

Factors Impacting Self-Report Item Responses

Nicholas L. Vasilopoulos and Jeffrey M. Cucina

The George Washington University

February 1, 2002

Self-Report Item Response Process

- ◆ Stimulus Encoding
- ◆ Stimulus Comprehension
- ◆ Response Decision
- ◆ Response Selection

Self-Report Measures Covered

- ◆ Personality Inventories
- ◆ Biographical Data Inventories
- ◆ Situational Judgment Inventories

Personality Inventories

- ◆ Popularity driven by the results of meta-analytic studies demonstrating predictive validity of the Big 5 personality dimensions
 - Agreeableness
 - Conscientiousness
 - Extroversion
 - Emotional Stability
 - Openness to Experience

Personality Inventories: Single Stimulus

- ◆ Typically include a single-stimulus stem with a 4 to 5 point response option

I always finish what I start

- a. strongly disagree
- b. disagree
- c. neither agree nor disagree
- d. agree
- e. strongly agree

Personality Inventories: Forced-Choice

- ◆ Forced-choice items present two statements identified as equal in desirability
 - A. I come up with good solutions
 - B. I act comfortably around others
- ◆ One option is related to the construct of interest while the other serves as a distracter
- ◆ Developed to minimize the impact of impression management

Biographical Data (Biodata) Inventory

- ◆ A self-report inventory designed to provide information about a respondent's life history
- ◆ Scoring Biographical Data
 - Rational scoring key - based on theories of human development
 - Empirical scoring key - based on results of criterion-related validation study

Mael (1991) Taxonomy of Biodata Items

HARD ITEMS



SOFT ITEMS

Methodological Facets (Ensure Correct Responding)

External (Observable actions)

Did you ever get fired from a job?

Objective (recall of events)

How many hours did you study for the SAT?

First Hand (examinee's perspective)

How punctual are you about coming to work?

Discrete (single, unique event)

In what year did you get your driver's license?

Verifiable (allows corroboration)

Were you ever suspended from school?

Internal (attitudes and thoughts)*

What is your attitude toward friends who smoke?

Subjective (interpretation of events)**

How adventurous are you compared to your peers?

Second Hand (observer's perspective)

How would your boss describe your work ethic?

Summative (average over time)*

How many hours do you study each week?

Non-Verifiable (can not be corroborated)

How often do you play computer games at work?

Mael (1991) Taxonomy of Biodata Items

HARD ITEMS



SOFT ITEMS

Historical

Historical (facts)

How old were you when you got your first job?

Future or Hypothetical (conjectural)

What position do you think you will hold in 10 years?

Moral or Ethical Concerns

Controllable (examinee had a choice)

How many times did you take the CPA exam?

Non-Controllable (examinee had no choice)*

How many siblings do you have?

Equal Access (open to all groups)

Do you belong to your local library?

Non-Equal Access (some groups excluded)*

Were you captain of the football team?

Job-Relevant

How many television sets did you sell last year?

Not Job-Relevant*

Are you proficient at crossword puzzles?

Non-Invasive (less personal)

How many clubs did you belong to in high school?

Invasive (highly personal)

How much money do you owe on your credit cards?

Situational Judgment Inventories (SJI)

- ◆ Includes a number of job-relevant scenarios presented along with a number of options representing possible behavioral responses to the situation
- ◆ Respondents select from among the options the behaviors they are most likely and least likely to exhibit
- ◆ The most and least likely responses are combined to compute a score for the scenario

Example Situational Judgment Item

While you are working at the service desk, a customer attempts to return a toaster. According to store policy, a customer cannot return a product without presenting a receipt. The customer claims that he accidentally threw out the receipt, but assures you that he purchased the toaster two days ago. When you tell the customer that you cannot allow him to return the toaster without a receipt, he becomes enraged and begins to yell at you.

1. Which of the following behaviors would you **MOST** likely exhibit?
2. Which of the following behaviors would you **LEAST** likely exhibit?
 - A. Call security to have the customer removed from the store.
 - B. Allow the customer to return the toaster.
 - C. Ignore the customer until he calms down.
 - D. Tell the customer you cannot help him unless he calms down.
 - E. Tell the customer that you will ask the manager to give him a refund.

Correlates of Self-Report Item Responses

◆ Characteristics of the Respondent

- Age
- Gender
- Culture / Race
- Social Desirability
- Job Familiarity

Correlates of Self-Report Item Responses

- ◆ Characteristics of the Self-Report Measure
 - Item Format
 - Item Context
 - Test Instructions

Age and Self-Report Measures

Ones and Viswesvaran (1998)

- ◆ Meta-analysis showing that older respondents (i.e., 40 years old and over) tend to score slightly higher on integrity tests than younger respondents

Goldberg, Sweeney, Merenda, and Hughes (1998)

- ◆ Older respondents scored higher on measures of conscientiousness than younger respondents

Gender and Self-Report Measures

Ones and Viswesvaran (1998)

- ◆ Women scored slightly higher on an integrity tests

Petrides and Furnham (2000)

- ◆ Men scored higher on an emotional intelligence test

Gender and Self-Report Measures

Costa, Terracciano and McCrae (2001)

- ◆ Women scored higher on measures of neuroticism, agreeableness, and openness to feelings
- ◆ Men scored higher on measures of assertiveness and openness to ideas
- ◆ Findings held across 26 cultures for both college-students and adults, with the largest effects found in American and Europe

Culture and Self-Report Measures

Levine and Franco (1981)

- ◆ Hispanics were less likely to provide information about their finances and personal tastes than Anglo-Americans

Alva (1985)

- ◆ Hispanics who scored high on acculturation scales reported having social values similar to those reported by Anglo Americans

Culture and Self-Report Measures

Chan (1997)

- ◆ No Black-White differences in the perceived predictive validity of a personality test

Collins and Gleaves (1998)

- ◆ The Five-Factor Model of personality fit equally well for Black and White job applicants

Culture and Self-Report Measures

Schmitt and Chan (1997)

- ◆ Compared to Blacks, Whites had higher situational judgment inventory (SJI) scores and perceived the SJI as more face validity
- ◆ Smaller racial differences were found when using a video-based rather than a paper-and-pencil test method

Whitney and Schmitt (1998)

- ◆ Black-Whites differences for 25% of biodata items
- ◆ Blacks scored higher than Whites on the biodata inventory

Social Desirability and Self-Report Measures

Paulhus (1984)

- ◆ Concerned that measure of social desirability did not correlate with each other
- ◆ Factor analyzed score from different social desirability measures
- ◆ Found that social desirability measures could be adequately represented by two factors

Two Types of Social Desirability

◆ *Self-Deception*

- Exaggerated claims of positive cognitive attributes driven by a subconscious need for ego defense/enhancement

◆ *Impression Management*

- Exaggerated claims of performing desirable behaviors
OR not performing undesirable behaviors
- Key concern for specialists using self-report measures to make human resource decisions

Impression Management and Validity

Ones, Viswesvaran, and Reiss (1996)

- ◆ Controlling for impression management does not significantly increase the relationship between scores on measures of personality and job performance

Ellingson, Sackett and Hough (1999)

- ◆ Controlling for impression management does not change the factor structure of items included on a personality measure

Impression Management and Utility

Rosse, Stecher, Miller and Levin (1998)

- ◆ Strongest relationships between IM and conscientiousness and emotional stability
- ◆ Respondents engaging in IM scored at the high end of the distribution
- ◆ Results in a high proportion of respondents identified as impression managers being hired when using top-down selection, especially when the selection ratio is low

Limitations of Impression Management Research

Anderson et al, (1984) and Pannone (1984)

- ◆ Found a substantial increase in the predictive validity of biodata inventories when using reported performance of “bogus tasks and experiences”
- ◆ Different than findings of studies using unlikely virtue scales

Holden, Kroner, Fekken and Popham (1992)

- ◆ Item response latencies added significantly to the prediction of impression managers when used with an impression management scale

Limitations of Impression Management Research

Barrick and Mount (1996)

- ◆ Overall, controlling for impression management did not increase criterion-related validity
- ◆ Examination of the correlation matrices presented revealed different effects for impression management in the two organizations providing data
- ◆ Did not consider the possibility that the organizations differed on a characteristic that moderates the relationship between impression management and performance

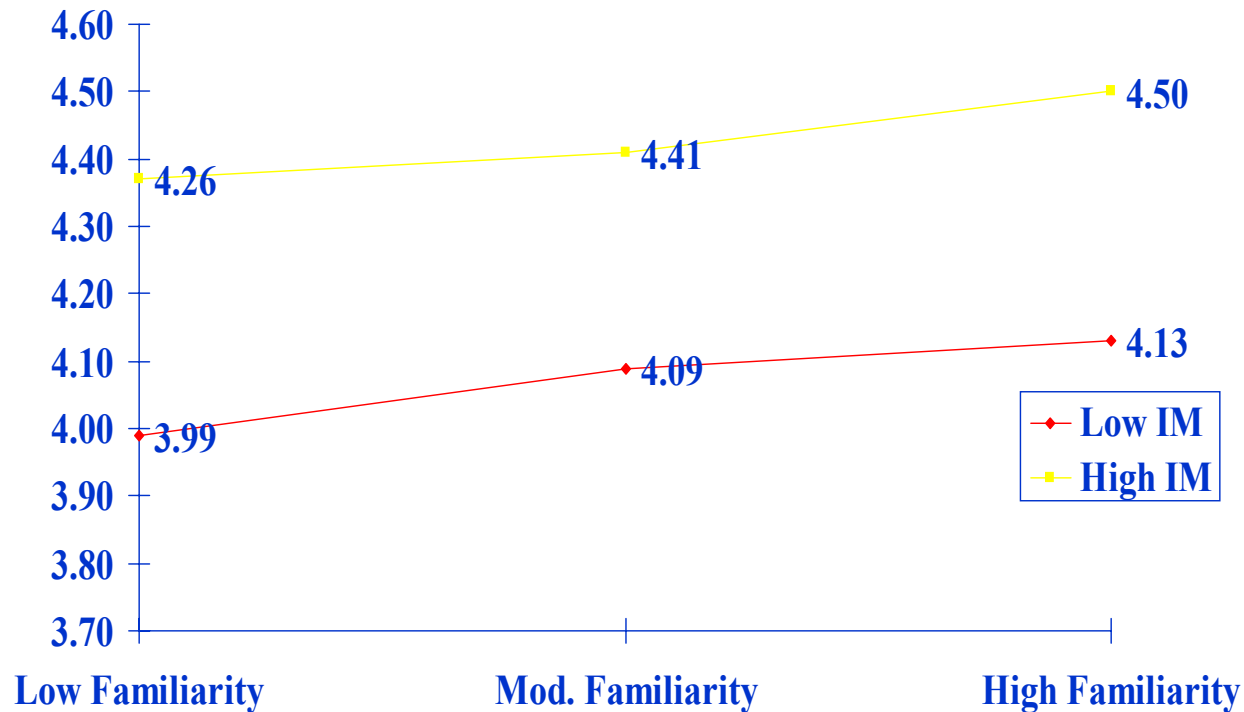
Job Familiarity and Self-Report Measures

Vasilopoulos, Reilly and Leaman (2000)

- ◆ Found that job familiarity influenced both honest and impression-managed responses
- ◆ Effect appeared be greater for impression managers (however the interaction was nonsignificant)

Job Familiarity and Self-Report Measures

Vasilopoulos, Reilly and Leaman (2000)



Job Familiarity and Honest Responses

Markus and Kunda (1986)

- ◆ Makes self-schemas for job-related attributes more salient in the respondent's "working self-concept"

Kunda and Sanitioso (1989)

- ◆ Higher self-ratings found for attributes believed to be related to performance

Job Familiarity and Impression Managed Responses

Holden, Kroner, Fekken and Popham (1992)

- ◆ Proposed that familiarity with the stereotypical qualified job applicant is used to develop and adopted schema

Vasilopoulos, Reilly and Leaman (2000)

- ◆ Proposed that the strength of an adopted schema for a qualified job applicant is a function of job familiarity
- ◆ Suggested by the finding that job familiarity was related to faster item response latencies among impression managers

Characteristics of the Self-Report Measure

- ◆ Item Characteristics
 - Frame-of-Reference
 - Forced-Choice v. Multiple Choice
 - Situational Judgment Item Responses
 - ◆ Least versus Most Likely
 - ◆ Should versus Would Do
- ◆ Test Characteristics
 - Bandwidth-Fidelity Dilemma
 - Warning of Response Verification

Forced-Choice and Self-Report Measures

Jackson, Wroblewski and Ashton (2000)

- ◆ Compared the validity obtained from forced-choice versus single stimulus response personality inventories.
- ◆ Respondents could inflate scores on both inventories when applying for a job
 - Single stimulus inventory score increased 1 SD
 - Forced-choice inventory score increased 1/3 of an SD

Forced-Choice and Self-Report Measures

Jackson, Wroblewski and Ashton (2000)

Validity of Forced-Choice and Single Stimulus Inventories

Condition	Single Stimulus	Forced-Choice
General	.48	.41
Job Applicant	.18	.36

Note. Criterion was self-reported frequency of delinquent behaviors at work

Most/Least Likely Situational Judgment Responses

McElreath and Vasilopoulos (2002)

- ◆ Least likely responses have a higher correlation with performance measures than most likely responses
 - Least likely responses involve decision processes similar to those used in the job
- ◆ Least likely responses have a higher correlation with cognitive ability
 - The least likely response decision process is more complex than the most likely response process

Most/Least Likely Situational Judgment Responses

McElreath and Vasilopoulos (2002)

	1	2	3	4	5	6	7
<u>Predictors</u>							
1. Least Likely SJI Responses	---						
2. Most Likely SJI Responses	.56**	---					
3. Inductive Reasoning	.20**	.11*	<u>.79</u>				
4. Deductive Reasoning	.22**	.11*	.71**	<u>.84</u>			
<u>Criteria</u>							
5. Training Score	.19**	.09	.55**	.58**	---		
6. Supervisor Rating	.04	.04	.20**	.16**	.27**	---	
7. Work Simulation	.21**	.16**	.49**	.49**	.49**	.26**	---

“Would/Should Do” Situational Judgment Responses

Ployhart and Ehrhart (2001)

- ◆ Compared the criterion-related validity obtained when asking respondent’s what they “would do” versus what they “should do”
- ◆ Found substantially higher validities for “would do” responses

“Would/Should Do” Situational Judgment Responses

Ployhart and Ehrhart (2001)

	Most/Least Likely Response ("would do")	Most/Least Effective Response ("should do")
1. GPA	.40*	.16*
2. SAT Total	.01	-.03
3. Self-Report Performance	.48*	.09
4. Peer Performance	.49*	.19*

Bandwidth-Fidelity Dilemma

- ◆ Broad v. Narrow Traits
 - Big five personality dimensions composed of facets
- ◆ Administrative demands can prevent measuring many traits with high reliability
 - Often impractical to emphasize both bandwidth and fidelity
- ◆ Two options
 - Emphasize bandwidth
 - Emphasize fidelity

Bandwidth-Fidelity Dilemma

Cronbach and Gleser (1957, 1965)

- ◆ Suggest using narrow traits to predict specific criteria
- ◆ Suggest using broad traits to predict broad criteria

Ones and Viswesvaran (1996)

- ◆ Advocate use of broad trait measures (i.e., fidelity)
- ◆ Do acknowledge benefit of using narrow traits in certain circumstances

Bandwidth-Fidelity Dilemma

Paunonen, Rothstein and Jackson (1999)

- ◆ Advocate use of narrow trait measures (i.e., bandwidth)
- ◆ Suggest that using narrow traits will give better prediction
- ◆ Broad traits aggregate facets that may have conflicting or different relationship to performance
- ◆ Recommend regressing performance on the narrow trait measures to maximize prediction
- ◆ Regression indicates the the relative weight that should be given to each narrow trait measure

Research on the Bandwidth-Fidelity Dilemma

Stewart (1999)

- ◆ Narrow traits add incremental validity at different times in employment

Vasilopoulos, Cucina, Goldenberg and Usala (2002)

- ◆ Narrow measures of conscientiousness and emotional stability were better predictors of training performance

Implications of the Bandwidth-Fidelity Dilemma

- ◆ Easier to implement use of broad trait measures
 - Although may not get best prediction
- ◆ Use broad traits to predict broad criteria
- ◆ Use narrow traits to predict narrow criteria

Warning of Response Verification

◆ Types of Warnings

- Detection
- Negative consequences
- Both detection and negative consequences

Example of Warning Instructions

As you complete the assessment, please describe yourself and your experiences honestly and accurately. Responses on this assessment are subject to verification. Deliberate attempts to falsify information may be grounds for not employing you or for dismissing you after you begin work.

Warnings and Impression Management

Dwight and Donovan (1998)

- ◆ Conducted a meta-analysis and found that warnings reduced the effect of impression management ($d = .25$)
- ◆ Found moderators of the warning/impression management relationship
 - type of sample
 - type of warning

Warning and Cognitive Ability

Cucina, Vasilopoulos and McElreath (2002)

- ◆ Adding warning of response verification may affect construct validity of personality scale scores
- ◆ No Warning: Personality and cognitive ability unrelated
- ◆ Warning: Personality and cognitive ability related

Questions and Comments

