Multiple-Choice Test Item Analysis: A New Look at the Basics

David G. Hamill

Paul D. Usala

U.S. Immigration and Naturalization Service

IPMAAC Panel Discussion: Developing Defensible Written Test Questions: Art, Science, and Some Guidelines.
July 3, 2002

Overview

- Item Analysis Basics
  - » Definition
  - » Why analyze test items?
- INS' Approach to MC Item Analysis
  - » Distractor analysis
    - – Hi/lo statistics
    - – Point Biserial (PB) correlations
  - » Internal consistency
- Examples

Item Analysis Basics

- What is Item Analysis?
  - "The term 'item analysis' refers to a loosely structured group of statistics that can be computed for each item in a test." (Murphy, K., & Davidshofer, 1991)

Why Analyze Test Items?

- Analyze Items to Identify:
  - Potential mistakes in scoring program
  - Ambiguous items
  - Equal distribution of all alternatives
  - Alternatives that don't work
  - "Tricky" items
  - Problems with time limits
  - Potential organizational policy conflicts

Assessment System Parameters

- Program Parameters
  - Assessments are used as part of a promotional assessment system
  - Each assessment is part of a comprehensive examination taxonomy
  - Structured method for test development
  - Continuous testing
    - No piloting of MC test items
  - Equate each new assessment with previous versions

Overview of the Assessment Process

- Develop Assessments
- Administer Assessments to Candidates
- Perform a “Key Clearance”
  - Review item statistics
    - BP correlations
    - Hi/Low statistics
    - Internal consistency correlations
- Score Assessments
- Equate with Current Version

Key Clearance

1. Review PB Data
   - Does a correct response relate to overall test performance?
2. Examine Distractors and Item Difficulty
   - How may candidates responded correctly?
   - Which responses did high & low performers choose?
3. Confirm Decisions with the KR-20 Internal Consistency Estimate of Reliability
   - If this item is removed, will it increase the reliability of the assessment?

Key Clearance

1. Item-Total Correlations (Point Biserial)
   - Purpose: to identify items that correlate with overall test performance

Key Clearance

2. Distractors and Item Difficulty
   - **Difficulty analysis**: to identify the proportion of examinees who choose the correct response option (*p-value)*

$$p_i = \frac{\#\,of\ Persons\,with\,Item\,Correct}{\#\,of\ Persons\,who\,took\,the\,Test}$$

Key Clearance

2. Distractors and Item Difficulty
   - **Method of extreme groups**: categorizes test takers into groups and compares item responses
     - Split test takers into 3 groups based on Scores
       - Upper group (usually top 27-35%)
       - Middle group
       - Lower group (usually bottom 27-35%)

Key Clearance

3. KR-20 Internal Consistency Reliability
   - Identifies items that would increase overall test reliability if dropped from battery

Example #1
Decision Making Assessment

The following passage presents a set of facts. The passage is followed by six conclusions. Read the passage and then decide whether each conclusion is:

A) ***True****, which means that you can infer the conclusion from the facts given*

B) ***False****, which means that the conclusion cannot be true given the facts*

C) ***Insufficient information****, which means that there is insufficient information for you to determine whether the conclusion is true or false*

When updating a training plan for station X, the Patrol Agent in Charge knew that certain courses would be offered during the next quarter. Classes included Fleet Management and Pursuit Management (both part of the Driver Training program), Report Writing, and a Statutory Authority and Legal Update. Based on feedback he had received from many journey-level Border Patrol agents, the PAIC decided that the two Driver Training courses would be required for all trainees. Trainees would be given priority placement in the Driver Training courses and there were enough slots so that all trainees would receive the training. Any additional slots not allocated for priority placements would be offered to other agents on a first-come first-served basis. Due to the amount of time required for the Driver Training classes, anyone who took these courses would not be able to take any other courses. Report Writing and the Statutory Authority and Legal Update courses were offered as optional to anyone interested.

Example #1
Decision Making Assessment Questions 1-3

1. Report Writing can be taken by anyone interested, including trainees such as Agent Alverez.

   A) True B) False C) Insufficient Information

2. Without exception, the agents who will take the Driver Training courses are trainees.

   A) True B) False C) Insufficient Information

3. If an agent is given priority placement in the Driver Training courses, then that agent must be a trainee.

   A) True B) False C) Insufficient Information

Example #1
Statistics for Decision Making Assessment Questions 1-3

Item	PB		Omit	A	B*	C
1.	.425	Freq.	0	29	150	3
		%	0	15.9	82.4	1.6
n = 182		PBR		-.398	.425	-.127
		Hi (%)		4.9	93.9	1.2
Alpha =	.6878	Lo (%)		26.8	70.7	2.4
Alpha if Item is Deleted: .6739						

Item	PB		Omit	A	B	C*	D
2.	.234	Freq.	0	64	74	45	2
		%	0	35.2	39.0	24.7	1.1
n = 182		PBR		-.396	.174	.234	.028
		Hi (%)		17.1	46.3	34.1	2.4
Alpha =	.6878	Lo (%)		50.0	35.4	14.6	0
Alpha if Item is Deleted: .6844							

Item	PB		Omit	A*	B	C*
3.	.166	Freq.	0	158	11	13
		%	0	86.9	6.6	7.4
n = 182		PBR		.007	-.166	.139
		Hi (%)		85.4	2.4	12.2
Alpha =	.6878	Lo (%)		82.9	11.0	6.1
Alpha if Item is Deleted: .6844						

Example #2
Statistics for the In-Basket Document #1

Item #	PB		Omit	A	B	C	D*	E
1	0.137	Freq.	12	210	150	107	317	4
		%	1.5	26.3	18.8	13.4	39.6	0.5
		PB	-0.123	-0.154	-0.025	0.073	0.137	0.013
		Hi (%)	0.3	20.6	17.5	16.9	44.2	0.6
		Lo (%)	2.2	32.5	20	11.4	33.3	0.6

Average Alpha: .4848
Alpha if Item is Deleted: .4929

Example #2
In-Basket Document #1

MEMO

TO: Jamie Lerner,
Associate Chief

From: Devil's Paw Tactical Center

This is to inform you about a problem we encountered while attempting to replace our fuel tanks. We discovered that the old tanks have rusted considerably and that the fuel has been leaking into the surrounding ground. The contractor informed us that, under EPA regulations, we must conduct an environmental impact study before we replace the tanks.

The study, and subsequent cleanup activity, if required, will delay the installation of fuel tanks by no less than three months. We have no funds budgeted for such contingencies. In addition, while the study and any cleanup are taking place, we will have no tanks in which to place bulk fuel delivered to the Tactical Center. We request $25,000 to cover the cost of conducting the environmental impact study.

Director Lippnan,
Please sign and return to me. I will forward this to HQ.
Cecily

Example #2
Question for the In-Basket Document #1

1. Cecily Molina has requested that you sign Document 23, which concerns the Tactical Center's leaking fuel storage tanks. Select the course of action that is closest to the one you would take after reading this memo.

A) Sign the memo and transmit it.

B) Revise the memo before leaving for Anchorage. Sign and transmit the memo once it has been retyped.

C) Revise the memo when you return from Anchorage. Sign and transmit the memo one it has been retyped.

D) Do not sign and transmit the memo. Delay further action until you have learned more about the problem.

E) Do not sign and transmit the memo. Tell Cecily Molina that you have decided not to request funds for the cleanup.

Example #3
Statistics for the In-Basket Document #2

Item #	PB		Omit	A	B	C	D	E*
2	0.214	Freq.	0	45	284	65	6	400
		%	0	5.6	35.5	8.1	0.8	50
		PB		-0.011	-0.153	-0.138	.083	0.241
		Hi (%)		6.1	27.2	4.7	0.3	61.7
		Lo (%)		5	41.9	11.7	1.4	40

Average Alpha: .4848
Alpha if Item is Deleted: .4782

Summary of INS' Item Analysis Methodology

- 3 Indices of Item Performance
  - PB correlations
  - Hi/Lo analyses
  - KR-20 internal consistency
- Review Test Statistics
  - Identify potentially problematic items
    - Make decision to eliminate or retain
  - Identify future anchor items
  - Equate assessments