

Investigator Characteristics as Moderators of Personnel Selection Research A Meta-Analysis

Craig J. Russell

Department of Management

Randall P. Settoon

Department of Management

Robert N. McGrath

Department of Management

A. Elyssa Blanton

Department of Management

Roland E. Kidwell

Department of Management

Franz T. Lohrke

Department of Management

Elton L. Scifres

Department of Management

George W. Danforth

Department of Management

ABSTRACT

Surrogate measures of original authors' reward context, primary motivation for doing the research, and knowledge, skills, and abilities (KSAs) were examined for their affect on criterion-related validities reported in the *Journal of Applied Psychology* and *Personnel Psychology* between 1964 and 1992. Number of years of experience (a surrogate KSA measure) displayed no moderating relationship. Type of organizational need (equal employment opportunity compliance, augmenting existing selection system, etc.) and investigator interests (e.g., theory testing) were related to criterion-related validities. Place of authors' employment (i.e., reward context) also displayed a moderating relationship (authors in private industry reported higher average validities in comparison with academics). Interaction effects on criterion-related validities were found between authors' experience and (a) place of employment and (b) primary motivation for conducting the research. Results are interpreted in view of possible differences in ability, motivation, and opportunity to do research across employment settings.

We thank James Combs and Michelle Dean for their assistance in data gathering and analysis. Philip Bobko, Stephen Gilliland, Kevin Mossholder, and Robert B. Stennett provided useful comments on drafts of the study, although all responsibility for any errors lies with the authors.

Correspondence may be addressed to Craig J. Russell, College of Business, Louisiana State University, Baton Rouge, Louisiana, 70803-6312.

Received: October 22, 1992

Revised: June 17, 1993

Accepted: June 18, 1993

One of the first steps in beginning a theory-driven study of performance prediction is a review of prior findings. Quantitative literature reviews of criterion-related validities in personnel selection (e.g., [Hunter & Hunter, 1984](#)) have contributed a powerful tool in systematizing this review effort. For example, F. L. Schmidt, J. E. Hunter, and their colleagues ([Hunter, 1986](#) ; [Hunter & Hunter, 1984](#) ; [Hunter & Schmidt, 1990](#)) presented convincing evidence that tests of general cognitive ability demonstrate positive criterion-related validity r_{xy} across a wide variety of jobs, organizations, and demographic groups. [Hunter \(1986\)](#) additionally speculated that cognitive ability will predict job performance independent of prior learning because individuals can never have learned every possible appropriate response for every possible job circumstance and must innovate in the face of new job demands. In support of this interpretation, [Hunter \(1986\)](#) reported evidence of higher criterion-related validity in studies conducted on more complex occupational groupings. If complex occupations are distinguished by innovative job demands, then these results support the contention that incumbent cognitive ability relates indirectly to job performance through prior learning and directly to job performance when novel job demands are encountered.

To be sure, alternative conceptualizations of causal relationships between cognitive ability and job performance exist. For example, [Crouse \(1979\)](#) argued that "it is not cognitive skill per se that affects later success. Rather, the stable motivations and aptitudes that lead to the development of cognitive skills also affect later success. A [cognitive ability] test's predictive power appears to derive in large part from its relationship to these stable underlying factors" (p. 85). This argument is also embodied in explanations of why biographical information demonstrates criterion-related validity (see [Owens's \[1968, 1971, 1976\]](#) developmental—integrative model or its descendant, [Mumford, Stokes, & Owens's \[1990\]](#) ecology model).

Regardless, as noted by [Schmidt and Hunter \(1980\)](#) , one of the direct results of meta-analysis procedures is to reject the tenet of situation specificity, providing original researchers the opportunity to pursue trait—performance relationships, biodata—performance relationships, or any other paths that hold promise for a

theory of performance prediction ([Campbell, 1990](#)). Before the widespread reporting of validity generalization (VG) results, authors hoping to systematically explore alternative theories of performance prediction had difficulty distinguishing between statistical artifacts and true variation in criterion-related validity across studies. Variance in criterion-related validity remaining after correcting for sampling error tends to be either small relative to sampling error ([Hunter & Hunter, 1984](#)) or small in an absolute sense ([McDaniel, Hirsh, Schmidt, Raju, & Hunter, 1986](#) ; [Schmitt, Gooding, Noe, & Kirsch, 1984](#) ; [Schmitt & Noe, 1986](#)).

Regardless, if nonzero variance in criterion-related validity remains after correction for statistical artifacts, subgroup or moderator effects may exist. [Wanous, Sullivan, and Malinak \(1989\)](#) noted that the inclusion of potential moderators in coding original research findings is a judgment call on the part of the VG investigator. If the VG investigators fail to code a moderator or the authors of the original research fail to provide this information, then the analysis may fail to detect true moderator effects. [Thomas \(1989, 1990\)](#) derived a likelihood-based procedure to estimate the number of independent ρ s underlying the observed distribution of criterion-related validities. [Lance, Stennett, and Searcy \(1992\)](#) applied this procedure to all studies of criterion-related validity reported in the *Journal of Applied Psychology* and *Personnel Psychology* between 1964 and 1982 (data were obtained from [Schmitt et al., 1984](#)). Results indicated that, at most, eight different population ρ (i.e., because of unknown statistical artifacts or study characteristics) best account for unexplained variation in criterion-related validity.

In this study, we explored what these previously unmeasured moderators might be. One might be judgment calls made by authors of original research. Specifically, [Wanous et al. \(1989\)](#) identified a number of judgment calls that might vary across VG researchers, causing results from VG analyses conducted on the same set of studies to yield different results. We argue that judgment calls made by the original investigators are of equal, if not greater, importance in deriving VG results. Beyond basic choices of which predictor and criterion measures to include, subtle and often unreported judgment calls are made at almost every step of the validation process. For example, the vast majority of published criterion-related validity studies on biodata do not report the method of empirical keying or why a particular method was chosen. [Devlin, Abrahams, and Edwards \(1992\)](#) found that this judgment call by the original investigator can have a meaningful effect on criterion-related validity. Many other judgment calls exist, ranging from deciding (a) whether, in the development of criterion measures, supervisors see enough of subordinate work activities to make behavioral ratings versus ratings of work outcomes to deciding (b) whether the general mandate of using multiple sources of job analysis information is adequately met with three versus four sources of information. Large portions of personnel selection textbooks address the advantages and disadvantages of such judgement calls (e.g., [Gatewood & Feild, 1990](#)), which might be thought of

as the craftsmanship that goes into a study, yet remains unreported in the final write-up.

Consequently, one might assume that (a) individual differences in investigator characteristics and (b) characteristics of the research project affect these judgment calls and subsequent research outcomes. Clearly, one of these classes of individual-differences variables is likely to be the cognitive abilities, skills, and knowledge possessed by the original investigators. [Schmidt, Hunter, Croll, and McKenzie \(1983\)](#) and [Hirsh, Schmidt, and Hunter \(1986\)](#) found that investigators with 3 or fewer years of experience (post-PhD) made subjective criterion-related validity estimates for job—test combinations containing about twice as much random error in comparison with more senior judges. Less expert judges also substantially underestimated observed validities for job—test combinations found in the literature. If investigators' experience is related to their estimates of criterion-related validity obtained from a selection procedure, it seems likely that their experience level is also related to their estimates of how likely a given job analysis procedure, research design, or performance appraisal system will contribute to the study of predictor—criterion relationships.

Other individual-difference characteristics of investigators have been examined outside of personnel selection. For example, [Sherwood and Nataupsky \(1968\)](#) found that investigators' biographical characteristics were highly related to results reported in studies of Black—White differences on intelligence tests. Examination of investigator characteristics is just starting to occur in the personnel selection literature. [Ryan and Sackett \(1989, 1992\)](#) reported substantial differences among those who conduct individual assessments for personnel selection purposes (i.e., psychologists and human resource professionals who are members of the American Psychological Association). For example, they reported that these individuals varied substantially in their (a) knowledge of legal constraints, (b) belief that validity of assessment is determined by the skill of the assessor, and (c) knowledge of guidelines for evaluating personnel selection validity. [Ryan and Sackett \(1992\)](#) noted that research on the validity of investigators' individual differences in personnel assessment practices is almost nonexistent. They speculated that the large observed differences in content and practice would likely be associated with large differences in criterion-related validity.

Finally, one might also expect that the reason or reasons why a research project was initially conducted would impact the criterion-related validity obtained. For example, an inquiry by the Equal Employment Opportunity Commission about test bias might cause investigators to make judgment calls that are very different in their research design and operationalizations in comparison with decisions made by investigators primarily interested in testing a model of performance prediction.

Our purpose in this study was to examine how two investigator individual-difference characteristics and the primary impetus behind the research project moderate criterion-related findings reported in the personnel selection literature. First, we obtained investigators' knowledge, skills, and abilities (KSAs) in the field with a surrogate measure used by [Hirsh et al. \(1986\)](#): number of years of experience in the field after receipt of PhD. Note that this is a very molar operationalization, quite different from what might be obtained from pencil-and-paper tests of authors' quantitative skills or mastery of the relevant research literature (which may erode or expand over time). We expected investigators with greater experience to know more about the research methods and substantive domains that bear on their specific research and, hence, to be more likely to conduct studies finding high criterion-related validity. This was even more likely to be true for authors submitting their studies to major research journals.

Second, we expected that the reward structure faced by investigators might affect the criterion-related validities obtained. Specifically, investigators employed in academic positions, private industry, consulting firms, or public sector organizations face very different reward contexts for conducting and publishing personnel selection research. One might speculate that "publish or perish" systems found in academia encourage faculty to submit almost all research efforts for publication. Hence, findings of low criterion-related validity in testing competing models of human performance or innovative selection systems might be submitted more frequently by academics. Private consultants, who depend on income generated from the design and implementation of successful selection systems, may be less motivated to publicly display mediocre results. However, we were unaware of any findings relating to investigators' reward structures and hence had no a priori expectations regarding how location of employment might moderate criterion validities reported in the literature.

Third, the original motivation or impetus to conduct a research project may impact the criterion-related validities obtained. Original investigators were surveyed about the original impetus behind the research project that led to the results they reported. Again, we were unaware of any findings relating research motivation to research results and hence could make no predictions about this relationship.

Finally, because most models conceive of human performance as a multiplicative function of ability and motivation ([Terborg, 1977](#)), we also examined relationships between years of experience and criterion-related validity within each motivational context. For example, relationships between our surrogate measure of investigator KSAs (i.e., experience) and criterion-related validity obtained may be much stronger when investigators work in environments with relatively large short-term rewards and punishments (e.g., the increase or decrease of income by consultants or promotions and merit pay increases received in private industry) versus environments with substantially less short-term impact (i.e., academia).

Method

Sample

All criterion-related validation studies of personnel selection systems published from 1964 to 1982 for the *Journal of Applied Psychology* and *Personnel Psychology* included in the meta-analysis reported by [Schmitt et al. \(1984\)](#) were updated through the spring of 1992 (39 studies added to Schmitt et al.'s original 99 yielded $N = 138$). Ninety-seven studies came from the *Journal of Applied Psychology*, and 41 came from *Personnel Psychology*. (References for the studies are available from Craig J. Russell.)

Procedure

All prior studies contained in the [Schmitt et al. \(1984\)](#) data set were reexamined to ensure correct recording of reliabilities, validities, study design, and so on. Only 99 coding differences were found in the original 15,706 pieces of information coded by Schmitt et al., resulting in an index of coding agreement greater than 99%. A total of 1,190 validity coefficients were coded (840 coming from [Schmitt et al., 1984](#)). As described by Schmitt et al., validities of measures within a single-predictor category and single-criterion category were averaged to produce a summary validity coefficient for each independent sample within a study, yielding a total of 443 summary validities. We did not make corrections for predictor reliability, criterion reliability, or range restriction, replicating Schmitt et al.'s procedures. All computations were performed on these summary validity coefficients.

In addition to findings reported in the studies, each author's place of employment was dummy coded (academia, private industry, consulting, public sector, and other), as was the number of years between publication of the study and each author's date of graduation with a PhD. Because both journals require authors to use the *Publication Manual of the American Psychological Association* ([American Psychological Association, 1983](#)) in formatting manuscripts for submission, we assumed that the authors' published affiliation reflected the authors' place of employment when they conducted the investigation.¹ Graduation dates were obtained from *Dissertation Abstracts International* and the *Directory of the American Psychological Association*. Of the information not available from these sources, approximately one third was obtained through personal telephone calls made by Craig J. Russell. The only a priori source of systematic bias in this experience information was due to the fact that, although we found information for many authors not based in the United States in *Dissertation Abstracts International*, we could not obtain graduation information for 17 authors affiliated with organizations in Israel.

Finally, a questionnaire was mailed to the first author of each publication. Authors were instructed that, although multiple causes may have led to a project, they

were to indicate the primary impetus behind the project that resulted in their study. Two sets of choices were available, labeled "organizational need" (i.e., need for employees who were higher or lower on a criterion, compliance with equal employment opportunity (EEO) concerns, compare criterion validities of different predictors in same sample) and "investigator desire to examine a research issue" (i.e., wanted to see how high of a criterion-related validity could be obtained or to compare criterion validities of different predictors in same sample vs. test some hypothesis from a theory or model). Space was also available for authors to describe other reasons as a primary impetus if none of the options presented seemed appropriate. We obtained responses from 78% of the authors (or coauthors when the first author could not be located or was deceased).

We weighted each summary validity coefficient by its sample size in computing average validity coefficients. Per the analyses conducted by [Schmitt et al. \(1984\)](#), we derived the variance of the summary validity coefficients σ^2_r , the variance resulting from sampling error σ^2_e , the variance remaining after subtracting variance resulting from sampling error σ^2_p , and the percentage of unexplained variance by using formulae found in [Hunter, Schmidt, and Jackson's \(1982\)](#) study. In addition, because number of years since receipt of PhD was a continuous moderator variable, we derived [Hunter et al.'s \(1982\)](#) formulae for the correlation between a continuous moderator and summary validities ($COR_{r,y}$ and $COR_{r,y}$; [Hunter et al., 1982, pp. 52—53](#)).

Results

Because mission statements and reviewer pools do not perfectly overlap for the two journals, relationships between source of publication and criterion-related validity were initially examined. Differences in mean criterion-related validity were not significant (mean $r_s = .283$ for *Journal of Applied Psychology* and $.259$ for *Personnel Psychology*). Hence, we assumed that source of publication did not moderate study results. We conducted all subsequent analyses on criterion-related validities pooled from both journals.

Author's Experience

We calculated the number of years since receipt of the PhD or experience, by subtracting the year of receipt of PhD from the year of each study's publication (authors who were doctoral students were coded as having 0 years of experience). This was done for the first author ($M = 8.1$ years, $SD = 8.3$ years) and the most experienced author (i.e., the one who had been out of school the longest, regardless of whether she or he was the first author [$M = 12.5$ years, $SD = 10.3$ years]). Average experience for all authors within the study was also calculated ($M = 9.5$ years, $SD = 9.3$ years). Summary validities were correlated with number of years of the most experienced author at $.079$, with number of years of experience by first author at $-.017$, and with average number of years of

experience of all authors at $-.042$. Correcting these correlations for sampling error yielded $.080$, $-.017$, and -0.43 , respectively. None of these correlations were significant.

Author's Place of Employment

[Table 1](#) contains comparisons of average validities broken down by place of employment for the first author. Average validities reported by first authors employed in private industry were significantly greater than those reported by first authors employed in academia ($.32$ vs. $.24$), $t(299) = 5.4$, $p < .001$ (using [Winer's \[1971\]](#) t test for small samples with unequal variances, pp. 41—42). Comparisons with other groupings cause problems because of small numbers of studies published by first authors who were employed by consulting firms and in the public sector. However, studies published by first authors employed in nonacademic settings yielded a significantly higher mean criterion-related validity when compared with studies by first authors employed in academia ($.29$ vs. $.24$), $t(329) = 5.43$, $p < .01$. Comparisons among the average validities for places of employment that did not involve private industry in the comparison were nonsignificant. Furthermore, site of the first author's employment did not covary with type of design, occupation, predictor, or criterion used in the study. Hence, differences in ρ reported by [Schmitt et al. \(1984\)](#) resulting from study design or type of occupation, predictor, or criterion were not responsible for the differences shown in [Table 1](#).

Primary Impetus Behind Research

[Table 2](#) shows comparisons of average validities broken down according to the primary impetus behind the research project: (a) organizational need versus researcher interest or (b) one of five specific reasons given for the research (i.e., maximizing criterion-related validity, complying with EEO guidelines, obtaining an incremental increase in criterion-related validity over the existing selection system, or testing or developing some theory). Projects conducted to address some organizational need yielded significantly higher mean criterion-related validities than projects conducted to address some investigator interest ($.32$ vs. $.24$), $t(335) = 5.62$, $p < .001$. Note that these mean validities are almost identical to those reported in [Table 1](#) for studies conducted, respectively, by researchers employed in either industrial or academic settings. There was not perfect overlap between these two moderators; 64% of first authors employed in industry reported organizational needs as the primary impetus, whereas 36% reported researcher interest as the primary impetus. Forty-six percent of first authors employed in academic settings reported organizational need as the primary impetus, whereas 54% reported researcher interest as the primary impetus.

Projects conducted to address EEO concerns yielded mean criterion-related validities comparable with studies conducted to obtain an increase in existing selection system validity ($.331$ vs. $.346$), $t(124) = 0.500$, $p > .05$. Projects

conducted to address EEO concerns and augment existing selection system validities both yielded mean validities greater those found for studies conducted primarily to maximize validity or to test theories: .331 versus .281 and .218, $t(167) = 2.099$, $p < .05$, and $t(104) = 5.526$, $p < .001$, respectively; and .346 versus .281 and .218, $t(201) = 2.363$, $p < .01$, and $t(144) = 5.24$, $p < .001$, respectively. Furthermore, studies conducted to maximize criterion-related validity yielded significantly higher validities (mean $r_{xy} = .281$) than studies conducted to test or develop some theory (mean $r_{xy} = .218$), $t(187) = 3.568$, $p < .001$.

No significant differences were found in [Table 2](#) for criterion related validities in studies conducted to maximize validity, regardless of whether the investigator or organization initiated the effort (.313 vs. .268), $t(111) = 1.241$, $p > .05$, or for studies conducted to determine any increase in validity over existing selection systems (.353 vs. .318), $t(78) = 0.8383$, $p > .05$. The only mean validity significantly and meaningfully below those reported for all other research purposes, regardless of how the project was initiated, was .218 for studies initiated by investigators to test some theory.

Author's Experience × Place of Employment

Relationships between authors' experience and criterion-related validity changed considerably when examined in view of first author's place of employment. A negative relationship between first author's experience and criterion-related validity was found for the 159 validities reported by first authors employed in academia $COR_{r,y} = -.167$ $p < .05$; $COR_{r,y} = -.169$. However, a positive relationship between first author's experience and validity was found for the 132 validities reported by first authors employed in private industry $COR_{r,y} = .224$, $p < .01$; $COR_{r,y} = .230$ and for the 35 and 29 validities reported by first authors employed in consulting and in the public sector, respectively $COR_{r,y} = .245$, ns ; $COR_{r,y} = .277$ and $COR_{r,y} = .097$, ns ; $COR_{r,y} = .097$. Identical trends in size of correlations and significance levels occurred when average authors' and most senior authors' experience level was related to the criterion-related validity reported in each study (these correlations and significance levels are available from Craig J. Russell on request). [Table 3](#) shows average validities broken down by place of employment and according to whether the first author had less than 6 years or 6 or more years of post-PhD experience (6 years was chosen because of its common use in promotion and tenure decisions in academia).

However, first author's experience and the interaction of experience and place of employment may have covaried with other study characteristics examined by [Schmitt et al. \(1984\)](#). An analysis of variance design was used to identify and remove the main effects of type of criterion, predictor, design, and occupational group, yielding residual variance in criterion-related validities that was unrelated to these study characteristics. Residual variance remaining in criterion-related validities was then correlated with first author's experience level within place of

employment. Results indicated that for academic authors the relationship between criterion-related validity and author experience changed to $-.010$, *ns* $COR_{r,y} = -.010$. However, the relationship remained large and positive between criterion-related validities and first author's experience when the first author was employed in private industry $COR_{r,y} = .206$, $p < .05$; $COR_{r,y} = .211$. Consequently, results suggested a strong positive relationship between experience of first author and validity when the first author was employed in private industry. No relationship existed for authors employed in academic settings.

Author's Experience x Primary Impetus Behind Research

First author's level of experience was unrelated to criterion-related validity observed in any of the studies conducted for reasons other than testing a theory. First author's experience correlated $.212$ ($p < .05$) with criterion validities reported in studies initiated for purposes of testing theory $COR_{r,y} = .219$. Furthermore, when other study characteristics were removed from criterion-related validities, first author's experience still correlated with residual variance $.222$ ($p < .05$). [Table 4](#) contains a breakdown of the mean correlations for first authors with fewer than 6 years of experience or 6 or more years experience in studies initiated for purposes of theory testing.

Discussion

Results suggested that there is no main effect of author experience level on criterion-related validity. A fairly strong main effect for place of employment was found. Authors employed in private industry reported significantly higher validities than those employed in academia. Authors concerned with EEO compliance and augmenting existing selection systems yielded significantly higher validities in comparison with those who simply wished to obtain a high validity. Authors concerned with testing a theory yielded mean correlations that tended to be significantly lower (by more than $.10$) than those of authors concerned with EEO issues or with augmenting existing selection systems. The magnitude of this difference is comparable to those used by [Hunter and Hunter \(1984\)](#) to infer that some types of predictors are more criterion valid than others.

Perhaps the only counterintuitive finding among the main effects was that authors wishing to augment existing selection systems yielded higher criterion-related validities than did authors simply wishing to maximize validity in an initial selection application. One possible explanation is that when authors enjoy the advantage of an existing selection system, job analyses, performance measures, and other hurdles required of field research will have already been negotiated. One might expect more reliable and less contaminated or deficient criterion measures under these conditions (cf. [Kemery, Mossholder, & Roth, 1987](#)).

Significant interaction effects were found on criterion-related validities between (a) first author's experience level and first author's place of employment and (b) first author's experience level and a theory-testing impetus for the research project. Authors employed in private industry with more than 6 years experience post-PhD reported validities averaging .08 above those reported by private sector authors with fewer than 6 years experience. For authors employed in academia, no effect was found after we controlled for other study characteristics. However, first authors with more experience were likely to generate validities approximately .13 higher than authors with less experienced when the primary purpose of the research was to test some theory.

Nevertheless, we originally speculated that the relationship between author experience and criterion-related validity would be positive for authors regardless of where they were employed or the purpose for which the research was originally initiated, although we thought emphasis on short-term rewards might make the relationship stronger in private industry than in academia. This initial speculation regarding the interaction effect is consistent with the results. Senior authors in the private sector may be more capable (because of their experience) and motivated (because of the reward context) to design systems that yield high criterion-related validities. More senior authors whose purpose is to test theories may simply have a greater mastery of the relevant literature, enabling them to make the right kind of judgment calls needed to appropriately test a theoretical proposition.

Alternative explanations abound for the finding that authors in academia obtain comparable criterion-related validities regardless of their experience levels. These generally fall under the headings of authors' motivation, ability, and opportunity to do research. For example, less experienced academics may be motivated by the "tenure clock" to get publications in refereed journals and, hence, are extremely focused on designing research efforts that yield respectable validities. Alternatively, receipt of promotion and tenure approximately 6 years into their careers effectively eliminates all major economic rewards available in academia, so more experienced academic researchers may have fewer rewards dependent on research outcomes and, hence, may be less motivated than their equally senior peers in private industry. Another possibility is that, after receiving tenure, academics may feel free to pursue more esoteric theoretical research that helps develop models of performance prediction while not yielding exceeding large validities. Finally, as academics acquire more experience, they may develop packaging skills that enable them to publish studies yielding low validities with a veneer of theory development that is, at best, post hoc. An alternative (and equally unpleasant) explanation for the positive relationship between experience and validities reported is that senior authors may co-opt the efforts of their less experienced (and more junior) colleagues ([Hunt & Blair, 1987](#)).

One limitation of the current results is the implicit assumption that authors of all experience levels in all employment settings are equally motivated to compose and submit their findings for refereed journals. The moderator effects reported above may be due to differential decisions to simply take the time to write up the study results. This potential bias in VG results might be expected to hold regardless of whether the write-up was for an archival technical report (such as the studies used in [Hunter & Hunter, 1984](#)) or for submission to a journal.

Regardless, no evidence in the current study explicitly supports any single explanation. Any or all of them could be correct in any single study included in the meta-analysis. Although the explanations may be interesting from a perspective of post-PhD career planning, we pose them more to amplify an observation made by [Sackett, Tenopyr, Schmitt, and Kehoe \(1985\)](#). Meta-analysis results are only as good as the individual studies that go into them. The current results, based on findings published in two of the leading refereed journals, suggest that the skill levels and reward contexts facing authors of original research also affect meta-analytic results.

Does this suggest that, if two hypothetical investigators were asked to examine the same predictor—criterion relationship, they would conduct their research so differently that dissimilar criterion validities will result? Possibly. The limits of field research prevent random assignment of investigators to employment venues or other research conditions; hence, causal implications cannot be drawn. It seems more likely, however, that two real-world investigators will self-select employment venues and have some voice in the predictor—criterion combination chosen for study. The point is that real-world investigators from industry, academia, consulting, and the public sector will probably not decide to conduct identical job analyses, to operationalize criteria the same way, or to use comparable test administration procedures (i.e., make the same judgment calls).

With this in mind, we reread all 138 original articles in hopes of detecting trends that might have gone undetected when coarse moderators like employment venue are used. For example, within each type of predictor we examined whether there were any noticeable differences in operationalization procedures, organizational circumstances, and so on, for authors in various employment settings. We were especially attentive to any sequences of replications and extensions of research conducted by authors in different employment settings and for different reasons. No discernible trends could be detected. In fact, one problematic observation from this effort was that few streams of research exist to which both academics and nonacademics contribute. For example, almost all assessment center criterion-related validity studies were authored by individuals in consulting or industry, primarily in the 1970s. Although many academic authors examined assessment center construct validity in the 1980s, almost none of these studies report criterion-related validities. Perhaps most disturbing was that since 1983 only three articles had been published with first authors employed in industry (2% of the total number of authors employed in industry), only three had

been published with first authors employed in the public sector (8.5% of total), and only one had been published with the first author employed in a consulting firm (3% of total). Fifty-one studies (35% of total) were published in this same time span with first authors employed in academia.

Regardless, the most important implication of these findings is not for career guidance in industrial—organizational psychology. The results suggest that low relative or absolute levels of σ^2_r ([McDaniel et al., 1986](#) ; [Schmitt & Noe, 1986](#)) may still hide meaningful moderator effects. However, as noted by [McDaniel et al. \(1986\)](#) , the meta-analyses we examined had been conducted on original research containing some very large sample sizes. Sampling error as a percentage of total variance in criterion-related validity is small relative to meta-analyses conducted on studies characterized by smaller samples. To the extent the sampling error widens the confidence bands around estimates of mean correlation, attempts to replicate the current moderator results in groups of studies characterized by smaller samples (e.g., [Hunter & Hunter, 1984](#)) will have less power.

The most important implication is for building theories of performance prediction. It appears that original research in personnel selection contains blind spots caused by combinations of investigators' KSAs and reward contexts. The blind spots identified in the current results are different from second-order sampling error ([Schmidt, Pearlman, Hunter, & Hirsh, 1985](#)). We call them *second-order production errors* to reflect the fact that it was not the sampling procedure that led to an unrepresentative cross-section of studies generating these VG results (indeed, the entire population of validity studies published in two journals from 1964 to 1992 was included). Instead, factors reflecting the authors' KSAs and reward contexts appear to have influenced the results that they produced. This suggests that even if the universe of all criterion-related validity studies ever conducted were included, VG results can still be influenced by the capabilities and motivational agendas of the original investigators and must be used with caution in guiding theory development.

References

- American Psychological Association. (1983). *Publication manual of the American Psychological Association* (3rd ed.).(Washington, DC: Author)
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology.(In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 687—732). Palo Alto, CA: Consulting Psychologists Press.)
- Crouse, J. (1979). The effects of academic ability.(In C. Jencks, S. Bartlett, M. Corcoran, J. Crouse, D. Eaglesfield, G. Jackson, K. McClelland, P. Mueser, M. Olneck, J. Schwartx, J. Ward, & J. Williams (Eds.), *Who gets ahead? The determinants of economic success in America* (pp. 85—121). New York: Basic

Books.)

- Devlin, S. E., Abrahams, N. M. & Edwards, J. E. (1992). Empirical keying of biographical data: Cross-validity as a function of scaling procedure and sample size. (*Military Psychology*, 4, 119—136.)
- Gatewood, R. D. & Feild, H. S. (1990). *Human resource selection*. (Hinsdale, IL: Dryden Press)
- Hirsh, H. R., Schmidt, F. L. & Hunter, J. E. (1986). Estimation of employment validities by less experienced judges. (*Personnel Psychology*, 39, 337—344.)
- Hunt, J. G. & Blair, J. D. (1987). Content, process, and the Matthew effect among management academics. (*Journal of Management*, 13, 191—210.)
- Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. (*Journal of Vocational Behavior*, 29, 340—362.)
- Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. (*Psychological Bulletin*, 96, 72—98.)
- Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. (Newbury Park, CA: Sage)
- Hunter, J. E., Schmidt, F. L. & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. (Beverly Hills, CA: Sage)
- Kemery, E. R., Mossholder, K. W. & Roth, L. (1987). The power of the Schmidt and Hunter additive model of validity generalization. (*Journal of Applied Psychology*, 72, 30—37.)
- Lance, C. E., Stennett, R. B. & Searcy, C. A. (1992, November). *On evaluating the situational specificity hypothesis in meta-analysis*. (Paper presented at the annual meeting of the Southern Management Association, New Orleans, LA)
- McDaniel, M. A., Hirsh, H. R., Schmidt, F. L., Raju, N. S. & Hunter, J. E. (1986). Interpreting the results of meta-analytic research: A comment on Schmitt, Gooding, Noe, and Kirsch (1984). (*Personnel Psychology*, 39, 141—148.)
- Mumford, M., Stokes, G. S. & Owens, W. A. (1990). *Patterns of life history: The ecology of human individuality*. (Hillsdale, NJ: Erlbaum)
- Owens, W. A. (1968). Toward one discipline of scientific psychology. (*American Psychologist*, 23, 782—785.)
- Owens, W. A. (1971). A quasi-actuarial prospect for individual assessment. (*American Psychologist*, 26, 992—999.)
- Owens, W. A. (1976). Background data. (In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 609—644). Chicago: Rand McNally.)
- Ryan, A. M. & Sackett, P. R. (1989). Exploratory study of individual assessment practices: Interrater reliability and judgments of assessor effectiveness. (*Journal of Applied Psychology*, 73, 568—579.)
- Ryan, A. M. & Sackett, P. R. (1992). Relations between graduate training, professional affiliation, and individual psychological assessment practices for personnel decisions. (*Personnel Psychology*, 45, 363—387.)
- Sackett, P. R., Callahan, C., DeMeuse, K., Ford, J. K. & Kozlowski, S. (1986). Changes over time in research involvement by academic and nonacademic psychologists. (*Industrial/Organizational Psychologist*, 24(1), 40—43.)
- Sackett, P. R., Tenopyr, M. L., Schmitt, N. & Kehoe, J. (1985). Commentary on

forty questions about validity generalizations and meta-analysis.(*Personnel Psychology*, 38, 697—798.)

Schmidt, F. L. & Hunter, J. E. (1980). The future of criterion-related validity.(*Personnel Psychology*, 33, 41—60.)

Schmidt, F. L., Hunter, J. E., Croll, P. R. & McKenzie, R. C. (1983). Estimation of employment test validities by expert judgment.(*Journal of Applied Psychology*, 68, 590—601.)

Schmidt, F. L., Pearlman, K., Hunter, J. E. & Hirsh, H. R. (1985). Forty questions about validity generalization and meta-analysis.(*Personnel Psychology*, 38, 697—798.)

Schmitt, N., Gooding, R. Z., Noe, R. D. & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics.(*Personnel Psychology*, 37, 407—422.)

Schmitt, N. & Noe, R. A. (1986). On shifting standards for conclusions regarding validity generalization.(*Personnel Psychology*, 39, 849—851.)

Sherwood, J. J. & Nataupsky, M. (1968). Predicting the conclusions of negro—white intelligence research from biographical characteristics of the investigator.(*Journal of Personality and Social Psychology*, 8, 53—58.)

Terborg, J. (1977). Validation and extension of an individual differences model of work performance.(*Organizational Behavior and Human Performance*, 18, 188—216.)

Thomas, H. (1989). A mixture model for distributions of correlation coefficients.(*Psychometrika*, 54, 523—530.)

Thomas, H. (1990). A likelihood-based model for validity generalization.(*Journal of Applied Psychology*, 75, 13—20.)

Wanous, J. P., Sullivan, S. E. & Malinak, J. (1989). The role of judgment calls in meta-analysis.(*Journal of Applied Psychology*, 74, 259—264.)

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.).(New York: McGraw-Hill)

1

[Sackett, Callahan, DeMeuse, Ford, and Kozlowski \(1986\)](#) examined a perceived decline in contributions by authors with nonacademic affiliations in *Journal of Applied Psychology* and in *Personnel Psychology* in 1985. They found that "virtually all the 1985 authorships by researchers listing industry or consulting affiliations" (p. 41) were either for studies actually conducted in an academic setting (i.e., a graduate student or faculty member conducted the research, left academia, and published the article under their nonacademic affiliation) or were studies by junior members of a research team employed at the field site of the sponsoring organization. However, it is not clear whether similar errors in authorship occur for investigators moving from nonacademic to academic employment settings. Regardless, if Sackett et al.'s results were to be interpreted at face value, our place of employment moderator variable may best be

considered to indicate career choice or inclination rather than actual employment venue.

Table 1.

Table 1
 Table 1: Distribution of Career Choice (Based on Employment Status)

Career Choice	No. of Respondents	Employment Status					Total
		Full-time	Part-time	Contract	Seasonal	Unemployed	
Government	15	10	5	0	0	0	15
Private	25	15	10	0	0	0	25
Self-employed	10	5	5	0	0	0	10
Total	50	30	20	0	0	0	50

Note: Data is preliminary and subject to change.

Table 2.

Table 2
 Table 2: Distribution of Career Choice (Based on Education Level)

Career Choice	No. of Respondents	Education Level					Total
		High School	Some College	Bachelor's	Master's	PhD	
Government	15	5	5	5	0	0	15
Private	25	10	10	5	0	0	25
Self-employed	10	5	5	0	0	0	10
Total	50	20	20	10	0	0	50

Note: Data is preliminary and subject to change.

Table 3.

Table 3
 Table 3: Distribution of Career Choice (Based on Age Group)

Career Choice	No. of Respondents	Age Group					Total
		18-24	25-34	35-44	45-54	55+	
Government	15	5	5	5	0	0	15
Private	25	10	10	5	0	0	25
Self-employed	10	5	5	0	0	0	10
Total	50	20	20	10	0	0	50

Note: Data is preliminary and subject to change.

Table 4.

Table 4
 Table 4: Distribution of Career Choice (Based on Gender)

Career Choice	No. of Respondents	Gender		Total
		Male	Female	
Government	15	10	5	15
Private	25	15	10	25
Self-employed	10	5	5	10
Total	50	30	20	50

Note: Data is preliminary and subject to change.