Content Valid Composites: The Empirical Elegance of Unit Weights

IPMAAC 2003

Maury Buster, Ph.D., State of Alabama

Philip Bobko, Ph.D., Gettysburg College

Philip Roth, Ph.D., Clemson University

Background

- Scant research in applied personnel management regarding test composite weighting strategies when using a content validation approach
- Much of the research on weighting methods has occurred in the statistical literature comparing regression and unit weighting techniques
- The applied literature has focused primarily on the effects of combining various types of predictors rather than how to combine the components within a type of predictor
- Common reliance on content validity, particularly in the public sector (e.g., job knowledge tests, T&Es, work sample tests, etc.)

Background (Continued)

- Under the content validation approach there are three general areas from which test composite weights are most often drawn

    1. Job analysis data

    2. Direct SME weights

    3. Unit weights

Study

- We examine the relationship between the three weighting methods across various jobs and exams

  - Jobs include:

    Programmer Analyst Associate
    Programmer
    Information Technology Operations Technician
    Right of Way Specialist (ROWS)
    Senior Right of Way Specialist (SROWS)
    Engineering Assistant (EA)
    Civil Engineer (Construction specialty) (CE-C)
    Civil Engineer (Design specialty) (CE-D)

Study (Continued)

- The jobs vary in complexity, number of incumbents, and number of qualifying applicants

- The exams vary from written multiple choice to 4 hour work samples

- We first provide examples of the 3 methods described previously: job analysis weights, direct SME weights, and unit weights

- We then report both the correlation between unit weights and the various other weighting methods and the change in potential adverse impact through investigation of the standardized difference (d) statistic

d Formula

- The d statistic is often referred to as the “standardized difference”
- It is defined as the difference between group means (e.g., White versus Black) divided by the pooled standard deviation of the groups (e.g., sample weighted, within-group standard deviation)
  - For example, a d of .33 means that the two groups differed, on average, by one third of a standard deviation

d Formula (Continued)

$$d = \frac{\overline{X}_W - \overline{X}_B}{\sqrt{\dfrac{(n_W - 1)(s_W^2) + (n_B - 1)(s_B^2)}{(n_W + n_B - 2)}}}$$

Where:

$\overline{X}$ = Mean score for the particular group sample (e.g., White or Black)

n = Sample size for the particular group

s^2 = Variance for the particular group sample

Example – Job Analysis Weights

Steps:

1. Link each exam item to the qualifying KSAs
2. Within each item, add up the importance scores associated with each linked KSA
3. Within each exam section, add up the importance scores across items, and then divide by the number of items

Importance Scale:

How important is this KSA for acceptable job performance?

0 = Not important	3 = Very important
1 = Slightly important	4 = Critical
2 = Important	

Example – Job Analysis Weights (Continued)

Example:

- Assume Section 1 has two items
- Item 1 links to 4 KSAs with importance scores 2.2, 2.4, 2.8, and 2.1
- Item 2 links to 3 KSAs with importance scores 2.3, 2.4, and 2.9

Section 1 weight equals:

$$\frac{(2.2 + 2.4 + 2.8 + 2.1) + (2.3 + 2.4 + 2.9)}{2} = \frac{17.1}{2} = 8.55$$

Example – Direct SME Weights

Steps:

1. Provide SMEs with a copy of the exam section to be rated and a copy of the qualifying KSAs

2. Allow SMEs to review the exam section and then respond to the rating scale

3. The mean rating across SMEs serves as the section weight

Example – Direct SME Weights (Continued)

Rating scale:

1 = The questions in this section are <u>minimally</u> related to successful job performance
2 = The questions in this section are <u>moderately</u> related to successful job performance
3 = The questions in this section are <u>strongly</u> related to successful job performance
4 = The questions in this section are <u>very strongly</u> related to successful job performance

Example - Unit Weights

Steps:

1. Standardize the x sections or components/subscores of the exam

2. Sum the resulting standardized scores

NOTE: Standardization is essential to obtaining unit weights. Summation of raw scores will result in a composite score weighted by the standard deviation of each section or component/subscore

Description of the Various Classes

Programmer Analyst Associate

- Beginning through mid-level professional programming and analysis work
- Examinees: $N_B = 19$, $N_W = 21$

Programmer

- Entry-level through full performance computer programming work
- Examinees: $N_B = 54$, $N_W = 39$

IT Operations Technician

- Entry through full performance operations support for data and voice communications equipment, mainframe and client server operations, and network operations
- Examinees: $N_B = 88$, $N_W = 59$

Description of the Various Classes (Continued)

SROWS

- Technical and professional supervisory work in relocation assistance, property management, and securing title to property needed for departmental purposes
- Examinees: $N_B = 10$, $N_W = 14$

ROWS

- Technical and professional work in relocation assistance, property management, and securing title to property needed for departmental purposes
- Examinees: $N_B = 33$, $N_W = 52$

EA

- This is sub-professional work assisting on engineering projects and related activities
- Examinees: $N_B = 266$, $N_W = 430$

Description of the Various Classes (Continued)

CE – Construction

- Paraprofessional technical work in an area of civil engineering which may not include supervisory duties upon entry, but will involve supervisory duties at the advanced level
- Examinees: $N_B = 4$, $N_W = 50$

CE - Design

- Paraprofessional technical work in an area of civil engineering which may not include supervisory duties upon entry, but will involve supervisory duties at the advanced level
- Examinees: $N_B = 5$, $N_W = 11$

Correlations Between Unit Weights and Other Weighting Methods

	r Job Analysis and Unit Weights	r Direct SME and Unit Weights
Programmer Analyst Assoc.	0.98	NA
Programmer	0.99	NA
IT Operations Technician	NA	0.97
SROWS	NA	0.99
ROWS	NA	0.99
EA	0.99*	0.99
CE-C	NA	0.99
CE-D	NA	0.99

* There were actually 2 different job analysis weights applied and both correlated 0.99.

d Values Corresponding to the Various Weighting Methods

	d Job Analysis Weights	d Direct SME Weights	d Unit Weight	Diff JA-Unit	Diff SME-Unit
Pr. An. As.	0.45	NA	0.51	-0.06	NA
Programmer	0.96	NA	1.00	-0.04	NA
IT Op. Tech.	NA	0.58	0.59	NA	-0.02
SROWS	NA	0.49	0.48	NA	0.01
ROWS	NA	0.77	0.82	NA	-0.04
EA	0.73	0.72	0.72	0.01	0.00
CE-C	NA	0.90	0.88	NA	0.03
CE-D	NA	1.45	1.30	NA	0.15
Weighted Average				0.001	0.001

Concluding Remarks

- The average differences (d) between unit weighting and the other weighing methods is miniscule (0.001)
- Additionally, the correlation between unit weighting and the other various weighting methods is very high (> 0.99 in most cases), indicating approximately equivalent ranks
- Unit weights have no sampling error / no variation from one group of SMEs to another; they are stable
  - Oftentimes we have small N's in our job analyses and/or our SME weights, thus, the above point is non-trivial
- Unit weights themselves are not influenced by outliers

Concluding Remarks (Continued)

- Unit weights are much more straight forward to apply and less time consuming
  - Especially in areas with high staff turnover
  - With non-I/O background personnel
- To the extent that task-based questionnaires, etc. have more components, you can expect very similar results
  - As the number of sections/subscores increases, the strength of the unit weighting method actually increases.

References

Arthur, W., Doverspike, D., & Barrett, G. (1996). Development of a job analysis-based procedure for weighting and combining content-related tests into a single testy battery score. Personnel Psychology, 49, 971-985.

Bobko, P. (1995). Correlation and regression: Principles and applications for industrial/organizational psychology and management. New York: McGraw-Hill, Inc.

Bobko, P, Roth, P.L., & Potosky, D. (1999). Derivations and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors, and job performance. Personnel Psychology, 52, 561-589.

Claudy, J.G. (1972). A comparison of five variable weighting procedures. Educational and Psychological Measurement, 32, 311-322.

Dorans, N., & Drasgrow, F. (1978). Alternative weighting schemes for linear prediction. Organizational Behavior and Human Performance, 21, 316-345.

References (Continued)

Einhorn, H.J., & Hogarth, R.M. (1975). Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 13, 171-192.

Ghiselli, E.E., Campbell, J.P., & Zedeck, S. (1981). Measurement theory for the behavioral sciences. San Francisco: W.H. Freeman and Co.

Ree, M., Carretta, T., & Earles, J. (1975). In top-down decisions, weighting variables does not matter: A consequence of Wilk's theorem. Organizational Research Methods, vol 1, no. 4, 407-420.

Ryan, A.M., Ployhart, R.C., & Friedel, L.A. (1998). Using personality testing to reduce adverse impact: A cautionary note. Journal of Applied Psychology, 83, 298-307.

Sackett, P.R., & Roth, L. (1996). Multistage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. Personnel Psychology, 49, 549-572.

References (Continued)

Schmidt, F.L. (1971). The reliability of differences between linear regression weights in applied differential psychology. Educational and Psychological Measurement, 84, 751-758.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 83, 213-217.