The Art & Science of Developing Hybrid Situational Judgment and Knowledge/Ability Based Exams

Bobbie Ames & Bridget Bailey

30TH ANNUAL IPMAAC CONFERENCE,
LAS VEGAS, NEVADA

JUNE 27, 2006

What is a hybrid exam?

- A computerized exam that combines two or more methods of measurement within a single exam to assess critical job requirements

Game Conservation Officer Supervisors

- Minimum Qualifications: “Three years of experience in either Land Management or Game Protection as a Game Conservation Officer with the Pennsylvania Game Commission”

The Job Study - Considerations

- Diverse jobs –
  - 5 program areas
  - Little crossover of function
- 29 incumbents
- Six regional offices / Regional differences
- Time constraints

SMEs

- Six Regional Directors (supervisors)
- Developed and operationally defined WBs and KSAs
- Rated WBs and KSAs
- Determined linkage
- Consolidated related KSAs to factors for assessment purposes

Establishing Content Validity

- Interviewed 15 and observed 8 incumbents
- Surveyed all incumbents for WB & KSA ratings
- Interviewed Bureau Directors (HQ) for all program areas

Example - Work Behavior

- Administers and plans the daily operations of Pennsylvania Game Commission (PGC) programs that promote habitat development and management, wildlife management ….
- Reviews, compiles and/or prepares reports concerning various activities . . . and program implementation

Examples - KSAs

- Knowledge of the programs and practices of the Pennsylvania Game Commission
- Knowledge of wildlife identification and habitat management practices and methods
- Ability to exercise proper judgment in various situations encountered on the job including those of a serious or unusual nature
- Ability to communicate effectively in writing

Identifying Job Requirements

- Ratings by incumbents, supervisors, and bureau directors (2nd line supervisors) to identify the important, entry-level job requirements
- The KSAs rated by > 80% as entry-level and received a combined rating of 2.5 on a scale of 1 to 3 in overall importance and in RSP were selected to be measured

Test Planning Considerations

- Homogeneous candidate group
- Estimate 120 eligible candidates
- Test security concerns
- Computer issues
- No oral exam
- Times & locations for testing

Test Design & Development

- Entry-level KSAs categorized into four factors to be measured
- Operationally defined each factor
- Determined most effective method of measurement for each factor
- Considered resources available, validity of method, ease of administration, scope of concept

Technical Job Knowledge

- Measured with standard multiple choice questions to determine knowledge of facts
- High reliability, easy to machine score, but can be difficult to write good questions
- No database items were suitable to SMEs so many were developed and refined
- 45 new TJK items were ultimately used

Effective Working Relationships

- Measured by applying the ability to evaluate the situation and determine what to do and say
- Situational Judgment Test selected – at least moderate reliability, some evidence that similar candidate experience and real-life job-related situations increase validity as a measurement method

Judgment

- Also selected to measure with SJT
- Acting effectively based on knowing facts, assessing a situation, and evaluating choices while considering the implications of problems and solutions
- Ruled out in-basket exercise which may have measured more static information concerning procedures for this job instead of a dynamic unfolding of a demanding situation

Written Communication

- Writing exercise is valid method to measure what candidate chooses to say, organizes it, and then presents it
- Face validity increases candidate acceptance
- Reliability can be questionable because of human rater tendency to subjectivity and time involved. Can control somewhat using boards of 2 raters doing independent ratings

Developing Test Items

- Technical Job Knowledge content identified by “What would a new GCOS specifically need to know about…?”
- List specifics using brainstorming, reference materials, critical incidents, and job observations
- Selected the essential and critical
- SMEs worked individually and in groups to draft items

Situations

- Effective Working Relationships and Judgment situations identified by “How would a new GCOS apply these abilities?”
- Listed specifics using brainstorming, reference materials, critical incidents and job observations
- Selected situations that evolved in steps to develop questions

Written Communication

- Selected situations from both lists and adapted to letter and essay question format
- SMEs wrote responses for rater suggested response guidelines

Technical Job Knowledge

- ✓35 questions
- ✓Select the one best answer choice
- ✓May change answer while on the screen
- ✓Once selection is made cannot return to the question to change answer
- ✓Must answer every question
- ✓ Wrong answers will not be penalized

Exercises

- Four scenarios remain in the test
- Assume the role of a newly appointed supervisor
- Given information supervisor may actually encounter on the job
- Must make a decision about what to do with this information

Exercises (continued)

- Score is the number of correct choices minus incorrect choices
- **Wrong answers penalized**
- Candidate must answer every question
- Once candidate moves to next question cannot go back to change response

Exercises (continued)

- A series of questions based on the developing scenario
- Each decision based on new information and circumstances
- Assume the role of a new Wildlife Education Supervisor
- Make decisions based on what they know at this point in time

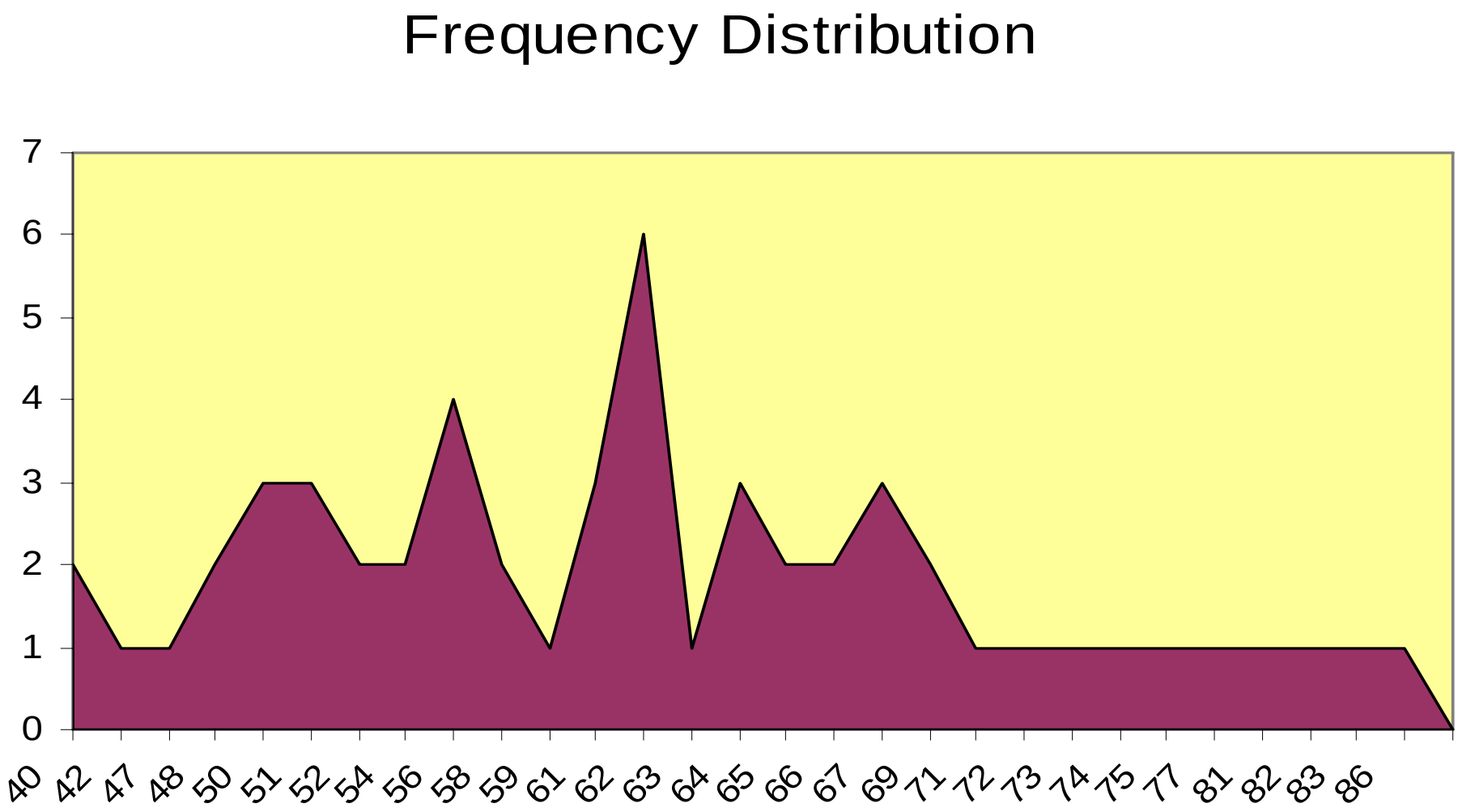
Written Exercises

- Write a letter
- Narrative response to a scenario
- Complete sentences & paragraphs are required
- Clear, concise and well-organized
- Proper grammar, punctuation and vocabulary

Scoring

- Technical items: +1 or 0
- Situations: +1 or -1
- Letter: 0 – 8, combined scores of 2 raters
- Essay: 0 – 32, combined scores of 2 raters
- Maximum 110 points
- Candidate scores ranged from 47 to 86
- Mean = 62.72, Median = 62, Mode = 62
- Std. Dev. = 9.2

Score Frequency Distribution



Scoring Issues

- Computer scoring + Human scoring
- No computer-generated item analysis available for multi-select items
- SPSS does not interface with ICE
- Passing point determination (+1?)

Item Analysis

Difficulty level for multiple-select/multiple choice items – two approaches

- First, for one correct choice per question
  - Ranged from
  - Average = Second, for all correct choices per question
  - Ranged from
  - Average =

Subtest Analysis

- Subtest to subtest correlations indicate each subtest independent of others in measurement – This is good!
- Highest correlation between Effective Working Relationship Subtest and Essay Question (.27)
- Lowest between Technical Job Knowledge and Essay (-.08)

Subtest Correlations

	TJK	STAF	HOA	CITAT	LETR	ESSY
RAW	.29*	.49*	.20	.31**	.34**	.76**
TJK		- .07	- .03	.21	- .09	- .08
EWR			- .02	.01	.27	.10
COM				-.09	.04	.06
JUD					-.10	.00
LET						.14

Correlations by Scores on Factors

	TJK	EWR	JUDG	WRIT
EWR	- .07			
JUDG	.21	.01		
WRIT	- .08	.10	- .10	
RAW SCORE	.29	.46	.31	.33

Correlations by Type of Measurement

	Written TJK Subtest	Situational Judgment Subtests
Essay Subtests	.07	.12
Situational Judgment Subtests	.07	

Reliability Issues

- Good agreement of independent ratings on both written exercises
- SJT and heterogeneous testing methods don’t lend themselves to split-half or internal consistency reliability
- Test-retest? Maybe later…

Validity

- High content validity – Test was based on new job analysis
- Items = high face validity and potentially high construct validity through test design of applying real-life situations to the KSAs (per SMEs)

The Challenges

- Labor relations involvement
- Scenario development Trial-and-Error
- Scoring determinations
- Discovered character number limitation in computer program for narrative responses

Other findings…

- Multiple-select/Multiple-choice items with differential scoring = large range of scores
- Difficulty levels of all subtests comparable
- Test was easy to administer and score
- Real-life job-related situations increased face validity and candidate acceptance

Questions

Contact Information

Bobbie Ames rames@state.pa.us

Bridget Bailey bridbailey@state.pa.us