DOCUMENT RESUME

ED 281 074 CG 019 807

AUTHOR Camara, Wayne J.
TITLE The Equivalence of Rater Sources on Job Analysis Ratings.
PUB DATE 25 Aug 86
NOTE 12p.; Paper presented at the Annual Convention of the American Psychological Association (94th, Washington, DC, August 22-26, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluators; *Generalizability Theory; *Interrater Reliability; *Job Analysis; Job Skills; Research Design; *Validity

ABSTRACT

Previous efforts to investigate the equivalence of rating sources for job analysis ratings have reported conflicting results. In the present research, correlational and generalizability analyses were conducted to examine the equivalency of rating sources for over 70 state civil service job classifications. Incumbent and supervisor ratings (N=697) were examined, along with ratings of 3 trained experts derived from narrative job descriptions. Separate survey instruments were individually developed to obtain objective skills and abilities for each job classification. Pearson correlation coefficients reported significant reliability for ratings across each rating source, and significant convergent validities across sources. Results of the generalizability analyses were inconclusive concerning the similarity of ratings, however, variance attributed to rating sources was negligible. The results indicated that rating source did not have a significant effect on job analysis ratings when using an ability-oriented instrument. Ratings provided by incumbents, supervisors, and experts were similar in 65 of the 70 analyses. These findings support Smith and Hakel's (1979) belief that rating source does not make any practical difference in job analysis rating. (NB)

THE EQUIVALENCE OF RATER SOURCES ON JOB ANALYSIS RATINGS

Wayne J. Camara

Human Resource Research Organization

Paper presented at the 94th Annual Convention of the American Psychological Association. Washington, D.C.

August 25, 1986

Abstract

Previous efforts to investigate the equivalence of rating sources for job analysis ratings have reported conflicting results. In the present research, correlational and generalizability analyses were conducted to examine the equivalency of rating sources for over 70 state civil service job classifications. Incumbent and supervisor ratings were examined, along with trained experts' ratings derived from narrative job descriptions. Separate survey instruments were individually developed to obtain objective skills and abilities for each job classification. Pearson correlation coefficients reported significant reliability for ratings across each rating source, and significant convergent validities across sources. Results of the generalizability analyses were inconclusive concerning the similarity of ratings, however variance attributed to rating sources was negligible. Additional credence is attributed to Smith and Hakel's (1979) belief that rating source does not make any practical difference in job analysis ratings.

Literature Review

The purpose of this study was to investigate the effects of different sources of job analysis ratings. The few studies that have examined such effects have used the traditional correlational approach (Burt, 1980); this study utilized the correlational approach and is the first to apply generalizability theory to such comparisons. Specifically this paper

attempts to examine the effects of rater source by (1) comparing correlational and generalizability approaches, (2) correcting previously inappropriate applications of the correlational approach with a prescribed method for future analyses, and (3) utilizing a large number of raters and different job classifications to increase generalizability of these findings. While several studies in the past have investigated the reliability of ratings, few have systematically examined the effects of different sources of ratings, and the more general concern over the equivalency of rating sources (Burt, 1980).

The reliability and validity of job ratings is essential, since occupational choice and organization choice decisions rely of the accuracy of the information available to the decision maker. Smith and Hakel (1979) indicate that there is no practical difference between rating sources, including naive raters, in terms of reliability and convergent validity. In comparing the ratings on several dimensions of occupations, with the Position Analysis Questionaire (PAQ), they found mean reliability coefficients ranged from .49 to .63 for each rating source. Much of their support for the equivalence of rating sources is based on the reported high convergent validities, i.e., the average correlation between pairs of raters from two different rating sources ($\bar{X}$ = .92).

However, Cornelius III, Denisis, and Blencoe (1984) note that individual jobs are the object of measurement in such a study, and reliabilities and convergent validities must be obtained at the level of the

individual job, not across jobs as conducted by Smith and Hakel (1979). They used a limited 9 job replication study with a sample of only 13 raters with the corrected method and lower convergent validity ($\bar{X}$ r =.58) resulted.

Several other attempts to measure different rater sources (Jones, Main, Butler & Johnson, 1982; Burt, 1980) have only continued to confuse the picture concerning a "shared stereotype." One limitation of previous studies is their almost exclusive use of the PAQ as the measurment instrument. Properties of the PAQ may artifically increase reliability and validity among raters (Smith & Hakel, 1979). Large numbers of items scored as "Does Not Apply" for specific jobs may have increased validities in these studies. This study attempted to investigate the equivalency of rater sources with an ability-oriented job analysis instrument.

Methods

Seventy job classifications that are listed in the professional or technical areas of the Illinois Civil Service System were utilized for this study. Each job was classified in one of seven subgroups based on job content.

Three rater sources were employed: incumbents, supervisors, and ratings of narrative job descriptions by expert raters. Initially, expert raters, comprised of three mid-level administrators at the Personnel Services Office of a major Illinois University, were trained in identifying objective abilities and skills required of several out-of-date job classifications.

Experts rated each of the 70 job classifications, identifying skills and abilities needed for each job. Descriptive phrases for each of four (0-3) points were developed through consensus agreement to help "anchor" the scale. Based on this initial analysis, separate surveys were developed for each job classification that contained items measuring each skill and ability required. Therefore, the number of items in each survey differed. All incumbents and supervisors employed in these job classifications were sent these surveys. The return rate for these surveys exceeded 60%, with a sample of 697 incumbents and supervisors across all jobs. The mean number of surveys returned for each job by incumbents and supervisors was 6.79 and 3.07, respectively. Each job classification was also analyzed by the three expert raters, providing a mean of 13 ratings per job.

Results

Reliability of the job analysis ratings for each rating source were computed at the level of the individual job. All possible pairwise correlations between raters within each source were computed and averaged for each job. The mean and median correlations for each of the seven job content areas, and the total of all 70 jobs were computed in a similair fashion. However, mean values for each job were weighted by the number of raters and items (on each specific survey). This weighting procedure was required since different numbers of raters and different numbers of items were used for jobs classifications. Pearson correlation coefficients across all jobs were .579 for narrative ratings, .466 for incumbents, and .440 for

supervisors. Median correlations were .540 for narrative ratings, .543 for incumbents, and .562 for supervisors. All three ratings sources were significant ($p<.001$) with mean and median figures. While these correlations are moderately smaller than those obtained from earlier studies (McCormick, et. al., 1972; Burt, 1980) they are relatively similar to correlations in studies investigating equivalency of rating source and narrative ratings (Jones, et. al, 1982; Smith and Hakel, 1979; Cornelius III et.al., 1984).

Convergent validities were computed in a similar fashion, with all pairwise correlations between rating sources computed and then averaged for each job. These convergent validities ranged from .477 to .518 using Smith and Hakel's (1979) method for comparison. Table 1 illustrates the reliabilities and convergent validities for each of the seven content areas of jobs. Differences between methods across jobs were not substantial.

Insert Table One About Here

Generalizability analysis was used to examine the relative amount of variance contributed by each of the factors used in this study (i.e., items, rating source, raters within rating source, interactions). Separate generalizability analyses were performed for each of the 70 job classifications using an unbalanced fixed effects, items crossed by raters who are nested within rating method, analysis of variance (ANOVA) model. Separate analyses were used to allow detailed examination of the instrument's ability to match an individual with each job classification.

Generalizability coefficients, estimated variance components, and error terms for each unbalanced design, with different raters and items, were computed through procedures outlined by Brennan (1982, p.111). While only 10 of 70 generalizability coefficients ($\hat{\rho}^2$) were within the criteria suggested by Cardinet et al.,(1982) of .80. over half of the coefficients exceeded .50 (ranging from .26 to . 87).

Estimated variance components show that rating source was not a significant factor in the ratings (mean o rs/RS=.0244). Table 2 illustrates the average estimated variance components derived from the generalizability analyses across the 70 jobs. Estimated variance components for items and jobs accounted for the greatest proportion of variance. However, unaccounted for error variance was also substantial for some jobs. Idiosyncratic differences in the interpretation of items by individual raters across rater sources could explain this error variance. The primary method for reducing this would involve instituting some type of rater training. Because variance contributed by rater source, and raters are so small across analyses, replications of this study would probably not lead to significantly different estimates for these sources (Doverspike, Carlisi, Barrett, & Alexander, 1983).

Insert Table Two About Here

Discussion

Results indicate that rating source does not have a significant effect on job analysis ratings when using an ability-oriented instrument. Ratings provided by incumbents, supervisors, and experts (based on narrative descriptions) were similar in 65 of 70 analyses ($\hat{\sigma}$ rs/RS <.07). However, this study does acknowledge that the true equivalency of rating sources has not been met when using strict measurment criteria (Gulliksen, 1968).

Consistent results derived from the generalizability analyses suggest that rating source does not contribute significant variance in job analysis ratings. The correlations and convergent validities, within and between sources, support this statement. Higher reliabilities have been occassionally obtained with the PAQ, however these may be caused by excessive use of "does not apply items", leniency error, job level, and generic design of the instrument. This study examined specific abilities in jobs with subtle differences (Accountant 1 vs Accountant 2, etc.).

In summary, no practical difference can be found between rating sources. The importance of this study lies in the findings that accurate job analytic data can be obtained from several sources for a wide variety of jobs. Narrative job descriptions provided an accurate source of job data within these strictly controlled conditions. Further studies might do well to address the "shared stereotype" explanation for rater source similarities with instruments other than the PAQ. Further, investigation of the

equivalency of rating sources with generalizability theory, and multivariate analyses is needed in areas within industrial/organizational psychology.

References

Brennan, R. (1982). Elements of generalizability theory. Iowa City; ACT Programs.

Burt, R. (1980). The equivalency of job analysis agents: a methodological reappraisal. Unpublished doctoral dissertation, University of Tennesse at Knoxville.

Cornelius III, E., Denis, A.S., & Blencoe, A.G. (1984). Expert and naive raters using the PAQ: does it matter? Personnel Psychology, 37, 453-464.

Doverspike, D., Carlisi, A.M., Barrett, G., & Alexander, R. (1983). Generalizability analysis of a point-method job evaluation instrument. Journal of Applied Psychology, 68, 476-483.

Guilliksen, H. (1968). Methods for determining equivalence of measures. Psychological Bulletin, 70, 534-544.

Jones, A., Main, D., butler, M., & Johnson, L. (1982). Narrative job descriptions as potential sources of job analysis ratings. Personnel Psychology, 35,813-827.

Smith, J. & Hakel, M. (1979). Covergence among data sources, response bias, and reliability and validity of a structures job analysis questionnaire. Personnel Psychology, 32,677-692.

Table 1
Interrater reliability by rater source and job content area (correlations were calculated for each individual job. Mean correlations were then weighted by items rated, and number of pairwise correlations).

Job Content Area	Incumb	Super.	Raters	R x S	R x I	S x I
Mean r 1.) Business	.435	.409	.562	.442	.486	.405
N	958	458	23	286	510	1270
Sig.L	.001	.001	.001	.001	.001	.001
Mean r 2.) Commun.	.538	.717	.718	.696	.569	.669
N	112	49	14	80	194	109
Sig. L	.001	.001	.001	.001	.001	.001
Mean r 3.)Science	.444	.572	.544	.431	.503	.489
N	352	106	16	164	201	224
Sig. L	.001	.001	.001	.001	.001	.001
Mean r 4.)Technical	.630	.647	.574	.613	.634	.569
N	231	37	14	102	221	199
Sig.L	.001	.001	.001	.001	.001	.001
Mean r 5.) Education	.536	.895	.564	.661	.595	.756
N	35	6	8	41	78	48
Sig.L	.001	.001	.001	.001	.001	.001
Mean r 6.) Arts	.610	.483	.965	.660	.733	.425
N	62	22	7	54	91	86
Sig.L	.001	.001	.001.	.001	.001	.001
Mean r 7.) Health	.417	.528	.173	.133	.284	.427
N	1939	322	21	251	525	1638
Sig.L	.001	.001	NS	NS	.01	.001
Mean r TOTAL	.466	.440	.579	.477	.518	.432
N	511	206	18	176	336	616
Sig. L	.001	.001	.001	.001	.001	.001

N = mean number of pairwise correlations computed between raters X n for each job.
n = mean number of items rated by each pair of raters.

Table 2
Average Estimated Variance Components for Generalizability Analysis of 70 Job Classifications *

Variance Component	MS	DF	$\hat{\sigma}$ (x|s)
Items	6.709	15.44	.548
Rater Source	1.585	2	.024
Raters within Source	1.442	9.88	.088
Items x Rater Source	.572	30.88	.039
Raters within Source x Items	.427	145.86	.427

* An average of estimated variance components.