A person whose job is to look after a school.
 3. out of reach
 4. In charge of
 5. The time when it gets light
 6. A group of judges who listen to the details of a case in court
 7. Go over
 8. Talk little

- The list of responses should be logically ordered.
 1. The impact of a meteorite with Ann Hodges
 2. The explosion of a large object above Tunguska
 3. The smashing of Titan into Earth
 4. The burning of a 150-ton meteorite while travelling through the upper atmosphere
 - A. 1908 (2)
 - B. 1954 (1)
 - C. 1972 (4)

- The directions for matching the premises and options should be clearly indicated.
- All the items should be placed on the same page.

Advantages and disadvantages

I. Advantages

A. Easy to construct and to score

B. Measurement of facts based on simple associations

II. Disadvantages

A. The problem of homogeneity

B. Rote learning of factual knowledge

Productive response items

1. Short answers or short response items
2. Fill in items or completion items

They measure simple specific learning outcomes such as knowledge of terminology and specific facts.

Advantages: easy to construct due to recall of memorized information and elimination of guessing due to supplying the item from memory

Incapability of measuring complex points and difficulty of scoring

Guidelines to make fill-in or completion items

❖ The required answer should be both brief and specific.

1. Poor: My mother was me. (kind to/
angry at/ talking to/ advising)

Better: My mother me because I had
broken her lovely vase. (punished)

2. Poor: A word which joins two clauses is called
a (subordinator/ coordinator)

Better: A word which joins two independent
clauses is called a (coordinator)

➤ There should be enough context to convey the intent of the item to students.

Poor: She had to home very early in the morning. (leave/return/stay)

Better: She had to home very early not to be late for her work. (leave)

➤ The blanks should be of standard length.

Poor: A. Hi Ali. What did the teacher students to do as homework assignment? (ask)

B. Nothing special. Next weekly meeting, we will be the contents of previous lessons. (reviewing)

Fill-in items must be phrased in statements.

Poor: How many syllables are there in the word “composition”?

Better: The word “composition” contains syllables.

Blanks in the fill-in items should occur later in the sentence.

Poor:is a very recent electronic form of communication via which people can keep in touch miles away. (Chatting)

Better: people can keep in touch via a very recent electronic form of communication called (Chatting)

❖ Too many blanks in fill-in blanks should be avoided.

Poor: A is a place in which various types of are kept for people to (zoo/animals/visit)

Better: The place in which various types of animals are kept for people to visit is called a (zoo)

Short answer or response items

- ✚ Short answer items had better be clearly phrased in questions.

Three factors have contributed to resurgence of cities in England. The first is the investment in reclaiming city centers. Another is jobs. The third is immigration.

According to the passage, What are the three things that have resulted in the growth of cities in England?

✚ Short answer items should be written so that the answer appears to the right of the question.

Poor: What is happening to the cities in England?

..... (They are growing again)

Better: What is happening to the cities in England? (They are growing again)

- ✚ Like fill-in items, short answer items should be formatted so that only one relatively concise answer is possible.

Reza is an English teacher. He teaches Elementary school students English. He has been teaching in this school since I was a school kid. He was born in London, England in 1990 about twenty years ago, but he was raised in Karaj, Iran.

Poor questions:

1. What does he do? He is an English teacher. He teaches English.
2. When was he born? He was born in 1990. He was born twenty years ago.

Better questions

1. What does he do exactly? He teaches English.
2. In which year was he born? In 1990

Task items

Tasks should be clearly defined.

An example of a well-defined written task:

Directions: Write a five-paragraph essay on the following topic, with your essay ranging in length between 500 and 700 words. You have two hours time to finish writing the essay in two hours. Your essay will be rated based on the following criteria:

1. Content: logical development of ideas
2. Grammar
3. Organization: introduction, body, and conclusion
4. Mechanics of writing: punctuation and spelling
5. Style and quality of expressions

Essay Topic:

In your opinion, what is the best way to choose a marriage partner?

Use specific reasons and examples why you think this approach is the best.

- ❖ The task should be sufficiently narrow (and/or broad) for the time available.
- ❖ Scoring procedures should be worked out in advance regarding rating the task: an analytic approach or a holistic approach
- ❖ Categories of the scoring approach should be decided on before hand.
- ❖ Descriptors should be provided so as to know what each score within each category means.
- ❖ Rating or scoring should be done anonymously.

An analytic rating scale

Essay rating sheet

Essay number:

Assessor's name:

	Very poor	Poor	Fair	Good	Very good	Excellent
1. Substance	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
2. Thesis development	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
3. Topic relevance	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
4. Introduction	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
5. Coherent support	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
6. Conclusion	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
7. Logical sequencing	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
8. Range	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
9. Word choice	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
10. Word form	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
11. Sentence variety	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
12. Overall grammar	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
13. Spelling	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
14. Essay format	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>
15. Punctuation/capitalization/handwriting	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>

A holistic rating scale

Test of Spoken English Rating Scale

Score

- 60** **Communication almost always effective: task performed very competently**
Functions performed clearly and effectively
Appropriate response to audience/situation
Coherent, with effective use of cohesive devices
Use of linguistic features almost always effective; communication not affected by minor errors
- 50** **Communication generally effective: task performed competently**
Functions generally performed clearly and effectively
Generally appropriate response to audience/situation
Coherent, with some effective use of cohesive devices
Use of linguistic features generally effective; communication generally not affected by errors
- 40** **Communication somewhat effective: task performed somewhat competently**
Functions performed somewhat clearly and effectively
Somewhat appropriate response to audience/situation
Somewhat coherent, with some use of cohesive devices
Use of linguistic features somewhat effective; communication sometimes affected by errors
- 30** **Communication generally not effective: task generally performed poorly**
Functions generally performed unclearly and ineffectively
Generally inappropriate response to audience/situation
Generally incoherent, with little use of cohesive devices
Use of linguistic features generally poor; communication often impeded by major errors
- 20** **No effective communication: no evidence of ability to perform task**
No evidence that functions were performed
No evidence of ability to respond appropriately to audience/situation
Incoherent, with no use of cohesive devices
Use of linguistic features poor; communication ineffective due to major errors

Stage 4: reviewing

- A. Subjective in nature
- B. Inspection of written items by at least one colleague

Stage 5: Piloting or pretesting

- A. Objective in nature
- B. Administration of the newly developed test to a population similar the one to take the test
- C. Two purposes of piloting
 1. Item analysis: individual item features
 2. Test validation: the whole test features

Item analysis: item uniqueness

- A. Item facility and item difficulty
- B. Item discrimination (item differentiation or item discriminability)
- C. Choice distribution (distracter efficiency or distracter tallies)

Item facility (item difficulty, item easiness, or IF)

➤ Definition: the proportion of correct responses to all the responses or the percentage of students answering an item

➤ Formulae for item facility: Formulae for item

1.
$$IF = \frac{N_{\text{correct}}}{N_{\text{total}}}$$

N_{correct} = number of students answering correctly

N_{total} = total number of students taking the test

2.
$$p = \frac{\sum C_r}{N}$$

3.
$$IF = \frac{\sum C}{N}$$

difficulty (ID):

1. $ID = 1 - IF$

2. $q = 1 - p$

q = the proportion incorrect

P = the proportion correct

Table 1

Testees	Items										Total
	1	2	3	4	5	6	7	8	9	10	
1	1	0	1	0	1	0	1	0	1	1	6
2	0	0	1	1	1	0	1	0	0	0	4
3	0	0	1	0	1	0	1	0	0	0	3
4	1	0	1	1	1	0	1	0	0	1	6
5	0	1	1	0	1	0	1	0	0	1	5
6	1	0	1	1	1	0	1	0	0	1	6
7	1	1	1	0	1	0	1	0	0	1	6
8	1	1	1	1	1	0	0	0	0	1	6
9	1	0	1	0	1	0	0	0	0	1	4
10	1	1	1	1	0	0	0	0	0	1	5
11	0	0	1	0	1	0	0	0	0	1	3
12	0	0	1	1	0	0	1	0	1	1	5
13	0	0	1	0	1	0	1	0	0	0	3
14	1	1	1	0	1	0	0	0	0	0	4
15	0	1	1	0	1	0	0	0	0	0	3
16	1	1	1	0	1	0	1	0	0	0	5
17	1	1	1	1	1	0	1	0	0	0	6
18	0	1	1	0	1	0	0	0	1	0	4
19	0	1	1	0	1	0	1	0	0	0	4
20	0	1	1	0	1	0	1	0	0	0	4
21	0	0	1	0	1	0	0	0	0	1	3
22	0	0	1	0	0	0	1	0	0	1	3
23	1	0	1	0	0	0	1	0	1	1	5
24	1	0	1	0	1	0	1	1	0	1	6
25	1	0	1	1	1	0	1	0	0	0	5
26	1	1	1	1	1	0	1	0	0	0	6
27	0	1	1	0	1	0	0	1	1	1	6
28	1	1	1	0	1	0	1	0	1	1	7
29	0	0	1	0	1	0	0	0	1	1	4
30	1	1	1	0	1	0	1	0	1	1	7
Total	16	15	30	9	26	0	20	2	8	18	

$$\text{IF}_{(3)} = \frac{30}{30} = 1 \quad \text{A useless item}$$

$$\text{IF}_{(2)} = \frac{15}{30} = 0.50 \quad \text{An ideal item}$$

$$\text{IF}_{(8)} = \frac{8}{30} = 0.27 \quad \text{A useless item}$$

$$\text{IF}_{(6)} = \frac{0}{30} = 0.00 \quad \text{A useless item}$$

Some points

- A. IF ranges from 0.00 to 1.00.
- B. Acceptable ranges for IF vary, but three are reported in the language assessment literature:
 1. **Between 0.33 and 0.67 (Henning, 1987)**
 2. Between 0.30 and 0.70 (J. D. Brown, 2005)
 3. Between 0.15 and 0.85 (D. Brown, 2004)
- C. Decisions are not necessarily based on IF index, but the following should be considered:

1. The issue of content validity:
2. The necessity of including easy items:
 - a. To create a sense of success among low-ability students
 - b. To set a warm-up
3. The issue of considering all ability levels
4. The issue of sufficient number of items

Item discrimination (ID)

- ❑ Definition of ID: the difference between high scoring group and low scoring group on an item divided by half of the total responses
- ❑ Formulae:

$$1. ID = IF_{upper} - IF_{lower}$$

ID = item discrimination for an individual item

IF_{upper} = item facility for the upper group on the whole test

IF_{lower} = item facility for the lower group on the whole test

$$2. D = \frac{H_c}{H_c + L_c}$$

D = Discriminability

H_c = the number of correct responses in the high group

L_c = the number of correct responses in the low group

$$3. ID = \frac{CH - CL}{\frac{1}{2}N}$$

ID = item discrimination

CH = the number of correct responses to a particular item given by the testees in the high scoring group

CL = the number of correct responses to a particular item given by the testees in the low scoring group

$\frac{1}{2}N$ = the total number of responses divided by 2

Table 2

		Items											Total
Testees		1	2	3	4	5	6	7	8	9	10	Total	
High-scoring group N (15)	28	1	1	1	0	1	0	1	0	1	1	7	
	30	1	1	1	0	1	0	1	0	1	1	7	
	1	1	0	1	0	1	0	1	0	1	1	6	
	4	1	0	1	1	1	0	1	0	0	1	6	
	6	1	0	1	1	1	0	1	0	0	1	6	
	7	1	1	1	0	1	0	1	0	0	1	6	
	8	1	1	1	1	1	0	0	0	0	1	6	
	17	1	1	1	1	1	0	1	0	0	0	6	
	24	1	0	1	0	1	0	1	1	0	1	6	
	26	1	1	1	1	1	0	1	0	0	0	6	
	27	0	1	1	0	1	0	0	1	1	1	6	
	5	0	1	1	0	1	0	1	0	0	1	5	
	10	1	1	1	1	0	0	0	0	0	1	5	
	12	0	0	1	1	0	0	1	0	1	1	5	
	16	1	1	1	0	1	0	1	0	0	0	5	
Low-scoring group N (15)	23	1	0	1	0	0	0	1	0	1	1	5	
	25	1	0	1	1	1	0	1	0	0	0	5	
	2	0	0	1	1	1	0	1	0	0	0	4	
	9	1	0	1	0	1	0	0	0	0	1	4	
	14	1	1	1	0	1	0	0	0	0	0	4	
	18	0	1	1	0	1	0	0	0	1	0	4	
	19	0	1	1	0	1	0	1	0	0	0	4	
	20	0	1	1	0	1	0	1	0	0	0	4	
	29	0	0	1	0	1	0	0	0	1	1	4	
	3	0	0	1	0	1	0	1	0	0	0	3	
11	0	0	1	0	1	0	0	0	0	1	3		
13	0	0	1	0	1	0	1	0	0	0	3		
15	0	1	1	0	1	0	0	0	0	0	3		
21	0	0	1	0	1	0	0	0	0	1	3		
22	0	0	1	0	0	0	1	0	0	1	3		
Total	30	16	15	30	9	26	0	20	2	8	18		

- $ID_{(3)} = 15 - 15 / 15 = 0.00$
- $ID_{(2)} = 10 - 5 / 15 = 0.33$
- $ID_{(8)} = 2 - 0 / 15 = 0.13$
- $ID_{(6)} = 0 - 0 / 15 = 0.00$

Item facility and item discrimination indices

Table 3

Items	1	2	3	4	5	6	7	8	9	10
IF	0.53	0.50	1.00	0.60	0.86	0.00	0.66	0.06	0.26	0.60
ID	0.80	0.33	0.00	0.33	0.00	0.00	0.26	0.13	0.13	0.40

Some notes about ID

- I. Unlike IF, ID also ranges from -1.00 to +1.00.
- II. Like IF, acceptable ID ranges also vary, but the following guidelines have been suggested in the literature:
 - A. 0.40 and up *very good items*
 - B. 0.30 to 0.39 reasonably good, but possibly subject to improvement
 - C. 0.20 to 0.29 marginal items, usually needing and being subject to improvement
 - D. Below 0.19 poor items, to be rejected or improved by revision

One more point

IV. There is a close relationship between item facility and item discrimination? Can you figure it out by looking at table three?

Point-biserial method

- Definition: correlation between item responses and total scores for any given test
- Formula:

$$R_{pbi} = \frac{M_p - M_q}{S_t} \sqrt{pq}$$

R_{pbi} = the point biserial correlation

M_p = the mean total score for examinees who pass the test

M_q = the mean total score for examinees who fail the item

S_x = the standard deviation of test scores

p = the proportion of examinees who pass the item

q = the proportion of examinees who fail the item

Some points

- I. Point-biserial correlation coefficients range from -1.00 to +1.00.
- II. Point biserial correlations of 0.25 and above are acceptable.
- III. Correlation magnitude is a function of sample size and ability range.

Choice distribution

- A. Qualitative and subjective
- B. The frequency with which choices (distracters) are chosen by examinees
- C. Examination of the efficacy of distracters serving a dual purpose
 - 1. To lure less able testees
 - 2. To be evenly distributed
 - A. To keep or drop distractors
 - B. Typology of distractors:
 - 1. Well-functioning or effective
 - 2. Mal-functioning
 - 3. Non-functioning or ineffective

A hypothetical example

Table 4

Item	Key	Choice distribution			
		A	B	C	D
1	A	35	10	0	5
2	C	19	3	25	3
3	B	6	18	8	18
4	D	7	8	25	10

Decisions to keep, modify, or reject items

I. The issue of theoretical considerations

Item facility and item discrimination for CRT is ideally 0.00 and 1.00 and -1.00 and +1.00, respectively at the beginning and end of instruction

II. The principle of practicality

Statistical indexes are only suggestive, not final.

- III. Consideration of IF, ID, CD at the same time
- IV. Inclusion of a wide array of items irrespective of item analysis indexes
 - A. Extremely easy items
 - B. Very easy items
 - C. Easy items
 - D. Fairly easy items
 - E. Items below average difficulty level
 - F. Items above average difficulty level
 - G. Fairly difficult items
 - H. Difficult items
 - I. Very difficult items
 - J. Extremely difficult items