PERSONNEL PSYCHOLOGY
2005, 58, 1041–1054

EXAMINING RATING SOURCE VARIATION IN WORK BEHAVIOR TO KSA LINKAGES

LAURA E. BARANOWSKI AND LANCE E. ANDERSON
Caliber, an ICF Consulting Company

We examined Work Behavior to knowledge, skill, or ability linkage ratings for 9 jobs to determine the degree to which differences in the ratings were due to rater type. We collected ratings from incumbents and 2 types of job analysts: project job analysts (analysts knowledgeable of the job) and nonproject job analysts (analysts with very little or no knowledge of the job). In our analyses of the data, we calculated means, standard deviations, effect sizes, and correlations for each rater type, as well as compared the reliability of the ratings. We also estimated variance components for each job by conducting generalizability analyses (Brennan, 1983; Shavelson, Webb, & Rowley, 1989). Our findings indicate that the level of linkage ratings is similar across rater types, that it is important to obtain ratings from multiple raters regardless of rater type, and that ratings from job analysts may be more reliable than those of incumbents.

One commonly obtained type of job analysis rating is the "linkage rating." Linkage ratings may require the rater to indicate for each work behavior whether each knowledge, skill, or ability (KSA) is needed to perform that work behavior. Conversely, a linkage rating may involve a rater estimating the degree of the relationship between a work behavior and KSA. The results of linkage ratings, however gathered, are useful for operationalizing each KSA because these ratings indicate how the KSA is used to perform the job. Having such an understanding is important to content-oriented test development (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978).

Despite the need to obtain high quality linkage ratings, there has been little systematic examination of the factors that may affect the quality of these ratings. Hughes and Prien (1989) demonstrated that for one job,

The authors thank Maury Buster for providing valuable feedback on earlier drafts of this paper. We also thank three anonymous reviewers and section Editor James Smither for insight and comments that greatly improved the quality of this paper.

Correspondence and requests for reprints should be addressed to Laura E. Baranowski, Caliber Associates, 10530 Rosehaven Street, Suite 400, Fairfax, VA 22030-2840; LBaranowski@caliber.com

James Smither served as guest editor for this paper.

incumbent subject matter expert (SME) judgments on linkage ratings were highly correlated for only about 75% of the KSA statements and then suggested that certain characteristics of the KSA statement may affect the quality of the ratings. However, no research published to date has provided any guidance on how the quality of linkage ratings could be improved without changing the content of the KSA statements.

One factor that may affect the quality of linkage ratings is the type of rater performing the linkages. Two groups are commonly used to produce linkage ratings, SMEs (including both job incumbents and supervisors) and job analysts. Sometimes, the job analysts that are used to perform this type of rating do so after interviewing incumbents and observing the job being performed; other times, job analysts may be asked to perform these ratings without these exposures. It is unclear which type of rater would provide the best ratings. This study provides empirical data to help answer this question.

Using incumbents, their supervisors, or other SMEs to make linkage ratings may provide superior ratings because SMEs have a better understanding of the job's work behaviors (Landy & Vasey, 1991). Research suggests that the quality of SME job analysis ratings for other types of ratings can be attributed to the type of rating being made by SMEs. For example, Lindell (1998) examined task ratings and found better chance level of interrater agreement among raters regarding "task importance," but lower interrater agreement for "time spent" on tasks.

One advantage of having job analysts perform linkage ratings is that job analysts, through their training and experience in analyzing jobs, may have a better understanding of the relationship between work behaviors and KSAs. Job analysts also may have observed or spoken to a variety of incumbents and supervisors and therefore may be able to analyze the job considering the variety in job tasks and required KSAs among the various positions, where the incumbent may respond based on what they know about their position only. In addition, job analysts may be more readily available and willing to perform the tedious activity of rating each work behavior in terms of each KSA (this activity often involves making over 300 ratings!). We could find no research examining the quality of job analyst's linkage ratings. Research suggests that for other types of job analysis ratings, such as those of task frequency or task importance, job analysts who have greater knowledge of the job provide ratings with greater interrater reliability, and their ratings differ from job analysts with less knowledge of the job (DeNisi, Cornelius, & Blencoe, 1987; Friedman & Harvey, 1987). A finding that the quality of job analyst ratings is equal or greater than that of incumbents could mean greater flexibility for practitioners and improvement in the quality of job analysis data.

EXHIBIT 1
Example: Work Behavior Statements

Performing administrative duties.
Providing financial and technical advice/assistance to others.
Reviewing and analyzing information and legal documents.
Preparing metal for welding.

In the present study, we analyzed the work behavior to KSA linkage ratings of three types of raters: "incumbents," "project job analysts," and "nonproject job analysts." We utilized incumbent data collected from a large state agency, as well as data from job analysts with knowledge of the job because they had performed interviews/observations of the job class (i.e., project job analysts) and from job analysts with very little or no knowledge of the job class (i.e., nonproject job analysts). The data were analyzed to determine the quality of the linkage ratings and to identify the sources of variance in the ratings.

Method

Jobs

Nine jobs were selected from a larger set of jobs performed in a large state government in the southeastern United States. The jobs were selected to provide variability in the nature of work (i.e., work behaviors and KSAs) and in the type of job (i.e., blue collar vs. white collar and supervisory vs. nonsupervisory).

Work Behavior and KSA Statements

Each linkage matrix included work behavior and KSA statements that were unique to the job. A work behavior statement defined a set of related tasks. A list of the tasks associated with each of the work behaviors accompanied each linkage matrix. Project job analysts had written the work behavior, task, and KSA statements after researching background information on the job and performing job observations and job interviews with incumbents. Both incumbents and supervisors provided feedback on all of the work behavior, task, and KSA statements before they were rated.

On average, each matrix included six work behaviors and 49 KSAs. Each of the work behaviors encompassed a number of related tasks. The nine jobs varied in the number of linkage ratings per job, averaging 316 linkage ratings. Examples of the work behavior and KSA statements are provided in Exhibits 1 and 2.

EXHIBIT 2

Example: KSA Statements

Knowledge of various types and properties of paper.
Knowledge of basic statistics, such as computation of the mean and standard deviation.
Skill in providing information orally that is clear and understandable.
Ability to read and comprehend written material, such as a work order or an equipment operator manual.

EXHIBIT 3

Importance of KSA to Performing a Work Behavior: Five-Point Rating Scale

How important is this knowledge, skill, or ability for performing this work behavior?
1 = *Not important*
2 = *Slightly important*
3 = *Moderately important*
4 = *Very important*
5 = *Extremely important*

EXHIBIT 4

KSA to Work Behavior Linkage Rating Scale: Two-Point Rating Scale

Is this knowledge, skill, or ability used to perform this work behavior?
0 = *No*
1 = *Yes*

Scales

The linkage ratings were provided on one of the two scales. Raters rating on the first scale provided linkages on a five-point rating scale, whereas raters who rated on the second scale rated on a two-point rating scale. Four of the jobs were rated on the first scale and the other five jobs were rated on the second scale. The ratings scales are provided in Exhibits 3 and 4.

Incumbents

The incumbent raters were individuals who rated their jobs as part of a job analysis process. Our sampling strategy varied by job. For six of the jobs, we asked all job incumbents to complete the ratings for their jobs. For the remaining three jobs, we selected a stratified random sample of incumbents where race and employing agency were our strata. All incumbents who made ratings had been in the job at least 6 months.

Project Job Analysts

The project job analysts for each job were professional job analysts from a consulting organization who had considerable experience with the respective job because they were responsible for performing background research on the job, interviewing/observing incumbents, generating work behavior, task, and KSA statements, and authoring the job analysis report.

Nonproject Job Analysts

The nonproject job analysts for each job were professional job analysts from the same consulting organization as the project job analysts, however, they were unfamiliar with the job being rated. Despite unfamiliarity with the job to be rated, the nonproject job analysts were familiar with the linkage rating process because they were responsible for similar job analysis activities for unrelated jobs within the state government.

Procedure

All incumbent linkage data were collected up to a year prior to the commencement of this study. On average, eight incumbents completed linkage ratings as part of a job analysis process for test development purposes. Incumbents completed linkage ratings through supervised job analysis workshops or through mail-out surveys. Incumbents were provided written information about the job as they completed the linkages, including work behavior and KSA lists. Task lists were also provided to help define the work behaviors being rated.

Eighteen job analysts were selected to participate in the study. Analysts were selected based on their knowledge of the job being rated and might have served as a project analyst for one or more jobs and as a nonproject analyst for other jobs. All job analysts volunteered to complete linkage ratings as part of the study. On average, job analysts rated a total of three jobs as either a project or nonproject job analyst. However, no job analyst rated more than four jobs as a project job analyst or four jobs as a nonproject job analyst.

For each job, three project job analysts were selected based on their exposure to the job through the job analysis process. Project job analysts were initially identified by the fact that they had performed job observations and job interviews, written work behavior, task, and KSA statements, and authored the job analysis report. Project job analysts verbally confirmed they were familiar with the job they were to rate. Nonproject job analysts were randomly selected from the pool of job analysts who did not meet the criteria of a project job analyst and asked to rate their familiarity with the

jobs they were to rate on a scale of 1 = *I know nothing about this job class* to 5 = *I have performed this job or know someone well who has performed it*. If a nonproject job analyst indicated a 3 or higher (I have had moderate exposure to this job class [e.g., read about the job, moderate knowledge of tasks, and KSAs]), another analyst was selected as a nonproject job analyst for the job from the remaining pool of analysts. In our selection of job analyst raters for each job, we matched project and nonproject job analysts for job analysis experience and educational levels as feasible.

Before completing linkage ratings, project job analysts and nonproject job analysts were provided written information about the jobs they were to rate. Raters were instructed to use this written information as they completed their ratings. This information included work behavior, task, and KSA lists for the job, as well as the rating scale. This information was identical to the information provided to incumbent raters. All ratings were completed on a paper copy of the linkage matrix. Project job analysts and nonproject job analysts completed their ratings independently and provided the completed linkage matrix to the researchers.

To help us understand and interpret the results, we first calculated basic descriptive data on the ratings, including the mean and standard deviation for each rater type for each job. Next, we calculated effect sizes to assess the differences in the level of the ratings provided by the different types of raters. Then, we calculated the reliability of the mean ratings to have a measure of consistency using the Shrout and Fleiss (1979) ICC (3,k) formula, which is equivalent to Cronbach's coefficient α (Cronbach, 1951). Next, we applied the Spearman–Brown formula (Brown, 1910; Spearman, 1910) to the data to determine the reliability of a single rater and then estimated the reliability of the mean ratings for all rater types. We also calculated the correlation between each rater type to determine the similarity of the ratings in terms of their profile or "shape."

To allow us to examine the sources of variation in the ratings, we developed a generalizability model (Brennan, 1983; Shavelson, Webb, & Rowley, 1989), computed variance components, and examined the percentage of variance associated with various facets. This model is a nested, two facet mixed model where the random facet of "rater" is nested within the fixed facet of "type of rater." Calculation of the variance components for this model and comparison of their size allows us to better understand how the ratings vary. For example, if the variance component associated with "cell" is large, one can conclude that a great deal of the variation in the ratings is due to the different combinations of work behaviors and KSAs that were rated in each cell. If the variance component associated with type of rater is large, then one can conclude that the type of rater affected the level of the ratings provided for different cells. This model also allows for comparison of variance components due to

interactions among facets. For example, if the variance component associated with the interaction of cell and type of rater (i.e., ct) was substantial, then the results would suggest that raters of different types provide different levels of ratings for different cells. We calculated a separate generalizability analysis for each of the nine jobs.

Results

Mean Differences

Table 1 provides, for each type of rater for each job and each type of scale, respectively, the total number of ratings, total number of raters per rater type, the mean and standard deviation of the linkage ratings, the reliability of the mean ratings, the reliability of the single rater for each job, and the reliability of the mean ratings for all of the raters combined. The table also provides the effect size (*d*) showing the size of the difference between incumbent ratings and the ratings provided by the two types of job analysts.

Correlations

Table 2 provides the mean effect size differences for each possible pairing of types of raters for both rating scales. The mean effect size difference provides a summary of the differences in the level of the ratings provided by the different types of raters. This table also provides the mean correlation of linkage ratings between incumbents, project job analysts, and nonproject job analysts for both rating scales.

Variance Components

We conducted a generalizability analysis for each of the nine jobs. The results of these analyses are shown in Table 3. The key finding in this table is the large portion of the variance in the rating associated with the cell, or the particular work behavior to KSA combination being rated, compared to the relatively low variance associated with type of rater or rater.

Discussion

In response to limited empirical evidence examining the factors that affect the quality of work behavior to KSA linkage ratings, we conducted a study examining the impact of rater type on the linkage ratings provided for nine jobs.

TABLE 1

Descriptive Statistics, Mean Differences, and Reliability Analyses for Linkages

Job number	Number of cells	Type of rater	Number of raters	Mean	Standard deviation	*d* Relative to incumbents	Reliability of mean ratings	Reliability of a single rater	Reliability of mean ratings for all raters
Five-point rating scale									
1.	325	1	5	2.33	1.37	NA	.83	.49	.92
		2	3	3.11	1.51	.54	.80	.57	
		3	3	2.84	1.56	.35	.85	.65	
2.	574	1	2	4.00	.82	NA	.43	.28	.82
		2	3	2.70	1.19	−1.29	.70	.43	
		3	3	3.11	1.43	−.79	.73	.45	
3.	288	1	4	3.35	1.31	NA	.49	.19	.82
		2	3	2.44	1.47	−.65	.68	.42	
		3	3	3.10	1.39	−.19	.65	.39	
4.	320	1	6	2.98	1.47	NA	.83	.45	.85
		2	3	3.32	1.50	.23	.73	.45	
		3	3	2.83	1.52	−.10	.73	.45	
Two-point rating scale									
5.	342	1	18	.79	.41	NA	.89	.31	.92
		2	3	.66	.48	−.29	.66	.39	
		3	3	.74	.44	−.12	.60	.33	
6.	294	1	7	.68	.47	NA	.61	.19	.79
		2	3	.74	.44	.13	.61	.34	
		3	3	.77	.42	.20	.65	.39	

7.	304	1	14	.67	.47	NA	.84	.27	.89
		2	3	.63	.48	−.08	.70	.43	
		3	3	.60	.49	−.15	.68	.42	
8.	168	1	8	.73	.45	NA	.64	.18	.83
		2	3	.69	.46	−.09	.68	.42	
		3	3	.56	.50	−.36	.78	.54	
9.	231	1	5	.64	.48	NA	.70	.32	.84
		2	3	.65	.48	.02	.72	.46	
		3	3	.66	.48	.04	.66	.39	
Means									
Means for all jobs on five-point rating scale	377	1	4	3.17	1.24	NA	.65	.35	.85
		2	3	2.89	1.42	−.29	.73	.47	
		3	3	2.97	1.48	−.18	.74	.49	
Means for all jobs on two-point rating scale	268	1	10	.70	.46	NA	.74	.39	.85
		2	3	.67	.47	−.06	.67	.41	
		3	3	.67	.47	−.08	.67	.41	
Means for all jobs	316	1	8	NA	NA	NA	.70	.30	.85
		2	3	NA	NA	−.16	.70	.43	
		3	3	NA	NA	−.12	.70	.45	

Note. The type of raters are 1 = *Incumbent*, 2 = *Project Job Analyst*, and 3 = *Nonproject Job Analyst*. The Spearman-Brown formula was used to estimate the reliability for a single rater.

TABLE 2
Comparison of Different Types of Raters

Types of raters compared	Effect size differences		Correlation between mean ratings	
	Mean	*SD*	Mean	*SD*
Incumbents and project job analysts	−.16	.54	.56	.08
Incumbents and nonproject job analysts	−.12	.33	.49	.13
Project job analysts and nonproject job analysts	.08	.26	.70	.06

TABLE 3
Generalizability Analyses for Each Job

	1	2	3	4	5	6	7	8	9
MS_c	12.3410	.8658	.6702	4.1943	6.8039	1.0502	8.5114	.9163	.8796
MS_t	20.6980	8.3948	2.9289	588.6426	212.8241	2.4866	59.8275	5.0302	.0978
MS_{ct}	1.4770	.1798	.1725	1.0915	1.2586	.2110	2.5368	.1995	.1926
$MS_{r:t}$	48.7659	6.2975	4.1337	73.0492	55.0896	6.5870	93.8105	4.5337	4.8410
MS_e	.8799	.1046	.1460	.8053	1.2288	.1425	1.1038	.1312	.1367
VC_c	.9877	.0286	.0383	.3874	.5545	.0419	.4979	.0512	.0624
VC_t	.1405	.0008	−.0009	.3364	.1643	−.0020	−.0265	.0006	−.0056
VC_{ct}	.1629	.0094	.0061	.1072	.0089	.0103	.3583	.0146	.0152
$VC_{r:t}$	.1623	.0181	.0136	.1259	.1870	.0212	.2897	.0262	.0204
VC_e	.8799	.1046	.1460	.8053	1.2288	.1425	1.1038	.1312	.1367
PV_c	42%	18%	19%	22%	26%	20%	22%	23%	27%
PV_t	6%	0%	0%	19%	8%	−1%	−1%	0%	−2%
PV_{ct}	7%	6%	3%	6%	0%	5%	16%	7%	7%
$PV_{r:t}$	7%	11%	7%	7%	9%	10%	13%	12%	9%
PV_e	38%	65%	72%	46%	57%	67%	50%	59%	60%

Note. MS = mean squared; VC = variance component; PV = percentage of total variance; c = cell; t = type of rater; ct = interaction of cell and type of rater; r:t = rater within type; e = error.

In general, there were no consistent differences between types of raters in the mean of their ratings. For all jobs and rating scales, the mean overall effect sizes differences were minimal. Effect sizes were low for all jobs rated on the two-point scale. However, higher effect sizes for project and nonproject job analysts relative to incumbents were calculated for three of the four jobs rated on the five-point scale. Based on this data, one cannot judge conclusively whether the higher effect sizes for these jobs could be attributed to the use of the five-point scale or some characteristic of the job.

The correlations of the mean ratings indicate a moderate to large positive relationship between the ratings made by each of the rater types. The highest correlation was between the project job analysts and nonproject job analysts, with a .70 correlation. The next highest correlation was

between the incumbents and project job analysts, with a .56 correlation. Finally, between the incumbents and nonproject job analysts there was a .49 correlation. The finding of a strong correlation (i.e., $r = .70$) between the project analyst and nonproject analyst ratings suggests that even the "naïve" project analyst is able to provide ratings linking a work behavior to a KSA. Unlike other job analysis, activities that require extensive knowledge of the job, such as determining the importance of a work behavior, it is believed that an analyst who may not be familiar with the specific job but has experience with KSAs and the work behaviors that require those KSAs can complete the linkage rating task successfully.

When examined by rater type, the average interrater reliability of the mean of the ratings across all nine jobs was generally acceptable (i.e., $r_{xx} = .70$ for each type with an average of eight incumbent raters and three each of the project and nonproject job analysts). Across the four jobs rated using the five-point scale, interrater reliability was generally acceptable (i.e., mean $r_{xx} = .71$ with an average of four incumbent raters and three each of the project and nonproject job analysts). Across the five jobs rated using the two-point scale, interrater reliability for each rater type was also found to be generally acceptable (i.e., mean $r_{xx} = .69$ with an average of 10 incumbent raters and three each of the project and nonproject job analysts). The average interrater reliability was much higher when all raters were combined to estimate the mean rating ($r_{xx} = .85$). To better understand how the reliability of the ratings was affected by the type of rater independent of the number of raters, we calculated the reliability of the ratings from each type for a single rater. The mean single-rater reliability coefficient was .30 for incumbents, .43 for project job analysts, and .45 for nonproject job analysts. The single-rater reliability of incumbent ratings was equal to or less than that for both types of job analysts for each of the nine jobs. These findings are consistent with that of other research examining importance and level job analysis ratings where the data suggest that analyst ratings tend to be more reliable than incumbent ratings (Childs, Peterson, & Mumford, 1999).

Analyst ratings may have been more reliable for a variety of reasons such as:

- Analysts may have better understanding of the theoretical relationship between tasks and KSAs.
- Analysts may have spent more time making the ratings.
- Analysts may have a better understanding of the rating task.
- Analysts, lacking the detailed information known by incumbents, may make their ratings based on a heuristic of the job. Such a heuristic may provide for more consistent or reliable ratings, but the validity of such a heuristic would be unknown.

The data provided from this study does not provide a clear explanation of the reasons for the differences.

The generalizability analyses show that for every job, the largest portion of systematic variance in the ratings can be attributed to the cell (i.e., the different combinations of work behaviors and KSAs being rated). The next largest source of systematic variance in the ratings was the individual rater, although the tendencies of individual raters toward leniency or stringency accounted for less than half as much of the variance that could be attributed to individual cells. The next largest source of systematic variance was the interaction between cell and rater type, indicating that a small portion of the variation in the ratings is due to the fact that different types of raters rate different cells differently, however, the variance attributable to type of rater alone was negligible, indicating that different types of raters could be expected to provide ratings at roughly the same level.

Limitations of the Study

Throughout the research process, we identified four limitations to this study. The first limitation is that we have no measure of the accuracy of the linkage ratings. This makes it difficult to make recommendations on whether job analyst linkage ratings are actually more valid than those provided by incumbents. However, we do not view this as a fatal flaw as we were able to demonstrate across several jobs that job analyst ratings were more reliable, differed little from those of incumbents, and correlated reasonably well with the ratings of incumbents. The second limitation is that we did not include other types of raters in the study (e.g., supervisors). Supervisors may have a better perspective than both incumbents and job analysts in that they often have observed multiple individuals performing the job over a period of time. The third limitation is the fact that we used a relatively small set of individuals to conduct the job analyst ratings ($n = 18$), and the findings may not generalize well to other job analysts. The fourth and final limitation is that a large proportion of the variance in the ratings remains unexplained.

To address these limitations, future research must examine the impact of rater type on the accuracy of linkage ratings. We present two ways that this could be accomplished. The first assumes the existence of a rater group serving as the job content expert rating group. The mean ratings from the expert group could be correlated with the mean ratings from another rater group to determine a convergent validity coefficient. Accuracy may also be determined by identifying a subset of work behaviors that a group of incumbents and job analysts rate through a consensus discussion as either unequivocally linked or not linked to KSAs. The ratings for this subset of work behaviors could be compared with the ratings already collected to

identify differences (i.e., errors). To understand more about SME ratings, future research should collect and analyze ratings from supervisors. In addition, future research should consider other factors, such as the type of scale or the setting in which the ratings are provided (e.g., whether the ratings were obtained in a workshop or via the mail). Future research could also manipulate the amount of information provided to raters. Finally, we believe that the study should be replicated with job analysts working under different circumstances (i.e., job analysts from the government agency).

Implications for Practice

The findings of this study suggest several implications for practitioners performing job analysis:

- Reliable linkage ratings can be obtained for a variety of jobs regardless of rater type.
- Practitioners should NOT be reluctant to use job analysts to make linkage ratings and may, in fact, obtain more reliable ratings from job analysts.
- It is not necessary to get ratings from a large number of incumbents or job analysts in order to ensure that linkage ratings are reliable.
- One promising approach may be to have both incumbents and job analysts complete linkage ratings forms and use the results of both ratings for job analysis purposes. This may be especially beneficial for job analyses that are conducted on jobs with few incumbents.

Finally, we would also like to add that anytime the job analysts are gathering and using linkage rating data, it is important that the rating task is clear to raters and is easy to complete. Researchers should also ensure that raters have time to complete the task. Finally, researchers should monitor the quality of the data by calculating the reliability of the linkage ratings.

REFERENCES

Brennan RL. (1983). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.

Brown W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322.

Childs RA, Peterson NG, Mumford MD. (1999). Occupational descriptor covariates: Potential sources of variance in O*NET ratings. In Peterson NG, Mumford MD, Borman WC, Jeanneret PR, Fleishman EA (Eds.), *An occupational information system for the 21st century: The development of O*NET* (pp. 237–245). Washington, DC: American Psychological Association.

Cronbach LJ. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

DeNisi AS, Cornelius ET, Blencoe AG. (1987). Further investigation of common knowledge effects on job analysis ratings. *Journal of Applied Psychology, 72*(2), 262–268.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*(166), 38290–38315.

Friedman L, Harvey RJ. (1987). Can raters with reduced job descriptive information provide accurate Position Analysis Questionnaire (PAQ) ratings? PERSONNEL PSYCHOLOGY, *39*(4), 779–789.

Hughes GL, Prien EP. (1989). Evaluation of task and job skill linkage judgments used to develop test specifications. PERSONNEL PSYCHOLOGY, *42*, 283–292.

Landy FJ, Vasey J. (1991). Job analysis: The composition of SME samples. PERSONNEL PSYCHOLOGY, *44*(1), 27–50.

Lindell J. (1998). Relationship between organizational context and job analysis task ratings. *Journal of Applied Psychology, 83*(5), 769–776.

Shavelson RJ, Webb NM, Rowley GL. (1989). Generalizability theory. *American Psychologist, 44*(6), 922–932.

Shrout PE, Fleiss JL. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Spearman C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271–295.