CPS HR CONSULTING

The Effects of Response Instructions on Situational Judgment Test Performance in a High-Stakes Employment Context

IPAC 2011 Conference
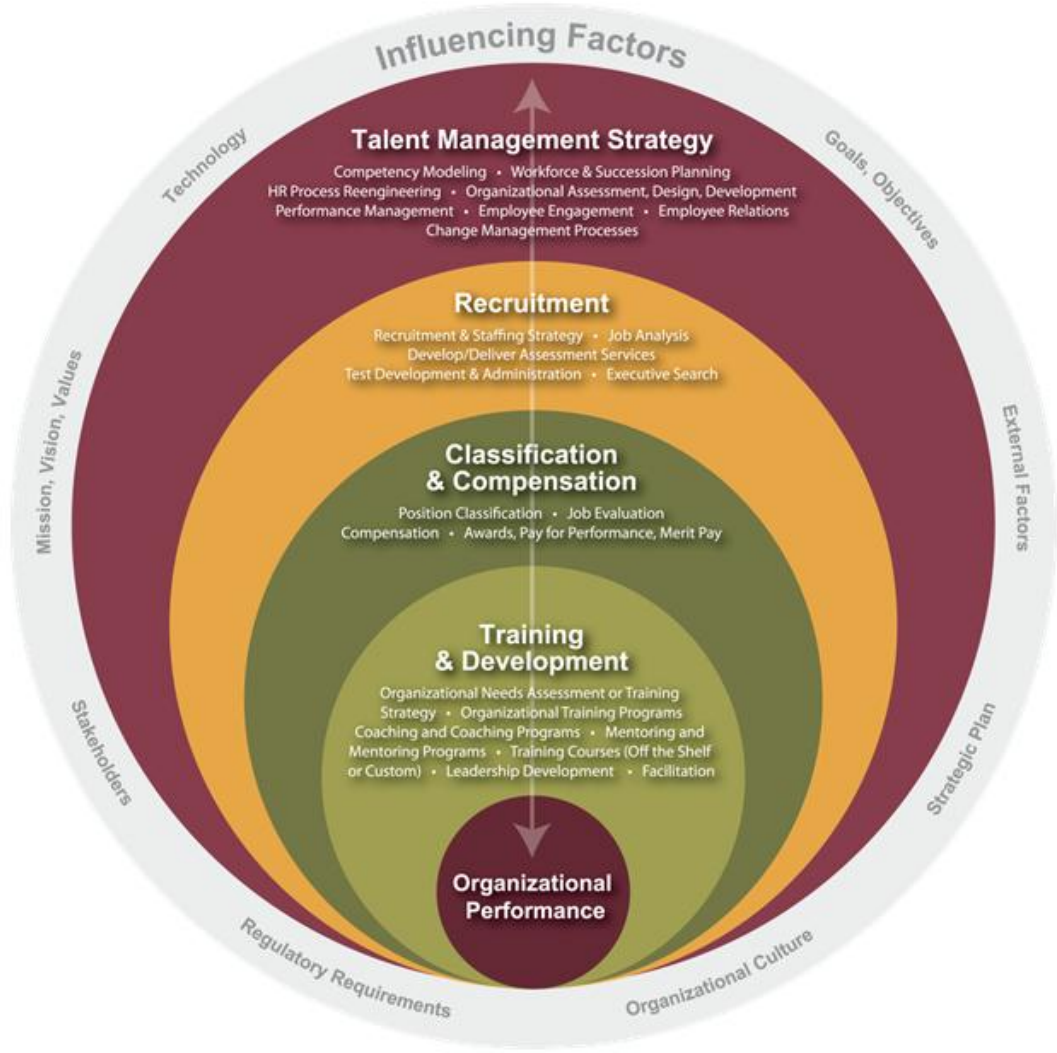
Washington, D.C.

July 19, 2011

Clinton Kelly, A.N.G.
Jason Schaefer, M.B.A., M.A.

Who We Are

CPS HR is a self-supporting public agency providing a full range of integrated HR solutions to government and nonprofit clients across the country. Our strategic approach to increasing the effectiveness of human resources results in improved organizational performance for our clients. We assist our clients in the areas of organizational strategy, recruitment and selection, classification and compensation, and training and development

Presentation Overview

- **SJT Background**

- **Knowledge vs. Behavioral Instructions**

- **Current Study in High-Stakes Situation**

- **Conclusions / Next Steps**

- **Questions / Discussion**

SJT Definition

- **Designed to asses an applicant's judgment regarding a situation encountered in the work place. A respondent is presented with work-related situations and a list of plausible courses of action.**

- **Example**

You are working on a project and are worried that you may not be able to complete the project by the deadline you were given by your supervisor. What is the best response in this situation?

- **a. Let you supervisor know that the deadline was unrealistic.**
- **b. Complete the most important parts by the deadline.**
- **c. Hand in what you were able to complete by the deadline.**
- **d. Ask for an extension of the deadline.**

SJT Background

- **Various administration methods**
  - **Paper and Pencil**
  - **Video**
  - **Hybrid**
  - **Interactive**

- **Been in use for over 100 years**

- **Are a measurement method and not a single construct measure**
  - **Examples of constructs measured**
    - **Interpersonal skills**
    - **Supervisory potential**
    - **Teamwork skills**

Response Instructions

- **Behavioral tendency**
  - **Respondents are asked how they would likely behave in a given situation**
    - **Would do**
    - **Most and least likely do**
    - **Have done**

- **Knowledge**
  - **Respondents are asked to evaluate the effectiveness of possible responses to a given situation**
    - **Should do**
    - **Most effective / best response**
    - **Best and worst**
    - **Rate effectiveness of each response**

Response Instructions

- **Knowledge instructions might best predict maximum performance**

- **Behavioral instructions might best predict stable/typical performance**

Ployhart & Ehrhart, 2003

- **Administered SJT to college students which presented difficult study situations**

- **Used six different instruction sets while keeping stem identical**
  - Three Behavioral: (1) most and least likely do, (2) option which best represents what you done in the past, (3) how likely you would perform each response
  - Three Knowledge: (1) what should you do, (2) most effective and least effective response, (3) rate the effectiveness of each response

Ployhart & Ehrhart, 2003 (cont.)

- **Intercorrelations among behavioral and knowledge SJTs were stronger than intercorrelations between each type**

- **Knowledge instructions had higher means, smaller standard deviations, and less normal distributions**

- **Internal consistency reliabilities did not vary much by type**

- **Behavioral instructions showed higher criterion-related validities**

Nguyen, Biderman, McDaniel, 2005

- **SJT and Big Five test were administered to college students who were instructed to respond honestly for one administration and "fake good" on another administration.**

- **SJT content remained the same and candidates responded using both knowledge and behavioral instructions**

Nguyen, Biderman, McDaniel, 2005 (cont.)

- **Candidates improved scores when faking with behavioral instructions**

- **Knowledge instructions appeared to be more immune to faking**

- **Knowledge instructions correlated more highly with cognitive ability**

McDaniel et al., 2007

- **Meta-analysis of response instructions and validity**

- **Only included studies where paper and pencil SJT was used and participants were employees or applicants (most were employees)**

- **SJTs with knowledge instructions had higher correlation with cognitive ability**

- **SJTs with behavioral instructions had higher correlation with three of the Big Five (Agreeableness, Conscientiousness, and Emotional Stability)**

McDaniel et al., 2007 (cont.)

- **Criterion-related validity was equal for both instruction types**

- **SJTs with both instruction types provided incremental validity above cognitive ability and Big Five**

Lievens, Sackett, Buyse, 2009

- **Administered 30 item SJT to perspective medical school students**

- **Content remained the same and half were given behavioral instructions and the other half knowledge instructions**

- **No meaningful score differences between knowledge and behavioral instructions**

- **Knowledge instructions had higher correlation with cognitive ability**

Lievens, Sackett, Buyse, 2009 (cont.)

- **There was no difference in criterion-related validity by instruction type**

- **Authors suggest knowledge instructions should be preferred because they make faking a non issue**

SJT Instruction Type Research

- **Most research has involved students or incumbents in low-stakes situations**
  - **Exception to this is Lievens, Sackett & Buyse, 2009**

- **There is a call for SJT research in high-stakes situations, specifically in personnel selection**
  - **McDaniel & Whetzel, 2009 and Ployhart & MacKenzie, 2010**

Current Study

- **High-stakes situation – Entry Firefighter paper and pencil test**

- **Test Sections**
  - **1: Reading Comprehension – 30 Items**
  - **2: Math – 30 Items**
  - **3: Mechanical Aptitude – 25 Items**
  - **4: Interpersonal (SJT) – 15 Items**
    - **First 8 items – Knowledge Instructions (In this situation you should)**
    - **Next 7 items – Behavioral instructions (In this situation you would)**

- **Two forms of the test – alternate instruction type**

Current Study

- **1,457 candidates applying for entry Firefighter completed one form of the test**

- **Still collecting data on the second form**

- **Hypotheses**
  - **Expect to find no difference between the instruction type regarding correlation with cognitive ability**
  - **Instruction type will not affect mean scores**

Current Study

- **Test Taker Demographics**
  - **Male = 80.3%**
  - **Female = 4.4%**
  - **Did not Indicate = 15.3%**

- **Ethnicity**
  - **Caucasian = 44.1%**
  - **Hispanic = 14.2%**
  - **Asian = 5.1%**
  - **Other = 4.3%**
  - **African American = 4.1%**
  - **Filipino = 3.2%**
  - **Native American = 1.2%**
  - **Did Not Indicate = 23.8%**

Current Study

- **Mean score on test = 73.5, SD = 11.8**
  - **Cognitive Ability (first three test sections)**
    - **85 Items**
    - **Alpha = .89**

- **Since we compared instruction sets across different items:**
  - **Controlled for Flesch-Kincaid Reading Level of each item**
  - **Reading level could increase cognitive loading**
  - **Controlled for item difficulty**

- **Average corrected point-biserial correlations**
  - **Should = .09 (SD = .05)**
  - **Would = .14 (SD = .03)**

Current Study

- **Average P-value**
  - **Should = .82 (SD = .10)**
  - **Would = .86 (SD = .09)**

- **P-values ranged from .64 to .95.**
  - **8 of 15 P-values in the .80s and .90s**

- **Average Flesch-Kincaid Reading Level ranged from 6.4 to 14.5**
  - **Should = 10.21 (SD = 1.91)**
  - **Would = 12.24 (SD = 2.07)**

Preliminary Findings

- **Would items were more highly correlated with cognitive ability when controlling for reading level and item difficulty**
  - **Opposite trend from what is typically found**

- **Inconclusive results due to small number of items in each condition**

Positives

- **Actual applicants vying for a position in the fire service**

- **Same candidates in both conditions**

- **Similar to exams that contain both MC cognitive ability and SJT items/instructions**

- **Data set based on responses of 1,457 test takers**

Negatives

- **Too few items in each condition**

- **Possible cuing based on both instruction sets in same test**

- **Item content varied between conditions as did difficulty and reading level**

- **Quality of distractors may have varied differentially across conditions leading to result**

- **Some items may not be members of population of "quality test items" PBISC < .10**

- **Range restriction on item difficulty, items tended to be easier**

CPS HR CONSULTING

Questions