# Evaluating the Reliability and Validity of the Federal Employee Viewpoint Survey (FEVS)

**Jeffrey M. Cucina,**

**Philip T. Walmsley,**

&

**Kevin A. Byle**

U.S. Customs and Border Protection

U.S. Customs and Border Protection

# Objective

- Describe applied research on the Federal Employee Viewpoint Survey (FEVS)
  - **Reliability:** Internal consistency, nesting
  - **Construct Validity:** Factor structure
  - **Criterion-Related Validity:** Prediction of turnover and performance
  - Address gap in applied research
  - Demonstrate how to apply psychometric techniques to surveys (Gast, 2008; 2010)

# Background

- Growing interest in employee engagement surveys
- U.S. Federal Government agencies required to administer a survey to employees each year.
- U.S. Office of Personnel Management (OPM) implemented Federal Employee Viewpoint Survey (FEVS)
  - Meets requirement mentioned above
  - Previously known as the Federal Human Capital Survey
  - Used as data source for Partnership for Public Service's *Best Places to Work in the Federal Government* rankings

# Background

- Several calls for research on the FEVS
  - Ferro (2015) called for linkages of FEVS data to other sources (e.g., performance ratings, turnover)
  - Fernandez et al. (2015) questioned measurement validity of FEVS and called for evaluating its reliability
- Little publicly-available research on FEVS reliability and validity
  - Most publicly-available reports designed for end users (e.g., leaders, supervisors, employees)

# Our Work

- We have been involved in several research studies examining reliability and validity of the FEVS

- Today we will present an overview of this work
  - Begin with short refresher on job satisfaction and employee engagement
  - Describe four studies
  - Study 1: Reliability
  - Study 2: Construct Validity
  - Study 3: Unit-level criterion-related validity
  - Study 4: Individual-level criterion-related validity*

# Job Satisfaction/Employee Engagement

- Job satisfaction is one of the primary job-related attitudes of interest to I/O psychologists

- Individual-level correlation between job satisfaction and job performance is .30 (Judge et al., 2001, true-score correlation)

- Unit-level correlations between job satisfaction and organizational productivity (.20), profit (.15), and turnover (-.36; Harter et al., 2002)

- Controlling for big five personality dimensions reduces validity to .19; controlling further for self-esteem variables reduces to .09 (Bowling, 2007).
  - Cook (2009) obtained similar results

- In general, there's a correlation between job satisfaction and performance, but it's not as high as often thought

- Employee engagement is strongly related to job satisfaction
  - Engagement includes some motivational aspects
  - However, unit-level true-score correlation is .91 (Harter et al., 2002)

# Study 1: Reliability

- ## Dataset:
  - 2008 FEVS administration
  - $n$ = 212,223 civilian full-time, permanent employees
  - Data from 73 agencies and 272 components
  - 63 items measuring job satisfaction and employee engagement
    - 11 additional questions on OPM's benefits were not included here

- ## Results:
  - Coefficient alpha (internal consistency) = .982
    - Very high value in "Excellent" range per U.S. Department of Labor (2000, p. 3-3) guidelines
    - All items had positive corrected point-biserials (mean = .672, range .454 to .804)
    - Average correlation between items was .461
    - Suggests that respondents provide very consistent responses across all items.

# Study 1: Reliability

- Results:
  - Inter-rater reliability/nesting/"groupiness"
    - Estimated an intraclass-correlation coefficient (ICC) for each item.
      - Measure of reliability when using raters
      - Used unconditional multi-level model
      - Essentially tells you correlation between two randomly chosen employees' responses to an item (averaged across all employees)
    - Mean ICC across 63 items was .039 at the agency-level
      - Ranged from .010 to .171
      - On average, two randomly chosen employees have responses that correlate .039
      - Percent of variance in individual-level FEVS data accounted for by agencies is .039
    - At component-level, mean ICC was .029 (range .013 to .106)
    - Suggests very little nesting in FEVS data
    - Over 90% of variance in individual-level FEVS data not explained by agencies or subcomponents

# Study 2: Construct Validity

- Research Question: What does the FEVS measure?  Is it unidimensional given the high coefficient alpha?
  - Unidimensional tests tend to have high alphas, but not necessarily vice versa (Cortina, 1993)

- Dataset:
  - 2011 and 2012 FEVS administrations
  - $n_{2011} = 266,376$; $n_{2012} = 687,687$
  - 285 subcomponents in 2011; 298 subcomponents in 2012
  - 71 items measuring job satisfaction and employee engagement
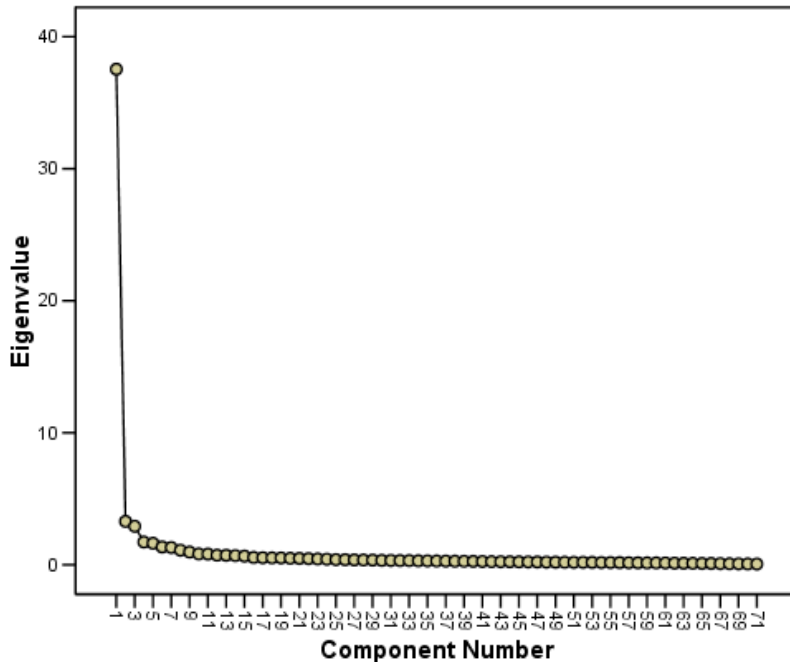
# Study 2: Construct Validity

- Results:
  - Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA)
    - Used 2011 dataset
    - EFA and PCA scree plots suggests presence of a large general factor
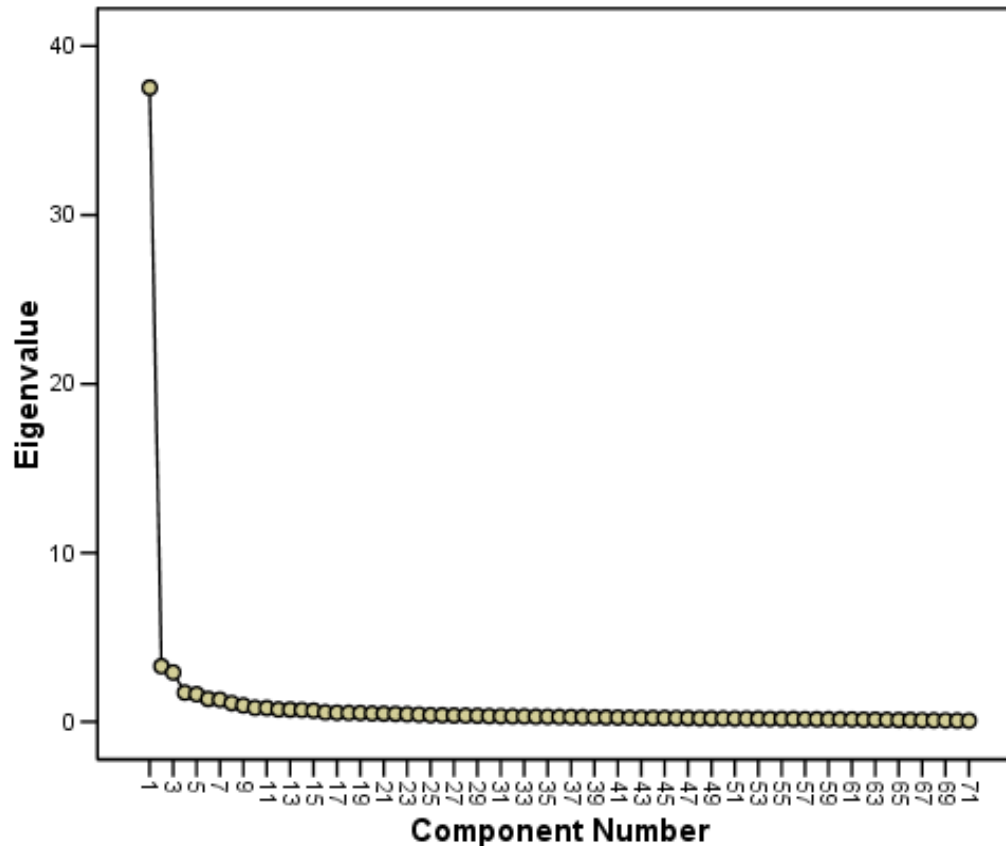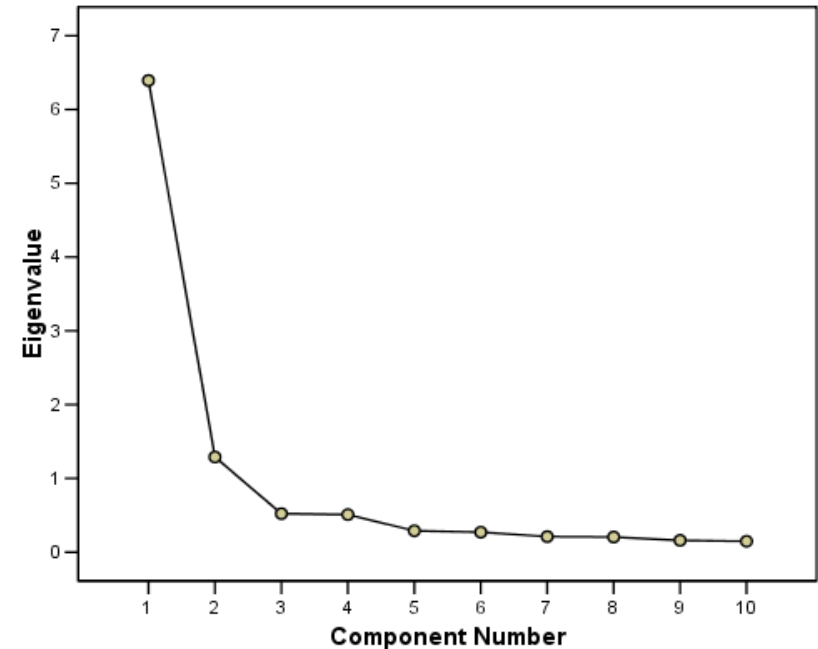
# Study 2: Construct Validity

- Results:
  - Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA)
    - For comparison, consider the FEVS scree plot versus those for cognitive tests, for which a general factor is largely accepted.

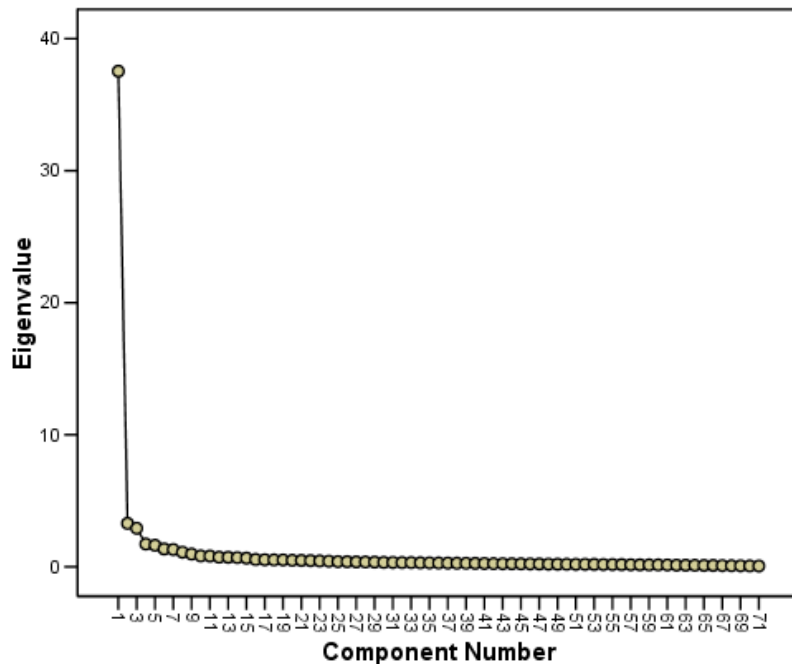FEVS                              vs.                              ASVAB
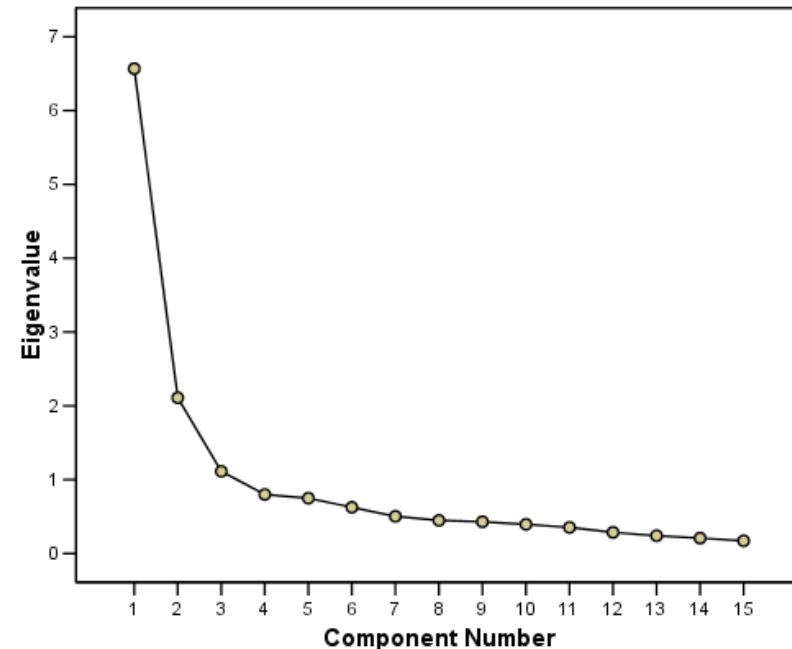
# Study 2: Construct Validity

- ## Results:
  - Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA)
    - For comparison, consider the FEVS scree plot versus those for cognitive tests, for which a general factor is largely accepted.

FEVS                                    vs.                              GATB

# Study 2: Construct Validity

- Results:
  - Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA)
    - For comparison, consider the FEVS scree plot versus those for cognitive tests, for which a general factor is largely accepted.
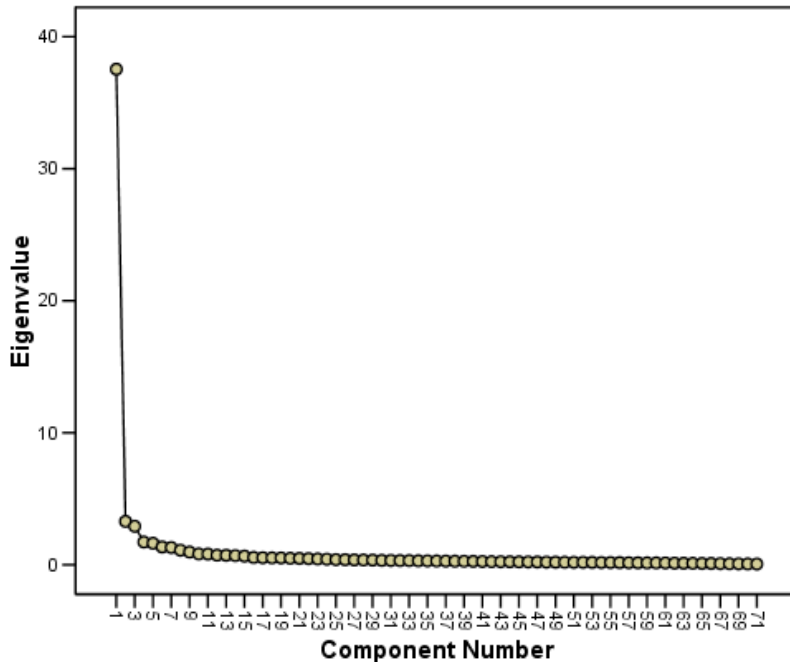
FEVS                          vs.                          EAS

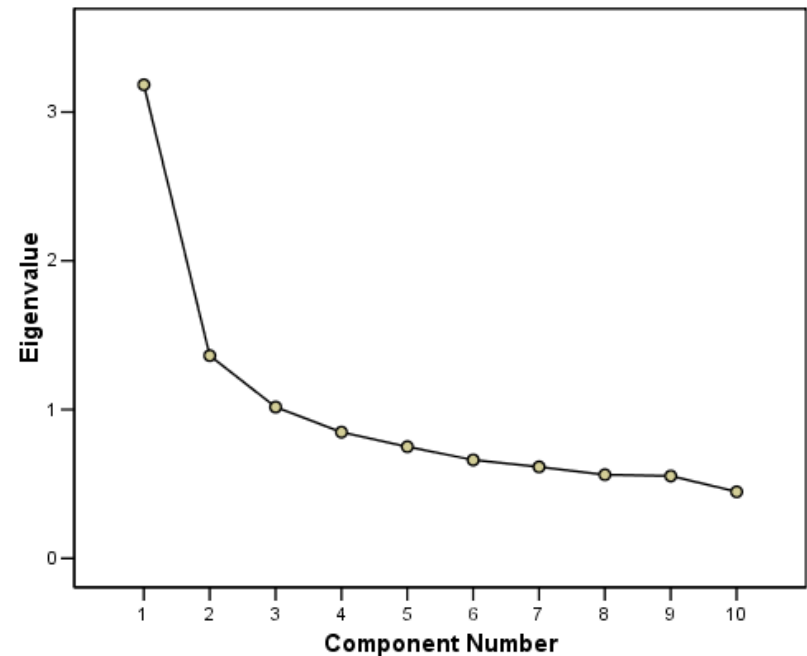# Study 2: Construct Validity

- Results:
  - Exploratory Factor Analysis (EFA) and Principal Components Analysis (PCA)
    - For comparison, consider the FEVS scree plot versus those for cognitive tests, for which a general factor is largely accepted.
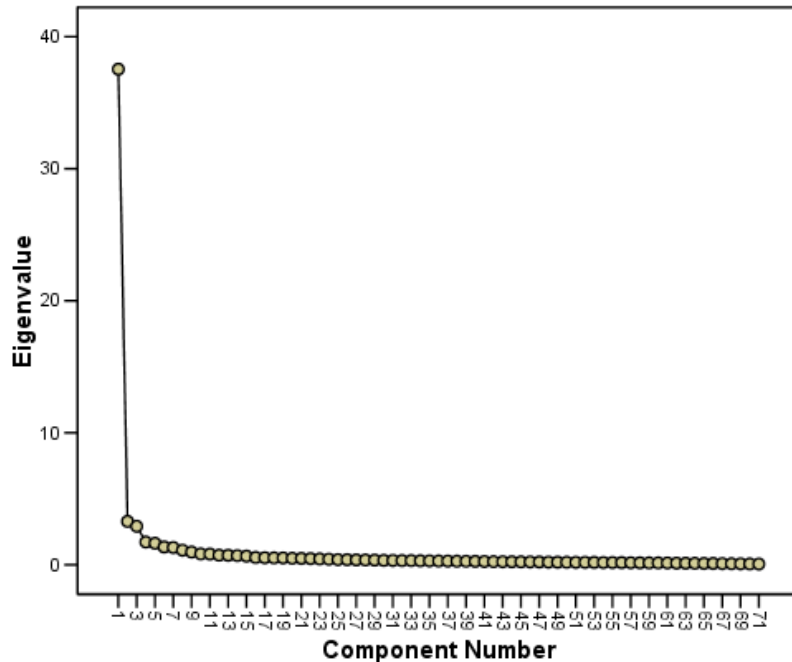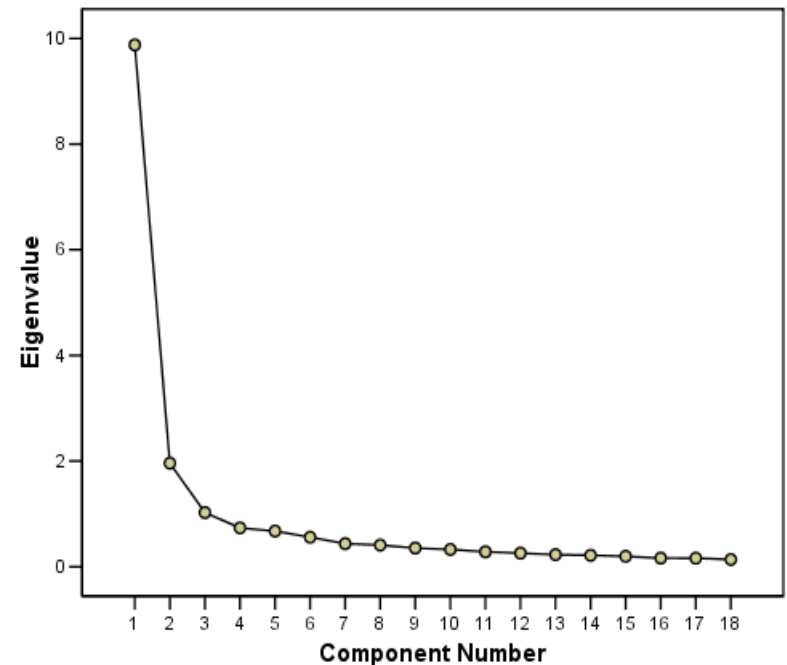
FEVS                    vs.                    DAT

# Study 2: Construct Validity

- Results:
  - Confirmatory Factor Analysis (CFA)
    - Used 2011 dataset to develop (but not validate) models
    - Used nine different *a priori* factor structures (linkages of items to factors)
    - EFA and PCA scree plots suggests presence of a large general factor
    - Conducted analyses using individual items and parcels as well as different subsets of items
    - Conducted analyses at individual employee level and on aggregated data (i.e., means for each subcomponent on each item served as cases)
  - OPM's assignment of 71 items to six factors provided the best fit
    - Found evidence that a higher-order model provided excellent fit (better than orthogonal and very close or nearly identical to oblique)
    - However, our empirical search for a model and use of EFA on the 2011 data capitalizes on chance
  - Cross-validated the model using 2012 data
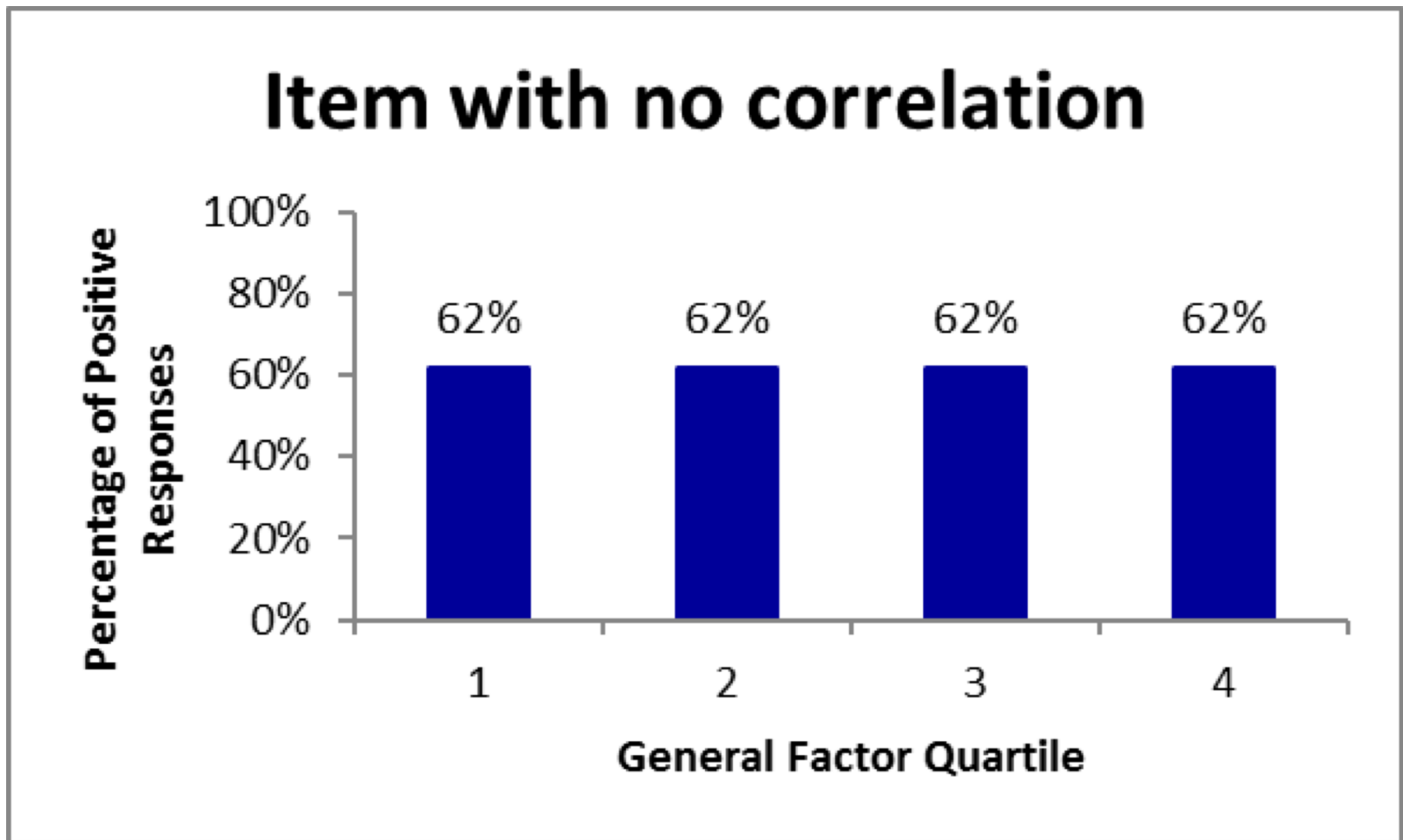
# Study 2: Construct Validity

- ## Results:
  - ### CFA using 2012 data
    - Six-factor higher-order model had excellent fit using maximum likelihood est.
      - NFI = .940; TLI = .936; CFI = .940; RMSEA = .080
    - Orthogonal model had lower fit
      - NFI = .914; TLI = .909; CFI = .914; RMSEA = .095
    - Oblique model had fit that was similar to higher order (suggests presence of general factor)
      - NFI = .941; TLI = .937; CFI = .941; RMSEA = .079
    - Using unweighted least squares estimation increased fit
    - Similar findings at the organizational level
  - ### Interesting factor factoids
    - General factor accounts for 53% variance at individual level, 64% at organizational level
    - General factor accounts for equivalent of 38 out of 71 items at individual level (45 at organizational level)
    - At individual level, average loading of items on general factor is .72 (range .45 to .86); at organizational level .77 (range .31 to .95)

# Study 2: Construct Validity

- Results:
  - More interesting factor factoids
    - Looked at residualized loadings using Schmid-Leiman decomposition
      - This is the loading of each item on the six non-general factors excluding all variance that is attributable to the general factor
    - Items have smaller loadings on six specific factors than general factor
    - At individual level, average loading of items on specific factors factor is .30 (range .15 to .55) at organizational level .28 (range .05 to .51)
    - Created expectancy charts for senior leaders showing relationship between general factor scores and percentage of positive responses to FEVS items
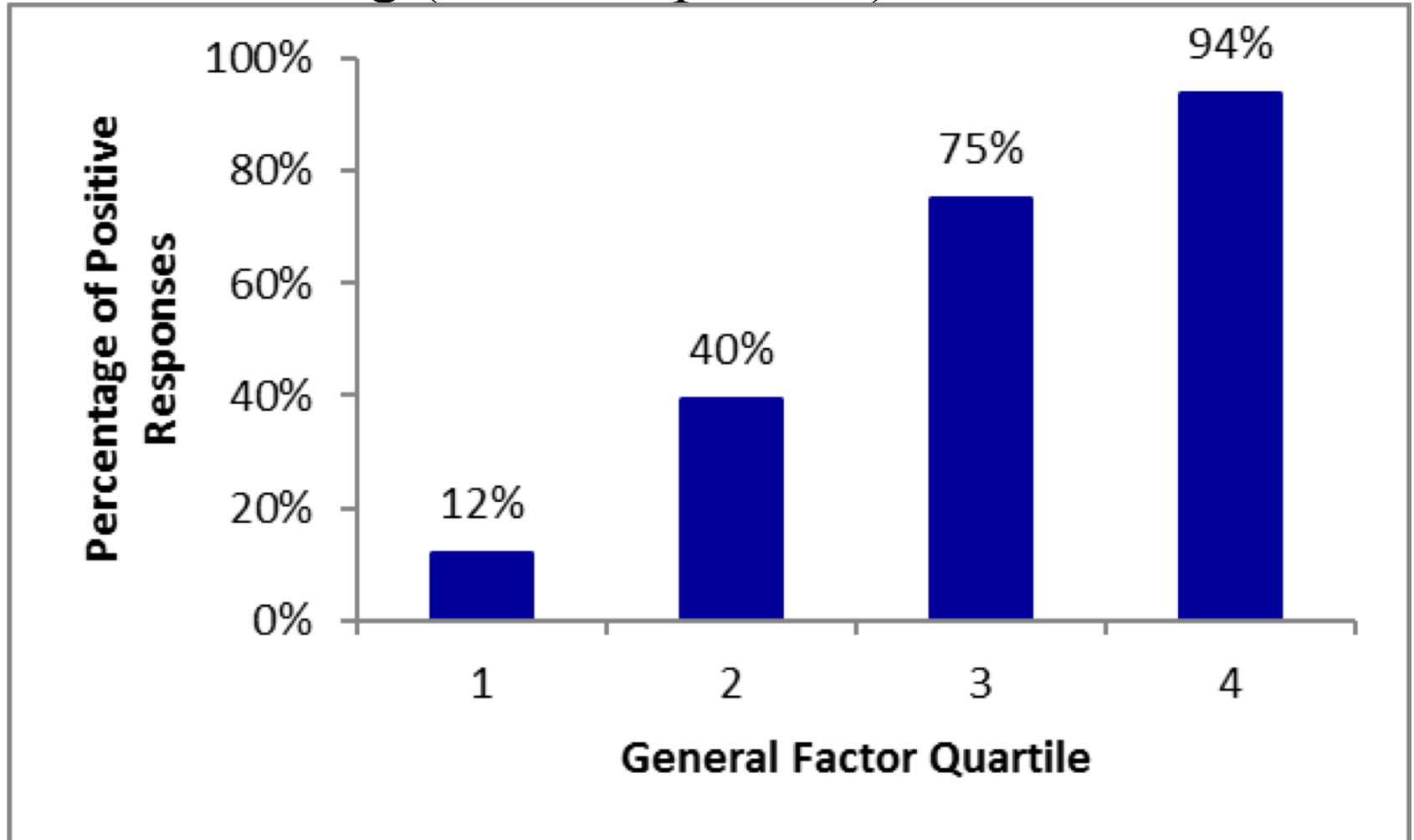      - This was done using data from the 2013 FEVS and only the 39 items in the HCAAF

# Study 2: Construct Validity

Expectancy chart for fictitious item with general factor loading of zero (and 62% positive) – Baseline control
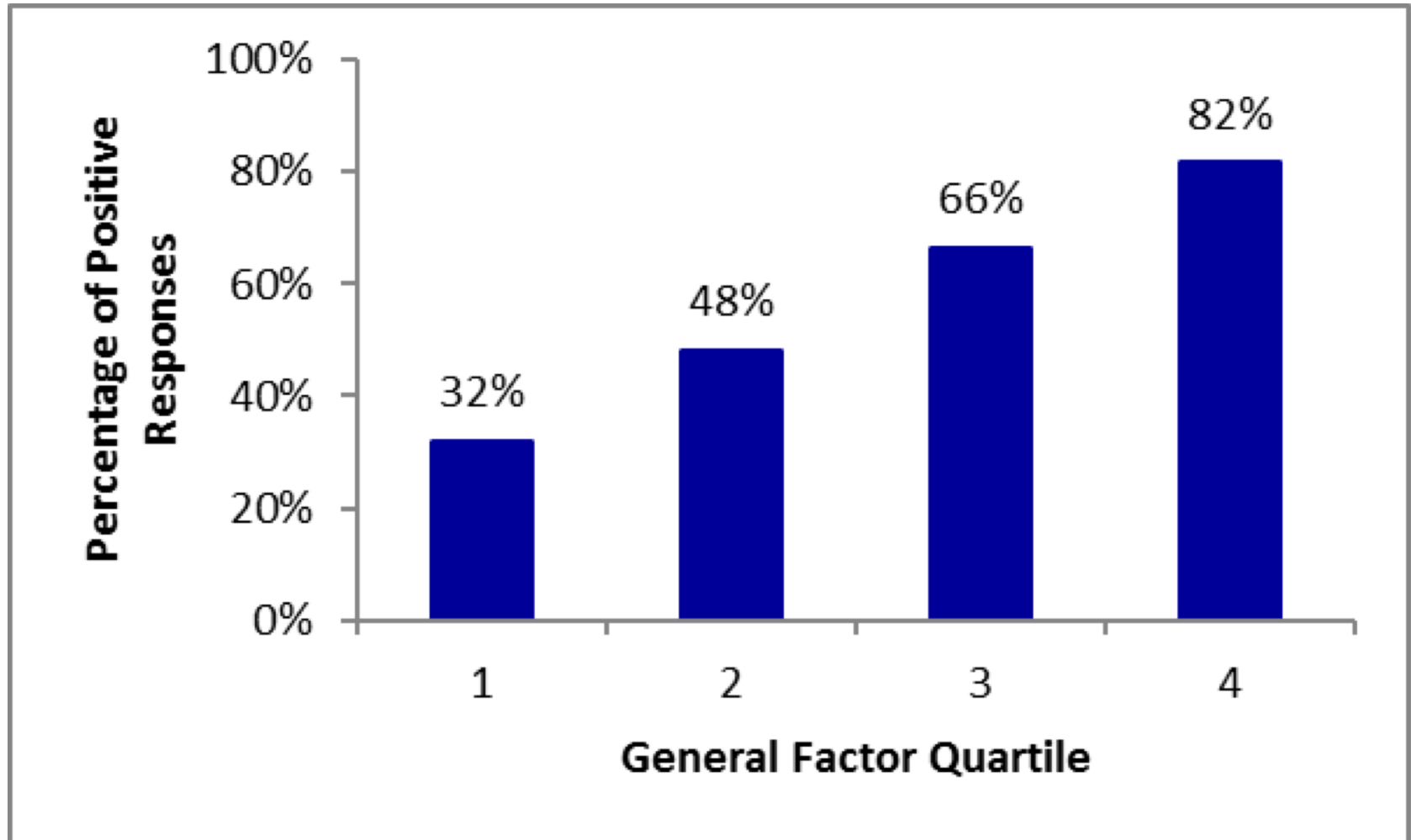


## Item with no correlation

# Study 2: Construct Validity

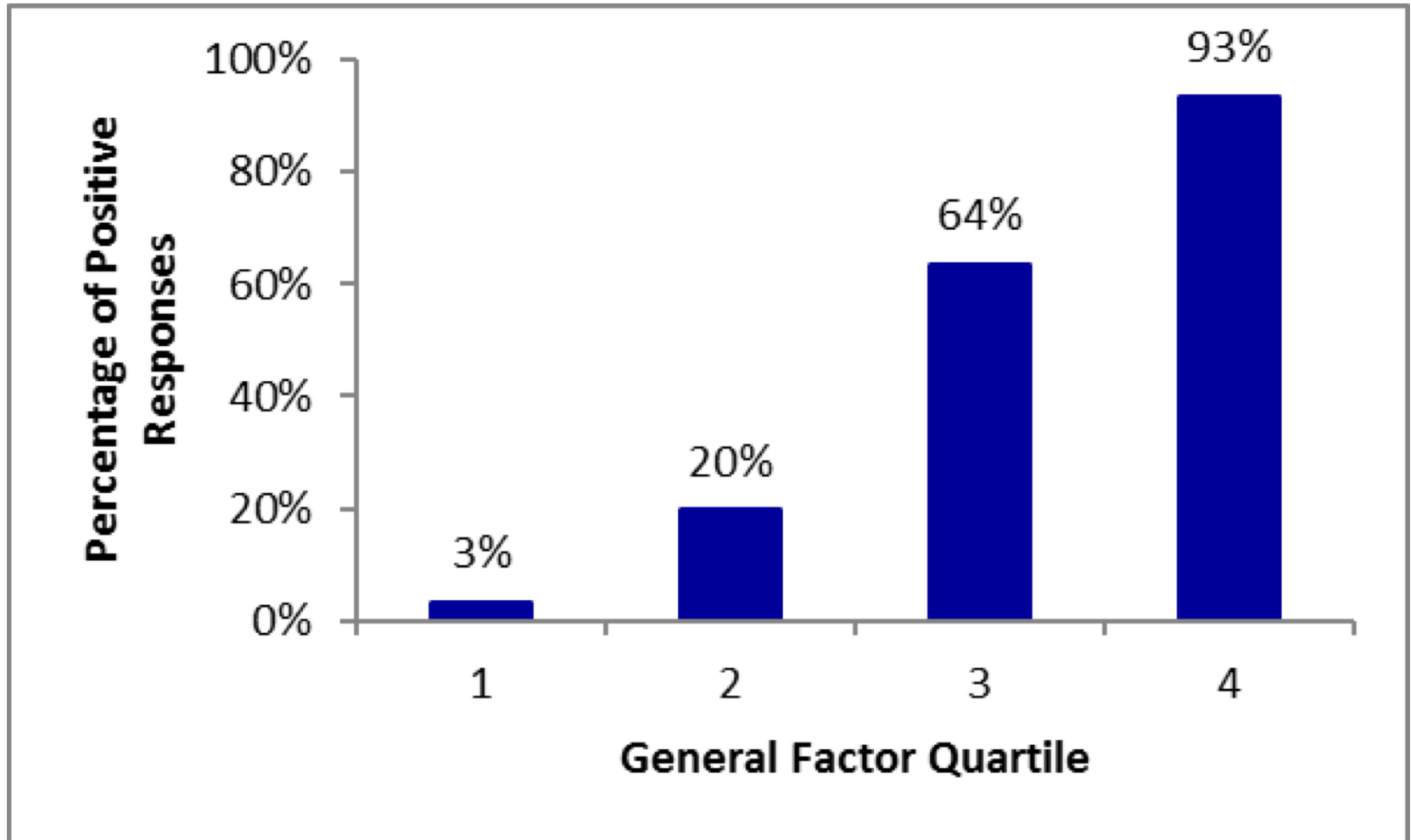Expectancy chart for item with median general factor loading (and 58% positive)

# Study 2: Construct Validity

Expectancy chart for item with lowest general factor loading (and 58% positive)

# Study 2: Construct Validity

Expectancy chart for item with highest general factor loading (and 48% positive)

# Study 3: Unit-level criterion-related validity

- Dataset:
  - 2012 FEVS administration
  - Aggregated (i.e., item mean) FEVS data from 298 subcomponents as predictor
  - 2012 Turnover rates (from FedScope) as criterion
    - Divided # voluntary quits by total number of employees

- Results:
  - Computed scores on general factor using PCA and PAF; both predicted turnover rates in expected direction
    - PCA: $r$ = -.133, $p$ = .022
    - PAF: $r$ = -.138, $p$ = .017
    - Similar to Heavey et al. (2013) meta-analytic correlation between unit-level job satisfaction and unit-level turnover (-.10); within their 95% confid. Interval
  - SEM using general factor to predict turnover
    - $\beta$ = .14 using both MLE (NFI = .83; TLI = .83; CFI = .85) and ULS (NFI = .99)

# Study 4: Individual-level criterion-related validity

- ## Dataset:

  - Study of clerical and technical Federal employees in 1993-1994

  - A total of 4,256 employees had complete data for our variables

  - 18-item job satisfaction survey

    - Of the 18 job-satisfaction items, 12 were identical or nearly identical to items in the FEVS; 4 were similar to existing FEVS items; 2 were not in the FEVS at all.

  - Job performance

    - An adaptation of the Descriptive Rating Scale, which was used in the GATB validation

    - 7 items measuring quantity and quality of work, job knowledge, creativity, positive service orientation, contribution to organizational goals, and the variety of assignments the employee could perform

      - Measured a general factor per a CFA (NFI = .99, CFI = .99, RMSEA = .06)

# Study 4: Individual-level criterion-related validity

- Results:
  - Factor structure of job satisfaction measure
    - Found evidence of four specific factors (using additional cases for which criterion was missing)
    - A dominant general factor also existed
  - Validity
    - General factor correlated with supervisory ratings of job performance
      - Mean correlation across 7 DRS items was .19 for PCA and .20 for PAF
      - Observed correlations similar to Judge at al. (2001) meta-analytic uncorrected r of .18.
    - SEM yielded $\beta$ = .282 for higher-order four-factor model (NFI = .98; TLI = .98; CFI = .98; RMSEA = .08) and $\beta$ = .262 for one-factor model (NFI = .96; TLI= .95; CFI = .96, RMSEA = .11)
    - Correlated individual job satisfaction items with the 7 DRS items
      - Mean correlations were .10; after partialling out general factor, mean correlation dropped to less than .01.

- Conclusion:
  - A subset of FEVS items and FEVS-like items have criterion-related validity
    - Most of the validity appears to be due to the general factor

# Conclusions

- FEVS has excellent *intra*-rater reliability
  - High internal consistency
- Little evidence of nesting at available groupings
- FEVS data appear to much more strongly be influenced by individual differences in job satisfaction than by agency or component-level effects
- There is a dominant general factor in the FEVS data
  - Six additional factors are present
  - Item results significantly impacted by general factor and six factors
  - This clouds interpretation of the items
    - An item's variance depends on general factor, specific factor, and unique content

# Conclusions

- FEVS has criterion-related validity
  - Unit-level FEVS results correlate with unit-level turnover
  - Subset of FEVS and related items predict job performance
- Next steps
  - Local criterion-related validity study of subset of FEVS items and job performance is underway
- Areas for future research
  - Unit-level test-retest
  - Reliability of unique variance in items
  - Individual-level turnover validity
  - Individual-level test-retest reliability
  - Shortening FEVS
- Causality?
  - Does action planning → increase FEVS scores → increase performance?

# Conclusions

- Recommendations for practitioners
  - Hard to interpret item-level survey results
    - Multiple factors influence results
    - We often see that items have similar patterns when comparing different units
  - Shorter surveys might be better
    - Less redundancy
    - Increase response rate
  - FEVS has criterion-related validity
    - Unit-level turnover
    - Individual-level performance

# Questions and Comments from the Audience