Journal of Applied Psychology
2003, Vol. 88, No. 6, 1068–1081

Copyright 2003 by the American Psychological Association, Inc.
0021-9010/03/$12.00 DOI: 10.1037/0021-9010.88.6.1068

A Meta-Analytic Study of General Mental Ability Validity for Different Occupations in the European Community

Jesús F. Salgado
University of Santiago de Compostela

Neil Anderson
University of Amsterdam

Silvia Moscoso
University of Santiago de Compostela

Cristina Bertua
Goldsmith College, University of London

Filip de Fruyt
Ghent University

Jean Pierre Rolland
University of Paris X

A comprehensive meta-analysis of the validity of general mental ability (GMA) measures across 12 occupational categories in the European Community (EC) is presented. GMA measures showed that there is validity generalization and large operational validities for job performance and training success in 11 occupational groups. Results also showed that job complexity moderated the magnitude of the operational validity of GMA tests across three levels of job complexity: low, medium, and high. In general, results were similar to those found in the United States, although the European findings showed a slightly larger magnitude of operational validity in some cases. Theoretical and practical implications of these findings for personnel selection are discussed.

In the past 80 years, hundreds of criterion-related validity studies have been carried out across the world, relating general mental ability (GMA) measures to work-related criteria. Following Schmidt (2002), GMA is defined effectively in this research as any measure that combines two, three, or more specific aptitudes, or any measure that includes a variety of items measuring specific abilities (e.g., verbal, numerical, spatial). Consequently, the observed validity of GMA for a specific study may be achieved using an omnibus GMA test (e.g., The Wonderlic Personnel Test, Otis Employment Test, Raven's Progressive Matrices Test) or using different specific tests combined in a battery (e.g. General Aptitude Test Battery [GATB], Differential Aptitude Tests [DAT], Primary Mental Aptitude [PMA]). In these primary studies, a wide variety of types of jobs were explored. Much of this research has been summarized in quantitative reviews of the validity of personnel selection measures. For example, Ghiselli (1966, 1973) wrote several now classic reviews. Ghiselli's main conclusion was that, in general, for training success and job proficiency, intelligence tests and spatial, mechanical, and perceptual ability measures are the best predictors. He also suggested that there was an intelligence or specific ability test with an at least moderate validity for each of the eight groups of occupations analyzed. This group of occupations included managerial, clerical, services, sales, and protective occupations, as well as vehicle operators, trades and crafts, and industrial occupations.

Meta-Analyses of Cognitive Ability Measures

After Ghiselli's (1966, 1973) reviews, several meta-analyses were conducted to examine the validity of cognitive ability measures as predictors in personnel selection. These meta-analyses are characterized by the fact that they analyzed the validity of general cognitive ability measures for broad occupational families (e.g., Hartigan & Wigdor, 1989; Hunter & Hunter, 1984) or by the fact that they meta-analyzed validity for only one specific occupation, such as clerical occupations, law enforcement occupations, computer programmers, sales jobs, and trades in the utility industry (Hirsch, Northrop, & Schmidt, 1986; Levine, Spector, Menon, Narayanan, & Cannon-Bowers, 1996; Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Vinchur, Schippman, Switzer, & Roth, 1998). For example, Hunter and Hunter (1984) classified the occupations described in the Dictionary of Occupational Titles (DOT) using a classification of five families. They found that criterion-related validity of cogni-

Jesús F. Salgado and Silvia Moscoso, Department of Social Psychology, University of Santiago de Compostela; Neil Anderson, Department of Psychology, University of Amsterdam; Cristina Bertua, Department of Psychology, Goldsmiths College, University of London; Filip de Fruyt, Department of Psychology, Ghent University; Jean Pierre Rolland, Department of Psychology, University of Paris X.

The research reported in the article was partially supported by Grant BSO2001-3070 from Ministerio de Ciencia y Tecnología (Spain) to Jesús F. Salgado.

We offer this article as a posthumous homage to John E. Hunter, whose seminal work served as a foundation for the research reported here. John E. Hunter and Frank Schmidt's thought inspired much of our work in this area. We also thank Heinz Schuler for his help in the literature search and Jerry Kehoe and two anonymous reviewers for their comments that improved a prior version of this article.

Correspondence concerning this article should be addressed to Jesús F. Salgado, Departamento de Psicología Social, Universidad de Santiago de Compostela, Santiago de Compostela, 15782, Spain. E-mail: psjesal@usc.es

tive tests was moderated by the occupational family membership. This membership was dependent on the job complexity of the specific occupation. Hunter and Hunter found that, for job performance and training success criteria, the validity of cognitive tests for more complex occupational families was larger than that of less complex occupations. The Panel of the National Research Council (Hartigan & Wigdor, 1989) reanalyzed Hunter and Hunter's data set, analyzed a new set of data, and found essentially the same findings. Therefore, these two large-scale meta-analyses came to the same conclusion: Job complexity is a moderator of the criterion-related validity of cognitive ability measures. However, these studies did not analyze specific or small occupational groups (e.g., managerial, clerical, sales) as Ghiselli did. Furthermore, there is a relevant difference between the study of Hunter and Hunter and the study of the Panel of the National Research Council. Hunter and Hunter corrected the validity for criterion unreliability and range restriction in predictor scores, whereas the Panel of the National Research Council only corrected the validity for criterion unreliability. In addition, there is a difference in the correction for criterion unreliability. Although Hunter and Hunter used a reliability estimate of .60 for job performance and .81 for training success, the Panel of the National Research Council used an estimate of .80 for both criteria. Consequently, there are important differences in the final results of these two meta-analytic studies. For example, in the case of the job performance criterion, Hunter and Hunter found that the validity ranged from .23 for Family II (feeding–off-bearing) to .58 for Family III (synthesize–coordinate), whereas the Panel of the National Research Council found that the validity ranged from .17 for Family I (setup) to .28 for Family III. For the training success criterion, Hunter and Hunter found that the validity ranged from .50 for Family III to .65 for Family I (setup) and the Panel of the National Research Council found that the validity ranged from .00 for Family V (copy–compare) to .60 for Family I. Differences may also be due to different sample sizes within each job family in both studies. Apparently, only Job Families I, IV (analyze–compile–compute), and V and Job Families IV and V were similarly represented in both studies for job performance and training success, respectively (see Hartigan & Wigdor, 1989, p. 161). A potential limitation of these two investigations is that they only considered studies carried out with the GATB.

Several other meta-analyses were conducted for specific occupational categories. For example, Pearlman et al. (1980) carried out a meta-analysis of the validity of cognitive ability measures for predicting job performance and training success in clerical occupations. They found that GMA and other specific cognitive measures, including memory and verbal, numerical, spatial, mechanical, and perceptual abilities, showed moderate to large criterion-related validity and that they generalized validity across samples and jobs. Hirsch et al. (1986) examined the validity generalization of cognitive ability measures for law enforcement occupations. They found validity coefficients of a modest magnitude for job proficiency criteria. However, except for spatial–mechanical ability measures, there was either no validity generalization, or its magnitude was very small. In the case of training success, they found a large validity generalization effect. Levine et al. (1996) carried out a meta-analysis using the criterion-related validity studies relating to crafts jobs in the utility industry (e.g., electrical assembly, mechanical utility jobs, telephone technical jobs). They found that the cognitive tests that included measures of verbal ability, numerical ability, abstract reasoning, and general intelligence showed an operational validity of .43 for predicting job performance. Schmidt et al. (1980) found that, in the case of computer programmer occupations, GMA measures had an operational validity of .73 for predicting job proficiency and of .91 for predicting training success. Finally, Vinchur et al. (1998) carried out a meta-analysis of the validity of cognitive measures for predicting job performance in sales jobs. They found an operational validity of .31 for overall cognitive ability and of .40 for GMA.

Considering the results of the meta-analyses mentioned above, together with Ghiselli's (1966, 1973) findings, it appears that there are some differences in the magnitude of the criterion-related validity of GMA for different jobs and occupational groups. This does not mean that the validity is specific for these jobs because 90% credibility values were positive and different from 0 in most cases. However, beyond the minimum of validity generalization for all jobs, validity may be larger or smaller according to the type of occupation. With regard to this, although there are a number of researches who have addressed this issue, some occupational groups have not been well represented in these meta-analyses. For example, no meta-analyses have been published in scientific journals reporting validity generalization analyses for managerial occupations, professionals (e.g., engineers, physicians, chemists), drivers, or even for apprentices. A second limitation of the prior meta-analyses is that they have only examined U.S. validity studies. Therefore, the conclusions of those meta-analyses could be valid only for the U.S. economy (e.g., Herriot & Anderson, 1997; Salgado, Viswesvaran, & Ones, 2001).

Consequently, the first objective of this research was to replicate and extend past validity generalization research conducted solely in the United States by examining the operational validity of GMA measures for predicting job performance and training success in the European Community (EC) and in a variety of occupations, including a broader range of occupations. Our second objective was to examine whether job complexity moderates the validity of GMA tests so that GMA shows larger validity for more complex occupational groups. We chose the EC as the region for examining this goal. We made this decision for three reasons: (a) There has been a notable lack of European meta-analyses, (b) our results would allow the comparison of the GMA validity in the EC with previous meta-analytic results from the United States, and (c) the EC comprises a number of culturally and socially diverse countries whose combined population is comparable with that of the United States. The 15 current members of the EC are not a completely homogenous group of countries. There is a certain degree of heterogeneity in the cultural values and in the organizational selection practices, but as Viswesvaran and Ones (2002) recently pointed out, EC countries are relatively homogeneous in comparison with the United States because they have less within-country diversity (i.e., ethnic and racial diversity), and the personnel psychologists have a very similar view of professional personnel selection processes (Levy-Leboyer, 1994; Roe, 1989). Recently, Salgado and Anderson (2002), in an examination of 16 European countries, showed that GMA tests are used more by organizations in EC countries than in the United States and that the majority of European companies are medium and small companies (fewer than 500 employees). Moreover, these authors also demonstrated that there are only small differences among the majority of the European countries in the popularity of GMA tests. They illustrated that the perceptions of applicants in the EC, with regard to the GMA tests, are very similar. In addition, there are initiatives to harmonize the legislative structures and the testing standards in Europe. A recent report of the European Foundation for the Improvement of Living and Working Conditions (EFILWC; Paoli & Merllié, 2001) showed that there is a great similarity among the members of the EC in attributes such as (a) proportion of employees in the workforce, (b) the proportion of women in the workforce, (c) percentage of employees having influence over their working hours, (d) the proportion of workers having to solve unforeseen problems, (e) the proportion of workers having to perform monotonous tasks, (f) the percentage of workers carrying out complex tasks, (g) the proportion of employees learning new things in their work, (h) the proportion of employees who think that their skills match the demands of their job, (i) the percentage of employees benefited from training provided by their company and the average duration of the training per person. According to EFILWC, other attributes in which the European countries are very similar are the duration of working hours (the average is 39.8 hr for full-time workers), the average commuting time (37.5 min), the average of paid vacation time (25.9 days per year). In contrast, according to the U.S. Department of Labor, Bureau of Labor Statistics, the paid vacation time ranges from 9.6 days for workers with 1 year of tenure to 20.3 days for the employees with 20 years of tenure, and the average working week is 34.2 hr. Consequently, it may be assumed that there is more homogeneity among the European countries than between the European countries and the United States.

Method

Search for Studies

On the basis of the goals of this research, a database was developed containing European validity studies. These studies had to meet three criteria to be included in the database: (a) report validity coefficients relating job performance and training success measures and general cognitive ability measures; (b) the samples should be applicants, employees or trainees, but not students; (c) only civilian jobs (sum of all private industry and government workers) would be considered, excluding military occupations and agricultural workers.

The search was made using six strategies. First, a computer search was conducted in the PsycLit database. Second, a manual article-by-article search was carried out in a large number of European journals (a complete list may be obtained from Jesús F. Salgado). Third, test manuals were inspected looking for criterion-related validity studies. Fourth, several test publishing companies were contacted and asked for reports in which validity studies were reported. Fifth, the reference section of the articles obtained was checked to identify further articles. Finally, several well-known European researchers were contacted in order to obtain additional articles and supplementary information related to published articles (e.g., complete matrix of correlation coefficients).

The literature search resulted in 89 articles, including 166 independent samples. These articles consisted of 71 (137 samples) published and 18 (29 samples) unpublished studies. Combining the published and unpublished studies conducted across the EC, the final database contained 81 independent samples with training success as the criterion and 85 samples with job performance as the criterion. The following countries contributed studies to the database: Belgium (2), France (18), Germany (9), Ireland (1), the Netherlands (10), Spain (13), and the United Kingdom (36).

Procedure

Two researchers served as judges, working independently to code every study. For each study the following information was recorded if available: (a) sample characteristics, such as gender, age, education, and so forth, (b) occupation and information related, (c) GMA measures used, (d) criterion used, (e) reliability of GMA, (f) criterion reliability, (g) range restriction value or data for calculating this value, (h) statistics concerning the relation between GMA and criterion, (i) correlation among the GMA measures when more than one was used, and (j) country in which the study was carried out. Disagreements were solved through discussion until the researchers agreed on the ability. We used the studies conducted in the United Kingdom (36% of the total) to calculate the reliability of the codification process. This was done for three reasons: (a) the U.K. studies usually contained more information for recording than the rest of the studies, (b) they were a very representative number of the total in terms that all the occupational groups were represented with at least one validity study, and (c) because all the U.K. primary studies used English as the language in which the study was written, the authors were more able to resolve any resultant coding differences and to attribute these to true coding interpretation differences rather than to language differences. Using the samples size, criterion reliability, predictor reliability, and validity coefficients, the overall agreement between researchers (prior to the consensus), estimated as a correlation, was .93. One validity coefficient was used per sample for each criterion–ability–occupation condition. When a study contained conceptual replications (i.e., two or more measures of the same construct were used in the same sample), linear composites with unit weights for the components were formed. Linear composites provide more construct valid estimates than the use of the average correlation. Nunnally (1978, pp. 166–168) and Hunter and Schmidt (1990, pp. 457–463) provided formulas for the correlation of variables with composites. Examples of the GMA measures used in the studies are (a) Batteries: DAT, GATB, T2, ASDIC, Intelligence Structure Test (IST-70), Wilde Intelligence Test (WIT), GVK, PMA, and Aptitudes Mentales Primarias (AMPE), (b) *g* tests: Raven's Progressive Matrices, Cattell's Fair Culture Tests, Otis Employment Test, Alpha Test, Logique Intelligence, CERP, Domino, D-48, NIIP-33.

Following a first inspection of the jobs represented in our database, we decided to create relatively homogeneous and conceptually meaningful occupational groups in accordance with the previous procedure used by Schmidt, Hunter, and Pearlman (1981) for the clerical jobs. To classify the jobs in occupational categories, we used DOT (U.S. Department of Labor, 1991). On the basis of the information included in the primary studies about the jobs, we grouped the jobs in the following 10 DOT occupational categories: 00 (occupations in engineering), 02 (chemist), 18 (managers), 20 (typing and filing occupations), 23 (information and message distribution occupations), 25 (sales occupations), 37 (police), 60 (mechanics), 82 (electrician), and 90 (drivers) plus two miscellaneous categories, skilled worker and apprentices. Therefore, the total number of occupational categories is 12. These DOT categories differ substantially in terms of job requirements, and the number of studies in our database allowed us to consider each one as a separate entity. The examples of jobs included in each occupational category appear in Table 1.

To achieve our second objective, all the jobs included in the studies were classified in terms of job complexity level. After an inspection of the jobs contained in the occupational groups, we decided to create only three levels of job complexity. Occupations classified as being high complexity consisted of the jobs with scores 0 and 1 in the data dimension of the DOT. The medium level of job complexity consisted of jobs with scores of 2, 3, and 4 in the same DOT dimension. The low level of job complexity consisted of jobs with scores of 5 and 6 in the data dimension. With this classification system, all the jobs included in our categories 00, 02, and 18 (i.e., engineers, chemists, and managers) were classified as jobs of a high level of complexity. All the jobs included in our categories 20, 25, 37, 60, 82, (i.e., typing and filing jobs, sales jobs, police, mechanics, and electrician),

Table 1
List of Representative Jobs Included in Each Occupational Category

Occupational category	Title	Jobs
00	Engineering	Engineer, IT technician, industrial engineer
02	Chemistry	Chemist
18	Managerial	Catering manager, banking manager, clerical manager, engineering supervisor, team supervisor, store manager, contract manager, sales executive
20	Typing & filing	Clerk, auxiliary clerk, civil service administrative, clerical staff
23	Information & message	Mail order clerk, telephone operator, messenger, postman
25	Sales	Salesman, insurance salesman, district sales agent, sales assistant, sales order
37	Police	Police officer, urban police, police sheriff, local police
60	Mechanics	Mechanic, machine operator, industrial worker, general mechanical fitter, air mechanic
82	Electrical Assistant	Electrician
90	Driver	Car driver, bus driver, truck driver, tramway driver, caterpillar driver
	Miscellaneous skilled worker	Carpenter, industrial painter, railways agent, railways worker
	Apprentice	Craft apprentice, electrician apprentice, turner apprentice, industrial painter apprentice, milling machine apprentice

Note. IT = information technology.

plus the apprentice category and the majority (over 80% of studies) of jobs included in our category 23 (information and message distribution jobs), were classified as jobs of medium complexity level. The jobs included in the category 90 (i.e., drivers) and the majority (over 80% of studies) of the jobs included in the miscellaneous group of skilled worker occupations were classified as jobs of a low complexity level. Jobs within a category but that did not fit to the complexity level of the majority of jobs included in the category were assigned to the correct group of complexity level. For example, messenger (DOT 230–663–010) included in the category 23 (information and message distribution jobs) was classified in the group of low complexity level; plotter (DOT 962–361–010) included in the miscellaneous category of skilled worker jobs was classified in the group of medium level of job complexity.

After the studies were collated and their characteristics recorded, the following step was to apply the psychometric meta-analysis of Hunter and Schmidt (1990, 2000). Psychometric meta-analysis estimates how much of the observed variance of findings across studies is due to artifactual errors. The artifacts considered here were sampling error, criterion and predictor reliability, and range restriction in GMA scores. To correct the observed validity for these last three artifacts, the most common strategy is to develop specific distributions for each of them. Some of these artifacts reduce the correlations below their operational value (e.g., criterion reliability and range restriction), and all of them produce artifactual variability in the observed validity (Carretta & Ree, 2000, 2001). In our analysis, we corrected the observed mean validity for criterion reliability and range restriction in the predictor.

Predictor Reliability

The reliability of predictors was estimated (a) from the coefficients reported in the studies included in the meta-analysis, and (b) using the coefficients published in the various test manuals. The appropriate coefficients are test–retest estimates of reliability (Schmidt & Hunter, 1999). We developed an empirical distribution of test–retest reliability for GMA measures. The average reliability was .83 (*SD* = .09). The predictor reliability estimates are only used to eliminate artifactual variability in the standard deviation of rho and are not used for correcting the observed validity, because our interest is in the operational validity of GMA and not in the theoretical value of GMA validity.

Criterion Reliability

In the present research, only studies using job performance ratings and training performance ratings as criteria were used. This choice was based on three reasons: (a) U.S. meta-analyses have used these two criteria extensively, and one of our objectives was to provide a comparison with those meta-analyses. Consequently, for this initial meta-analysis of EC studies, it was important to retain the same criteria. (b) Other criteria, such as tenure, turnover, promotions, or output, were used in a very small number of studies in our database; therefore, we would not be able to carry out meta-analyses in these cases. (c) Job performance and training success have been used extensively in the industrial–organizational psychology literature as early performance criteria, and in establishing the operational validity of GMA tests, these measures are of paramount importance. Not all studies provided information regarding the criteria reliability (i.e., job performance ratings and training ratings), and consequently, we had to develop empirical distributions for these two criteria. Fortunately, a number of studies provided reliability coefficients for estimating criterion reliability.

For job performance ratings, the interrater reliability coefficient is the one of interest (Hunter, 1986; Schmidt and Hunter, 1996) because if this type of reliability is used in the correction for attenuation, it will correct most of the unsystematic errors in supervisor ratings (Hunter & Hirsh, 1987). We found 19 studies reporting interrater coefficients of job performance ratings (see Table 2). The average coefficient weighted by sample size was .52 (*SD* = .19). This coefficient is exactly the same coefficient found by Viswesvaran, Ones, and Schmidt (1996) in their meta-analysis of the interrater reliability of job performance ratings. In addition, Rothstein (1990) showed that the asymptotic value for job performance ratings was .52.

It must be pointed out that the two lowest values of job performance reliability reported in Table 2 are .29 and .26, which are unusually low reliabilities and might be considered outliers. Some readers might think that they are leading to higher estimated rho values and higher estimates of variance accounted for by artifacts. However, the effects of these two are very small. The distribution without these two values would have a mean of 0.59. Because the square root of the mean would be used for correcting the observed validity, the difference is 0.72 (19 reliability values, *M* = 0.52) compared with 0.76 (17 reliability values, *M* = 0.59), which is practically irrelevant. For example, correcting 0.25 for 0.72 is 0.347, whereas correcting for 0.76 is 0.329. In other words, the effect is less than 0.02. The effects on the variance accounted for by artifacts are equally irrelevant.

In the case of training success, ratings by trainers or supervisors were used in most primary studies included in this meta-analysis. A small number of primary studies used pass–fail qualifications given by the supervisor or used the ratings given by teachers or external examiners in both theoretical and practical examinations. Therefore, all training success

Table 2
Interrater Reliability Empirical Distribution of Training Success and Job Performance Ratings in the European Studies

	Training success rating		Job performance ratings	
	Sample	r_{yy}	Sample	r_{xx}
	152	.73	61	.80
	259	.71	170	.78
	156	.65	45	.75
	91	.65	96	.74
	182	.64	67	.74
	44	.61	8	.73
	333	.57	60	.72
	361	.57	90	.65
	89	.52	34	.60
	347	.50	36	.60
	171	.50	151	.58
	51	.50	91	.58
	524	.46	64	.52
	90	.46	30	.49
	47	.31	64	.48
			69	.42
			306	.39
			109	.29
			351	.26
M	0.56		0.52	
SD	0.09		0.19	
N	2,897		1,936	

scores used in the present meta-analysis can be considered as a form of rating. In fact, when a theoretical and a practical examination were given to the trainees, the correlation between ratings of examinations ranged from .46 to .73. We found 15 studies reporting training success reliability (see Table 2). The weighted sample-size average reliability was .56 (*SD* = .09). This figure is remarkably lower than the value of .81 used by Hunter and Hunter (1984). We considered that our estimate is more representative of training success reliability than Hunter and Hunter's estimate for three reasons: (a) It was empirically developed using civil studies, whereas Hunter and Hunter's estimate was an assumed one; (b) Hunter's estimate was found for training success measures in the U.S. Navy (Hunter, 1986), and these measures typically consisted of objective examinations (Vineberg & Joyner, 1989); (c) all the coefficients found in this research were calculated using ratings given by trainers or supervisors. Vineberg and Joyner (1989) affirmed that written tests and hands-on tests are the two types of measures most frequently used to evaluate trainee achievement in military contexts and that instructor ratings of trainee performance are very rarely used. Campbell, McHenry, and Wise (1990) reported a reliability of .53 (split-half) for hands-on test scores in Project A when they were used as a training success criterion.

Range Restriction Distributions

The distributions for range restriction were based on the following two strategies: (a) Some range restriction values were obtained from the studies that reported both restricted and unrestricted standard deviation data, and (b) another group of range restriction values was obtained using the reported selection ratio. To use the reported selection ratio, we used the formula reported by Schmidt, Hunter, and Urry (1976). This double strategy produced a large number of range restriction estimates. We grouped the range restriction values in three categories, depending on the type of job used in the primary study. The three groups parallel Hunter and Hunter's (1984) job complexity classification, but we used three groups instead of five. The range restriction values of GMA appear in Table 3. There are two reasons behind this procedure. First, the selection process results in a greater restriction for highly complex jobs than for lower complex jobs because of the fact that applicants tend to gravitate toward jobs with a complexity similar to their cognitive abilities and skills. In addition, selection processes for highly complex jobs screen out all except the very intelligent, but this is not true for the low complex jobs. Second, the three complexity groups are three practically nonoverlapping clusters of the occupational groups that are differentiated by job complexity. Consequently, if we use a single range restriction distribution for all the jobs, we would be undercorrecting the validity for high complexity jobs and overcorrecting the validity for medium and low complexity jobs. In the present study we used the range restriction distribution of high complexity level for the categories 00, 02, and 18 (i.e., engineers, chemists, and managers). We used the range restriction distribution of medium complexity level for the categories 20, 23, 25, 37, 60, 82, (i.e., typing and filing jobs, sales jobs, police, mechanics, and electrician), plus the apprentice category. Finally, we used the range restriction distribution of low complexity level occupations for the category 90 (i.e., drivers) and the miscellaneous group of skilled worker occupations.

Results

Validity Generalization Results for GMA–Occupation–Job Performance Combinations

The results of the meta-analyses for the nine occupational categories with job performance as criterion appear in Table 4. We did not find validity studies for chemical jobs, mechanic jobs, and the apprentice group; therefore, these occupations are not included.

Table 3
Range Restriction Distributions for General Mental Ability in Three Degrees of Job Complexity and Two Criteria

	High complexity			Medium complexity			Low complexity		
	u	F	C	*u*	F	C	*u*	F	C
	.95	1	J	.93	1	J	.93	1	T
	.66	1	T	.92	1	J	.87	1	J
	.60	1	J	.91	1	J	.78	1	J
	.55	1	J	.84	1	J	.73	1	T
	.45	1	J	.83	1	T	.36	1	J
	.41	1	J	.76	1	T			
	.40	1	J	.68	1	J			
	.37	1	J	.66	1	T			
	.35	1	T	.64	1	J			
	.28	1	J	.48	1	J			
	.27	1	T	.45	1	J			
	.27	1	J	.44	1	J			
				.44	1	T			
Job performance									
M	0.47			0.69			0.67		
SD	0.20			0.19			0.22		
N	761			1,046			272		
Training success									
M	0.43			0.67			0.83		
SD	0.17			0.15			0.10		
N	609			1,364			354		

Note. *u* = range restriction ratio, calculated as s/S, where s is the restricted standard deviation and S the unrestricted standard deviation. F = frequency; C = criterion; J = job performance; T = training success.

Table 4
Meta-Analysis of General Mental Ability–Occupation Combinations for Predicting Job Performance Ratings

Source	K	N	r	S_r^2	SD_r	% VE	S_e^2	ρ	SD_ρ	90% CV	L_ρ	NSD	LCV
Driver	5	394	.22	.025	.159	64	.012	.45	.191	.21	.40	.225	.11
Electrician	3	280	.28	.023	.150	65	.009	.54	.171	.32	.45	.215	.17
Information clerk	5	890	.31	.006	.081	100	.005	.61	.000	.61	.56	.175	.33
Engineer	9	837	.23	.014	.118	100	.010	.63	.000	.63	.57	.181	.34
Manager	6	783	.25	.039	.197	40	.007	.67	.407	.15	.61	.433	.06
Police	5	619	.12	.015	.121	31	.007	.24	.151	.05	.22	.159	.01
Sales	5	394	.34	.018	.133	100	.010	.66	.000	.66	.60	.189	.36
Skilled worker	7	994	.28	.002	.045	100	.006	.55	.000	.55	.51	.161	.30
Typing	23	1,870	.23	.032	.179	47	.011	.45	.256	.12	.41	.276	.05

Note. K = number of studies; r = observed mean validity; S_r^2 = observed variance; S_e^2 = sampling error variance; SD_r = observed standard deviation; ρ = operational validity; SD_ρ = standard deviation of the operational validity; % VE = variance accounted for by artifactual errors; L_ρ = lowest (hypothetical) rho; NSD = hypothetical standard deviation of L_ρ; LCV = lowest (hypothetical) CV.

The number of studies for each GMA–occupation–criterion combination ranged from 3 to 23, and the single sample sizes ranged from 280 to 1,870. In the table, from left to right, the first two columns show the number of validity coefficients and the total sample size. The next five columns are the average observed validity weighted by the sample size, the observed variance weighted by the sample size, the observed standard deviation weighted by the sample size, the percentage of variance accounted for by all the artifactual errors (i.e., predictor and criterion reliability, predictor range restriction, and sampling error), and the sampling error variance. The three columns that follow show the operational validity (observed validity corrected by criterion reliability and predictor range restriction), the standard deviation of the operational validity, and the 90% credibility value.

The largest operational validity was found for the managerial occupations. In this case, the operational validity was .67, and 40% of the observed variability was accounted for by artifactual errors. The 90% CV in this case was .15. Therefore, GMA measures were shown to be valid predictors of job performance for this occupational group, and they generalized validity across samples and countries. The next occupational category with the largest validity was sales occupations. The operational validity was .66, and all the observed variability was accounted for by artifactual errors. However, in the sales occupational group, the predicted artifactual variance was larger than 100 (we rounded the variance accounted for down to 100%), indicating that there is a second-order sampling error. Therefore, it is possible that the set of coefficients is not completely representative of the population. However, Hunter and Schmidt (1990) suggested that a second-order sampling error usually has only a small effect on the validity estimates and that it affects the meta-analytic estimates of standard deviations more than it affects the estimates of means.

The validity for the engineer occupations was .63, which was the third largest estimate. For engineer jobs, the 90% CV was .63 because all the observed variance was explained by artifactual errors. Therefore, there is evidence of validity generalization for this occupational group. The next occupational category with the largest operational validity was the information and message distribution clerk. The validity was .61 and all the observed variance was explained by the artifactual errors. Ranked immediately below, two occupational groups showed coefficients of .54 and .55, and there is validity generalization in the two cases. These two occupational categories were the electrical assembling jobs and the miscellaneous category of skilled workers. In the case of skilled workers, all the observed variability was accounted for by artifactual errors, and the explained variance was 65% for the electrical assembling jobs. The next occupational category, the typing and filing occupations, showed an operational validity of .45, and the 90% CV was .12. The artifactual errors accounted for 47% of the observed variability in these occupations. For drivers, the operational validity was .45 and the 90% CV was .21. In this case, the artifactual errors explained 64% of the observed variance. The police occupations had the lowest value of operational validity. In this group, we found an estimate of .24, and the 90% CV was .05. The artifactual errors explained 31% of the observed variance. Therefore, there is evidence of validity generalization for this occupational group.

Because of the fact that there were a relatively small number of studies in some distributions and that second-order sampling error was shown for some occupational groups, we conducted a so-called file-drawer analysis (Rosenthal, 1979; Hirsh et al., 1986) to ensure that all the studies in our review were relevant. With regard to this point, Ashworth, Osburn, Callender, and Boyle (1992) developed a method for assessing the vulnerability of validity generalization results to unrepresented or missing studies. Ashworth et al. (1992) suggested calculating the effects on validity when 10% of the studies are missing and when their validity is 0. Therefore, we calculated additional estimates to represent what the validity would be if we were unable to locate 10% of the studies and if these studies showed zero validity. The last three columns of Table 4 report these new (hypothetical) estimates for every occupational group: lowest rho value, new standard estimation, and lowest 90% CV. As can be seen, the increase in 10% of studies showing zero validity had only minimal effects on the operational validity and affected the estimates of the standard deviation little more. The hypothetical lowest 90% CVs were also all positive and different from 0. Consequently, even in the case of there being an addition of 10% of studies with zero validity, there was still validity generalization for all occupational groups in the case of the job performance criterion.

An important finding was that, for eight out of nine occupational groups across the EC, GMA tests showed high to very high operational validity, with the exception of the police category for which the operational validity was somewhat lower. Comparing our results with those reported in the U.S. meta-analyses for similar occupational groups, in general, our findings are similar or somewhat larger than the U.S. findings, although the content of the occupational categories is not necessarily the same in the EC and the United States. For example, the reanalysis of Ghiselli's (1966, 1973) findings by Hunter and Hunter (1984) reported operational validity estimates of .54, .28, .53, .42, .61, .48, and .46 for clerical, driver, manager, protective, sales, service, and trade occupations, respectively. In other meta-analyses, the results have also been similar. For example, Pearlman et al. (1980) reported validities of .50 for typing and filing occupations, Hirsh et al. (1986) reported an operational validity of .22 for law enforcement occupations, and Levine et al. (1996) reported an operational validity of .53 for electrical general occupations in the utility industry, although the validity only reached .31 for electric assembly jobs.

In summary, the meta-analyses of the occupation–job performance combinations for GMA measures showed that they are valid predictors for all occupational categories and that they generalized validity across samples and countries of the EC. The results also showed that the magnitude of the operational validity was notably high for some occupational groups, namely managerial, sales, engineer, and informational and message distribution clerks jobs. Only in the case of police occupations was the magnitude of the validity relatively small. The fact that our results largely replicated the U.S. findings for these occupational groups is of great importance because it shows that the validity of GMA tests is not limited to the United States.

Validity Generalization Results for GMA–Occupation–Training Success Combinations

The results of the meta-analysis for the 10 occupational groups with training success as criterion appear in Table 5. In the present analyses, we did not find an acceptable number of studies for carrying out meta-analyses for managerial and sales occupations. The number of studies for each GMA–occupation–training success combination ranged from 3 for police to 12 in the miscellaneous group of skilled workers. The single sample sizes ranged from 353 for the electrical assembly category to 2,276 for the skilled workers. The columns in Table 5 are the same as for Table 4.

The largest operational validity for predicting training success was found for the engineer occupational category, with the magnitude of the validity equal to .74. The 90% CV was .74, and 100% of the observed variability in validities was accounted for by artifactual errors. Therefore, the results indicate that GMA measures have generalized validity across samples and countries for this occupational group. The next group with the largest validity was the chemist job category. For this occupational category, GMA measures showed an operational validity of .72; the 90% CV was also .72 because the explained variance was 100%. Therefore, for this occupational group, there is also validity generalization across countries and samples. The next occupational category with the largest operational validity was the information and message distribution clerks. In this case, GMA measures showed an operational validity of .69, and all the observed variability was accounted for by the artifactual errors.

GMA measures were also found to be valid predictors of training success for a group of five occupational categories with the operational validity ranging from .40 to .63. The operational validity for the electrical assembling jobs was .63, the 90% CV was .48, and the percentage of explained variance was 73%. For the typing clerks category, the operational validity was .57, the 90% CV was .24, and the explained variance was 31%. The next occupational group was the apprentices, which showed an operational validity of .49, the 90% CV was .39, and the artifactual errors accounted for 84% of variability. For the driver occupational category, the operational validity was .40, the 90% CV was .32, and the artifactual errors accounted for 72%. Last, for the mechanics jobs, the validity was .40, the 90% CV was .29, and 82% of variability was explained by artifactual errors. Consequently, in all of these occupational categories, GMA showed validity generalization across samples and countries for predicting training success.

The last two occupational categories analyzed were police jobs and the miscellaneous group of skilled workers. In the case of police jobs, the operational validity was .25, and the 90% CV was −.07. Consequently, we cannot conclude that there is validity generalization for this occupational category. This counterintuitive result for the police job is probably due to the effect of a possible outlier value. If this value is deleted, the operational validity is .48, and the 90% CV is .48. Finally, for the skilled workers, the

Table 5
Meta-Analysis of General Mental Ability–Occupations Combinations for Predicting Training Success

Category	K	N	r	S_r^2	SD_r	% VE	S_e^2	ρ	SD_ρ	90% CV	L_ρ	NSD	LCV
Apprentice	9	1,229	.26	.010	.101	84	.006	.49	.075	.39	.45	.158	.24
Chemistry	4	1,514	.28	.003	.051	100	.002	.72	.000	.72	.65	.206	.39
Driver	9	2,252	.26	.006	.077	72	.003	.40	.062	.32	.36	.129	.20
Electrician	4	353	.35	.017	.131	73	.009	.63	.121	.48	.57	.214	.30
Information clerk	4	579	.46	.002	.041	100	.004	.69	.000	.69	.63	.198	.37
Engineer	8	1,051	.28	.013	.115	100	.006	.74	.000	.74	.65	.207	.39
Mechanics	4	549	.21	.010	.100	82	.007	.40	.079	.29	.36	.137	.19
Police	3	392	.13	.026	.162	31	.007	.25	.255	−.07	.23	.253	−.10
Skilled worker	12	2,276	.17	.012	.108	47	.005	.27	.121	.12	.25	.139	.07
Typing	12	1651	.31	.029	.171	31	.006	.57	.259	.24	.52	.296	.14

Note. K = number of studies; r = observed mean validity; S_r^2 = observed variance; S_e^2 = sampling error variance; SD_r = observed standard deviation; ρ = operational validity; SD_ρ = standard deviation of the operational validity; % VE = variance accounted for by artifactual errors; L_ρ = lowest (hypothetical) rho; NSD = hypothetical standard deviation of L_ρ; LCV = lowest (hypothetical) CV.

operational validity was .27, and the 90% CV was .12. Therefore, there is validity generalization for this occupational category.

The results of the file-drawer analyses using the Ashworth et al. (1992) method appear in the last three columns of Table 5. Adding 10% of the studies with zero validity had no effect on our conclusions for the training success criterion because the magnitude of the new rho estimates was very similar and it was shown that there is validity generalization for all occupational groups, except for police jobs.

We can now compare our results with previous U.S. results, as we have done for the job performance criterion. Our results for electrical assembly category are lower than the value found by Levine et al. (1996) because these authors reported an operational validity of .77. Pearlman et al. (1980) reported an average operational validity of .80 and .54 for GMA measures in typing and filing clerks and information and message distribution clerks, respectively. Therefore, we have found a larger value for information clerks and a lower value for typing clerks. For police occupations, Hirsh et al. (1986) reported an average validity of .56 for GMA measures, whereas our results showed a remarkably lower operational validity.

In summary, we found that GMA measures are valid predictors for the 10 occupational groups we analyzed here. They showed large operational validities ($\rho > .50$) for 5 out of the 10 groups and moderately large operational validities for 3 additional groups (ρ = .40, .45, and .48, respectively). We also found that GMA measures generalized validity across samples and countries for 9 out of the 10 occupational groups. In general, our results are similar, although slightly lower than the operational validity estimates found in the U.S. meta-analyses for the same occupational groups.

Meta-Analysis of the Job Complexity–Criterion Validity Relations

In Table 6, the results of the meta-analyses carried out for the three job complexity levels and for the two criteria used in the present research are presented. The results for job performance show that job complexity has a positive effect on the validity magnitude of GMA measures because there are different sizes for each complexity level.

As can be seen, the validity for the high level of job complexity is .64. For the medium complexity level, the validity magnitude is .53, and for the low level of job complexity the operational validity is .51. These sizes parallel the values found by Hunter and Hunter (1984). For example, we found a value of .53 for the medium level of job complexity, which is exactly the same value reported by Hunter and Hunter. This is an especially relevant finding because the medium level of job complexity grouped 62.7% of the jobs (Hunter & Hunter, 1984). Another relevant replication is that we found a value of .51 for the low level of job complexity, and Hunter and Hunter found a value of .50, which is practically the same. Finally, Hunter and Hunter found a value of .59 for the highest level of job complexity, and we found a value of .64, which is very close. Therefore, we replicate the previous findings for job performance.

In the case of training success, job complexity also shows effects on the GMA validity, but these effects are more remarkable. The validity for the high level of job complexity is .74, for the medium level it is .53, and for the low level it is .36. Therefore, it is clear that the validity of GMA increases when job complexity increases, showing a linearly positive relationship between GMA and job complexity. These results also replicate previous findings by Hunter and Hunter (1984). Consequently, the previous conclusion that job complexity is a powerful moderator of GMA validity is supported by the present results.

Discussion

GMA measures have been shown to be highly valid predictors of job performance and training success criteria in personnel selection in the United States. Furthermore, meta-analytical research has provided empirical evidence that GMA measures show validity generalization across tests, samples, and organizations. In addition, it has been shown that the magnitude of the operational validity is different for the various occupational categories. However, such investigations only included primary studies carried out in the United States. Our examination of the operational validity of GMA measures in EC countries with similar and additional occupational categories was a relevant research question from both a theoretical and an applied point of view. The EC member countries represent an important replication test for earlier U.S. meta-analytical findings because they are all developed, postindustrial economies similar to the United States, with a population of over 400 million, but with different cultural, historical, and employment legislation frameworks (Herriot & Anderson, 1997).

Table 6
Meta-Analysis of Job Complexity–General Mental Ability Combinations for Predicting Job Performance and Training Success

Category	K	N	r	S_r^2	SD_r	% VE	S_e^2	ρ	SD_ρ	90% CV	L_ρ	NSD	LCV
Job performance													
High complexity	14	1,604	.23	.024	.154	66	.008	.64	.242	.33	.58	.295	.20
Medium complexity	43	4,744	.27	.031	.175	43	.008	.53	.255	.21	.48	.287	.11
Low complexity	12	864	.25	.021	.146	87	.012	.51	.102	.38	.46	.175	.24
Training													
High complexity	13	2,619	.29	.009	.094	100	.004	.74	.000	.74	.67	.213	.40
Medium complexity	35	4,304	.29	.024	.155	40	.007	.53	.224	.24	.48	.258	.15
Low complexity	21	4,731	.23	.014	.117	34	.004	.36	.145	.18	.33	.173	.11

Note. K = number of studies; r = observed mean validity; S_r^2 = observed variance; S_e^2 = sampling error variance; SD_r = observed standard deviation; ρ = operational validity; SD_ρ = standard deviation of the operational validity; % VE = variance accounted for by artifactual errors; L_ρ = lowest (hypothetical) rho; NSD = hypothetical standard deviation of L_ρ; LCV = lowest (hypothetical) CV.

The findings of the present research showed that GMA measures are valid predictors of job performance and that they generalized validity across samples and occupations within the EC for predicting job performance and training success. In the case of job performance, GMA measures showed large operational validities for nine occupational groups, including driver jobs, electrical assembling jobs, information and message distribution clerks, engineer and managerial occupations, sales jobs, typing and filing clerks, and police jobs. Furthermore, the results showed larger validity for the miscellaneous group of skilled worker jobs. In general, the results were similar, and in a few cases slightly higher, to those found in the United States. On the basis of the 90% CVs, it is possible to conclude that there is international validity generalization for GMA measures for predicting job performance. A possible explanation for this difference between the EC and the U.S. results is that there are differences in the conceptualization of work performance. Another possibility is that the jobs included in each occupational category are more cognitively complex jobs than the ones included in the U.S.-based meta-analyses. These findings are especially noteworthy given the cultural and historic variety of countries in the EC. That the magnitude of operational validity of GMA measures was equal to, or slightly greater than, those found in earlier meta-analyses carried out in the United States suggests even more convincingly that GMA measures are robust predictors of future job performance and training success across occupational categories, job complexity, and national cultures (Schmidt, Ones, & Hunter, 1992; Salgado & Anderson, 2002). This is, to our knowledge, one of the only occasions in which U.S. findings of GMA effects sizes have been subjected to international verification in a multicultural context such as the EC (Anderson, Lievens, van Dam & Ryan, in press). Such findings contribute to an increasingly global nature of science and practice in industrial–organizational psychology in that it can be ascertained whether earlier findings from the United States generalize to other continents, countries, and cultures. These findings present unambiguous evidence, based on a large sample of primary studies across seven European countries, that crucially important earlier findings are generalizable, and indeed in some cases operational validity of GMA measures in selection may be slightly higher in the EC than in the United Sates.

Regarding the training success criterion, the findings show that GMA measures were valid predictors for all occupational groups across the EC, except for police jobs. The findings demonstrate generalized validity for 9 out of 10 occupational groups examined here, including apprentice jobs, chemist jobs, driver jobs, electrical assembling jobs, information and message distribution clerks, engineer occupations, mechanic jobs, skilled worker jobs, and typing and filing clerks. On the basis of the 90% CVs, there is empirical evidence for concluding that there is international validity generalization of GMA measures for predicting training success in the EC. In relation to the explained variance, we found that it was accounted for to a large extent by artifactual errors, ranging from 36% for the typing clerks to 100% for the chemist jobs and the information and message distribution clerks. In addition, the pattern of the validity coefficients was similar to those found in the U.S. meta-analyses, although in some cases, the magnitude of the coefficients was smaller in the European studies. Consequently, when taking the results for training success in Europe and the United States as a whole, it is clear that there is international validity generalization for this criterion. Police jobs were the only occupational group that showed a negative 90% credibility value and, consequently, there was no evidence for validity generalization. However, as we already pointed out, this result could have been due to an outlier value, and therefore we suggest more single studies on the validity of GMA measures in this occupational group.

Another relevant finding is related to the fact that the magnitude of the operational validity was in general larger for training success than for job performance ratings. The findings of the U.S. meta-analyses, in which the effects of the occupational group were analyzed, suggests that the validity was also larger for training success than for job performance, although the results are not completely conclusive. For example, Ghiselli's (1966, 1973) results, as reanalyzed by Hunter (1986; Hunter & Hunter, 1984), indicated that the validity for training success was larger in eight out of nine occupational groups. Managerial occupations were the exception. Hirsh et al. (1986) and Pearlman et al. (1980) found the same results for police jobs and clerical occupations that we found for these two categories. In the present meta-analysis, we found that the validity was larger for training success than for job performance in six out of seven occupational groups: driver jobs, electrical assembling jobs, engineer occupations, information and message distribution clerks, police jobs, and typing and filing clerks. The exception was the miscellaneous group of skilled worker jobs. These last results may be due to the fact that the type of job is not exactly the same for the job performance criterion than for the training success criterion. Therefore, in general, our results supported the hypothesis that the operational validity is larger for training success performance than for job performance success when the validity coefficients are grouped based on the occupational category.

Limitations and Strengths of the Present Study

The present study also has several limitations. First, the number of studies and the total sample size are smaller in some occupational categories than in others. Another limitation is that some occupational groups may have also been underrepresented (e.g., police jobs) or not represented at all in some criteria (e.g., manager–training combination) in our database. A third limitation is that we were not able to examine whether there is validity variance associated with the countries because we do not have a sufficient number of jobs (studies) per country in order to do this analysis. In this sense, the occupational categories in our database may be more heterogeneous than the categories used in the U.S. meta-analyses. Another limitation is that our study was conducted only with civilian samples, and therefore, the results may not be generalized to military occupations.

Counter to these limitations, the present meta-analysis also possesses several strengths. First, we were able to obtain primary criterion-related validity studies across seven member states of the EC, ranging from northern countries (e.g., the Netherlands, Germany) to southern countries (e.g., Spain). Second, our computer-based, manual, and in-person searches to locate primary studies resulted in a final database covering 12 occupational categories, which ranged considerably in job complexity. Third, the present meta-analysis benefited greatly from recent advances in meta-analytic methods made by pioneering scholars of this technique in the social sciences in the United States (e.g. Hunter & Schmidt, 2000). In our procedure, we incorporated these methodological advances at each stage and whenever possible explained our rationale for decisions made throughout this meta-analytic investigation. Thus, in balance, despite some unavoidable limitations to the present study, we argue that these strengths should also be borne in mind, given that this is one of the first attempts to comprehensively verify the international generalizability of the magnitude of criterion-related validity of GMA measures for personnel selection.

Scientific and Theoretical Implications

Our findings, together with the U.S. meta-analytic findings, have implications from a theoretical point of view. In effect, as with the U.S. meta-analyses, we found that job complexity is a pervasive moderator of the validity of GMA. Hunter and Hunter (1984) found evidence supporting this hypothesis. In our case, we found that GMA showed larger validity for the occupational categories theoretically associated with higher job complexity. For the job performance criterion, the larger validity coefficients were for managerial jobs and engineer occupations. In the case of training success, the larger validity coefficients were for engineer and chemist occupations. Furthermore, we found that job complexity positively moderates the validity magnitude of GMA tests for both job performance and training success, with values practically identical to the ones found by Hunter and Hunter.

Another finding with theoretical implications is that in some cases the magnitude of the validity was somewhat larger in the European studies than in the United States for job performance. However, for training success, the findings are contrary, although of smaller magnitude. Consequently, a variable must exist which moderates the validity in the European studies but not in the U.S. studies (or to a lesser extent). In the case of training success, one possible explanation is that there were differences in the way in which this criterion was measured. In Europe, training success is typically assessed using ratings, whereas in the United States it is commonly assessed with objective measures (Ackerman & Humphreys, 1991).

Practical Implications for Organizational Selection Processes

Our results also have implications from an applied point of view. One unequivocal implication from this large-scale meta-analysis is that selection practitioners in the EC countries may use GMA measures across all the occupational groups because they are highly valid predictors for job performance and training success. However, the findings are also relevant for the selection of expatriate individuals because the main findings indicate that there is validity generalization across countries in Europe and the United States. GMA tests are therefore likely to be robust predictors for expatriate assignments across these two continents. However, it is important to note that for expatriate assignments on other continents, caution in extending the present meta-analytic findings is in order. Future research is called for to examine the global generalizability of the present U.S. and EC meta-analytic findings to other continents and countries worldwide, given the postindustrial cultural similarities between the U.S. and the EC. Nevertheless, it should be taken into account that we examined only two criteria, and future studies should explore whether the pattern of validity coefficients is true for other relevant organizational criteria (e.g., turnover, tenure, absenteeism, output, career development). Previous meta-analytic findings for the North American continent and now our findings for the European continent unequivocally support the use of tests of GMA for personnel selection regardless of job complexity or job occupation being selected for.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ackerman, P., & Humphreys, P. I. (1991). Individual differences theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology, Vol. 1* (pp. 223–282). Palo Alto, CA: Consulting Psychologist Press.

*Amthauer, R. (1973). *I-S-T 70. Intelligenz-Struktur-Test* [I-S-T 70. Intelligence-Structure-Test]. Gottigen: Hogrefe.

*Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (1999). *I-S-T 2000. Inttelligenz-Struktur-Test 2000.* [I-S-T 2000. Intelligence-Structure-Test 2000]. Gottingen, Germany: Hogrefe.

Anderson, N., Lievens, F., van Dam, K., & Ryan, A. M. (in press). Future perspectives on employee selection: Key directions for future research and practice. *Applied Psychology: An International Review.*

*Anstey, E. (1963). *D-48. Manual.* Madrid, Spain: Tea.

Ashworth, S. D., Osburn, H. G., Callender, J. C., & Boyle, K. A. (1992). The effects of unrepresented studies on the robustness of validity generalization results. *Personnel Psychology, 45,* 341–361.

*Banister, D., Slater, P., & Radzan, M. (1962). The use of cognitive tests in nursing candidate selection. *Occupational Psychologist, 36,* 75–78.

*Beaton, R. (1982). Use of tests in a national retail company. In K. Miller (Ed.), *Psychological testing in personnel assessment* (pp. 137–148). Epping, United Kingdom: Gower Press.

*Blanco, M. J., & Salgado, J. F. (1992). Diseño y experimentación de un modelo de selección de personal para conductores de materias peligrosas (factores humanos y conducción) [Design and experimentation of a personnel selection model for drivers of dangerous substances]. In M. J. Blanco, J. F. Salgado, F. J. Alvarez, C. del Rio, R. Prada, & G. Buela-Casal (Eds.), *Aportaciones de la psicología a la seguridad vial.* Madrid, Spain: Mapfre.

*Boerkamp, R. G. (1974). *Een criteriuminstrument voor de bedrijfsselectie-situatie* [A criterion measure for applicant-selection in companies]. Unpublished doctoral dissertation, University of Amsterdam, the Netherlands.

*Bonnardel, R. (1949). Examens psychometriques et promotion ouvriere (Étude portant sur un groupe d'ouvriers électiciens en cours de perfectionnement) [Psychometric examination and worker promotion (A study with a group of electrician workers in a training course)]. *Le Travail Humain, 12,* 113–117.

*Bonnardel, R. (1949). L'emploi des méthodes psychométriques pour le controle des condictions psychologiques du travail dans les ateliers [The use of psychometric methods for controlling the psychological condictions of work in factory]. *Le Travail Humain, 12,* 75–85.

*Bonnardel, R. (1949). Recherche sur la promotion des ouvriers dans le cadres de maitrise [Research on the promotion of workers to supervisors]. *Le Travail Humain, 12,* 245–256.

*Bonnardel, R. (1954). Appréciations profesionnelles et notations psychométriques étude portant sur un groupe de jeunes ouvriers [Professional qualifications and psychometric scores in a group of young workers]. *Le Travail Humain, 17,* 119–125.

*Bonnardel, R. (1954). Examen de chauffeurs de caminos au moyen de tests de réactions [Examination of drivers using tests of reactions]. *Le Travail Humain, 17,* 272–281.

*Bonnardel, R. (1956). Un example des difficultés soulevées par la question des critéres professionnels [An example of the difficulties related to the question of job criteria]. *Le Travail Humain, 19,* 234–237.

*Bonnardel, R. (1958). Comparaison d'examens psychométriques de jeunes ingenieurs et de cadres administratifs [Comparison of psychometric examinations of young engineers and administrative supervisors]. *Le Travail Humain, 21,* 248–253.

*Bonnardel, R. (1959). Liaisons entre résultats déxamens psychométriques et réussite dans le perfectionnement profesionnel ouvrier [Relationships between the psychometric examinations and the success in the improvement of professional worker]. *Le Travail Humain, 22,* 239–246.

*Bonnardel, R. (1961). Recherche sur la promotion des ouvriers dans la maitrise [Research on the promotion of workers in industry]. *Le Travail Humain, 24,* 21–34.

*Bonnardel, R. (1962). Recherche sur le recruitment a des cours d'étude du travail au moyen de techniques psychométriques [Research on the recruitment by psychometric procedures]. *Le Travail Humain, 25,* 73–78.

*Bruyere, M. J. (1937). Quelques données sur l'intelligence logico-verbale et les aptitudes techniques pour l'orientation vers la carrier d'ingenieur [Soma data on logic-verbal intelligence and technical aptitudes for the orientation to the engineer studies]. *Bulletin de l'Institut Nationalle d'Orientation Professionelle, 9,* 141–147.

*Burt, C. (1922). Test for clerical occupations. *Journal of the National Institute of Industrial Psychology, 1,* 23–27 & 79–82.

Campbell, J. P., McHenry, J. J., & Wise, L. L. (1990). Modeling job performance in a population of jobs. *Personnel Psychology, 43,* 313–333.

*Campos, F. (1947). Notas del Laboratorio Psicotecnico de la institución “Virgen de la Paloma” [Notes of the psychotechnic laboratory of the “Virgen de la Paloma” institution]. *Revista de Psicología General y Aplicada, 2,* 545–549.

Carretta, T., & Ree, M. J. (2000). General and specific cognitive and psychomotor abilities in personnel selection: The prediction of training and job performance. *International Journal of Selection and Assessment, 8,* 227–236.

Carretta, T., & Ree, M. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325–335.

*Castillo J., Bueno, R., Moldes, F., Fernández, J., & Bacas, M. (1969). Estudio estadístico del test de matrices progresivas de Raven, escalas general y especial [Statistic study of Raven Progressive Matrices Test: General and special scales]. *Revista de Psicología General y Aplicada, 24,* 1004–1009.

*Castillo-Martin, J. (1963). Estudio del Otis en sujetos de la industria eléctrica [A study of Otis test in electric industry subjects]. *Revista de Psicología General y Aplicada, 18,* 1015–1020.

*Castle, P. F. C., & Garforth, F. I. (1951). Selection, training and status of supervisors. *Occupational Psychologist, 25,* 109–123.

*Cattell, R. B. (1981). *Manual del test Factor G. Escalas 2 y 3.* [Manual for the Factor G Test, Scales 2 and 3]. Madrid, Spain: Tea.

*Childs, R. (1990). *Graduate and managerial assessment data suplement.* Windsor, Berkshire, United Kingdom: ASE, NFER-Nelson.

*de Buvry de Mauregnault, D. (1998). *Evaluatie selectieprocedure automatiseringsmedewerkers, validatieonderzoek aan de hand van task performance en contextual performance* [Evaluation of the selection procedure of automation employees: Validation research of task and contextual performance]. Unpublished manuscript, University of Amsterdam, the Netherlands.

*Evers, A. (1977). *De constructie van een testbatereij voor de selectie van leerling-verplegenden aan de 'De Tjonsgerschans'* [The construction of a test battery for the selection of student nurses in De Tjonsgerschans]. Unpublished manuscript..

*Farmer, E. (1930). A note on the relation of certain aspects of character to industrial proficiency. *British Journal of Psychology, 21,* 46–49.

*Farmer, E. (1933). The reliability of the criteria used for assessing the value of vocational tests. *British Journal of Psychology, 24,* 109–119.

*Farmer, E., & Chambers, E. G. (1936). *The prognostic value of some psychological tests. Industrial Health Research Board Report No. 74.* London, England: HM Stationary

*Feltham, R. (1988). Validity of a police assessment centre: A 1–19 year follow-up. *Journal of Occupational Psychology, 61,* 129–144.

*Frankford, A. A. (1932). Intelligence tests in nursing. *Human Factor, 6,* 451–453.

*Frisby, C. B. (1962). The use of cognitive tests in nursing candidate selection: A comment. *Occupational Psychology, 36,* 79–81.

*García-Izquierdo, A. (1998). *Validación de un proceso de selección de personal en un centro de formación del sector de la construcción* [Validity of a personnel selection process in a training center of the construction industry]. Unpublished doctoral dissertation, University of Oviedo, Oviedo, Spain.

*Germain, J., Pinillos, J. L., García-Moreno, E., & Aberasturi, N. L. (1969). La validez de unas pruebas selectivas para conductores [Validity of drivers' selection tests]. *Revista de Psicología General y Aplicada, 24,* 1067–1114.

*Goguelin, P. (1950). Recherches sur la sélection des conducteurs de véhicules [Research on the selection of drivers]. *Le Travail Humain, 13,* 9–35.

*Goguelin, P. (1951). Étude du poste délectricien de tableau et examen de selection pour le poste [A study of electrician jobs and examination of selection for the job]. *Le Travail Humain, 14,* 15–65.

*Goguelin, P. (1953). Étude du poste de dispatcher dans l'industrie eléctrique et de la selection pour ce poste [A study of dispatcher jobs and the selection for this job]. *Le Travail Humain, 16,* 197–205.

*González, A., Sanz, A., & Ferrer, A. (1984, May). *Prevención de accidentes mediante una batería psicofisiológica: nuestra experiencia en la compañía telefónica* [Accident prevention with a psycho-physiological battery: Our experience in a telephone company]. Paper presented at the Primera Reunión Internacional de Psicología y Seguridad Vial, Valencia, Spain.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests.* New York: Wiley.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology, 26,* 461–477.

*Greuter, M. A. M., & Smit-Voskuyl, O. F. (1983). *Psychologish onderzoek geëvalueeerd, een valideringsonderzoek naar de psychologische selectie bij de VNG* [Psychological assessment investigated: A validation study of the psychological selection process in VNG]. Amsterdam, the Netherlands: University of Amsterdam, Department of Psychology.

*Handyside, J. D., & Duncan, D. C. (1954). Four years later: A follow-up of an experiment in selecting supervisors. *Occupational Psychology, 28,* 9–23.

*Hänsgen, K. D. (2000). *Evaluation des eignungstests für das medizinstudium in der schweis-suverlässigkeit der vorhersage von studienerfolg* [Evaluation of aptitude test for medicine studies in Switzerland. Prognosis reliability of the studies success]. Fribourg, Switzerland: University of Fribourg, Department of Psychology, CTD Centre for the Test Development and the Diagnostic.

Hartigan, J., & Wigdor, A. K. (1989). (Eds). *Fairness in employment testing. Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

*Henderson, P., & Boohan, M. (1987). SHL's technical test battery: Validation. *Guidance and Assessment Review, 3,* 3–4.

*Henderson, P., & Hopper, F. (1987). SHL's TTB: Development of norms. *Guidance and Assessment Review, 3,* 5–6.

*Henderson, P., Lockhart, H., & O'Reilly, S. (1987). SHL's technical test battery: Test–retest reliability. *Guidance and Assessment Review, 3,* 4–5.

*Henderson, P., & O'Hara, R. (1990). Norming and validation of SHL's personnel test battery in Northern Ireland. *Guidance and Assessment Review, 6,* 2–3.

Herriot, P., & Anderson, N. (1997). Selecting for change: How will personnel selection psychology survive? In N. Anderson & P. Herriot (Eds.), *International handbook of selection and assessment* (pp. 1–34). London: Wiley.

Hirsh, H. R., Northrop, L., & Schmidt, F. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39,* 399–420.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Intelligence, 29,* 340–362.

Hunter, J. E., & Hirsh, H. R. (1987). Applications of meta-analysis. *International Review of Industrial and Organizational Psychology, Vol. 2* (pp. 321–357). Chichester, United Kingdom: Wiley.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative 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. (2000). Fixed effects vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8,* 275–292.

*Jäger, A. O., & Althoff, K. (1994). *Der WILDE-Intelligenz Test (WIT)* [WILDE Intelligence Test]. Gottigen, Germany: Hogrefe.

*Keizer, L., & Elshout, J. J. (1969). *Een onderzoek naar de validiteit van selectierapporten en enkele andere voorspellers* [A study of the validity of selection protocols and other predictors]. Unpublished report.

*Krielen, F. (1975). *De validiteit van de testbatterijen 'Middelbaar Algemeen' en 'Middelbaar Administratief '* [The validity of the test batteries "Middelbaar Algemeen" and "Middelbaar Administratief"]. Unpublished doctoral dissertation, University of Amsterdam, the Netherlands.

*Ledent, R., & Wellens, L. (1935). La selection et la surveillance des conducteurs des tramsways unifiés de Liége et extensions [Selection and vigilance of tramways drivers in Liége and area]. *Le Travail Humain, 3,* 401–405.

Levine, E. L., Spector, P. E., Menon, S., Narayanan, L., & Cannon-Bowers, J. (1996). Validity generalization for cognitive, psychomotor, and perceptual tests for craft jobs in the utility industry. *Human Performance, 9,* 1–22.

Levy-Leboyer, C. (1994). Selection and assessment in Europe. In H. C. Triandis, M. D. Dunnette, & L. M. Hough (Eds), *Handbook of industrial and organizational psychology, Vol. 4* (pp. 173–190). Palo Alto, CA: Consulting Psychologists Press.

*Mace, C. A. (1950). The human problems of the building industry: Guidance, selection and training. *Occupational Psychology, 24,* 96–104.

*McKenna, F. P., Duncan, J., Brown, I. D. (1986). Cognitive abilities and safety on the road: A re-examination of individual differences in dichotic listening and search for embedded figures. *Ergonomics, 29,* 649–663.

*Miller, K. (1982). Choosing tests for clerical selection. In K. Miller (Ed.), *Psychological testing in personnel assessment* (pp. 109–121). Epping, United Kingdom: Gower Press.

*Montgomery, G. W. G. (1962). Predicting success in engineering. *Occupational Psychologist, 36,* 59–80.

* Moran, A. (1986). The reliability and validity of Raven's standard progressive matrices for Irish apprentices. *International Review of Applied Psychology, 35,* 533–538.

*Morin, J. (1965). Notation psychométrique et profesionnelle de jeunes ingénieurs [Psychometric and professional scores of young engineers]. *Le Travail Humain, XX,* 201–211.

*Munro, M. S., & Raphael, W. (1930). Selection tests for clerical occupations. *Journal of the National Institute of Industrial Psychology, 5,* 127–137.

*National Institute of Industrial Psychology. (1932). The selection of salesmen. *Human Factor, 6,* 26–29.

*Nelson, A., Robertson, I. T., Walley, L., & Smith, M. (1999). Personality and work performance: Some evidence from small and medium-sized firms. *Occupational Psychologist, 12,* 28–36.

*Notenboom, C. G. M., & van Leest, P. F. (1993). Psychologische selectie werkt! [The psychological selection works]. *Tijdschrift voor de Politie, 3,* 65–67.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed). New York: McGraw-Hill.

*Nyfield, G., Gibbons, P. J, Baron, H., & Robertson, I. (1995, May). *The cross-cultural validity of management assessment methods.* Paper presented at the 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.

*Pacaud, S. (1946). Recherches sur la selection profesionnelle des opératrices de machines a perforer et de machines comptables. La « seléction omnibus » est-elle possible en mecanographie? [Research on the professional selection of punching clerks: Is it possible the omnibus selection in typing?]. *Le Travail Humain, 9,* 74–86.

*Pacaud, S. (1947). Sélection des mécaniciens et des chaufeurs de locomotive. Etude de la validité des tests employés et composition des batteries selectives [Selection of train mechanics and drivers: A study of validity of tests used and composition of the selection batteries]. *Le Travail Humain, 10,* 180–253.

Paoli, P., & Merllié, D. (2001). *The third European survey on working conditions, 2000.* Luxembourg: European Foundation for the Improvement of Living and Working Conditions, Office for Official Publications of the European Communities.

*Patin, J., & Vinatier, H. (1962). Un test verbal: Le CERP 15 [A verbal test: The CERP 15]. *Bulletin du CERP, 11,* 327–342.

*Patin, J., & Vinatier, H. (1963). Un test d'intelligence: Le CERP 2 [An intelligence test: The CERP 2]. *Bulletin du CERP, 12,* 187–201.

*Patin, J., & Vinatier, H. (1963). Le test CERP 14 du docteur Morali-Daninos [Docot Morali-Daninos' CERP 14 test]. *Bulletin du CERP, 12,* 381–390.

*Patin, J., & Nodiot, S. (1962). Le test mecaniques de P. Rennes [P. Rennes' mechanic test]. *Bulletin du CERP, 11,* 61–97.

*Patin, J., & Panchout, M. F. (1964). L'admission des candidats à une formation profesionelle de techniciens. Approche docimologique, validités du concours et de l'examen psychotechnique [Candidate admission for a training course of technics. Rationale, concurrent validity and psychometric examination]. *Bulletin du CERP, 13,* 147–165.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65,* 373–406.

*Petrie, A., & Powell, M. B. (1951). The selection of nurses in England. *Journal of Applied Psychology, 35,* 281–286.

*Roe, R. A. (1973). *Een onderzoek naar de validiteit van een selectieprocedure voor lager en uitgebreid lager administratieve functies* [A validity study of a selection procedure for lower level administrative jobs]. Unpublished manuscript, University of Amsterdam, the Netherlands.

Roe, R. A. (1989). Designing selection procedures. In P. Herriot (Ed.). *Assessment and selection in organizations* (pp. 127–142). Chichester, United Kingdom: Wiley.

*Roe, R. A., & Boerkamp, R. G. (1971). *De selectie vooor functies uit de groep Uitgebreid Algemeen. Een validatiestudie* [The selection for functions of the cluster "Uitgebreid Algemeen"]. Technical report, Amsterdam, the Netherlands: University of Amsterdam, Department of Psychology.

*Roloff, H. P. (1928). Ueber eignung und bewährung [On aptitude and validity]. *Beihefte zur Zeitschrift für Angewande Psychologie, 148* (Abstract in L'Année Psychologique, 1928, 29, 897–899).

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 238–241.

*Ross, J. (1962). Predicting practical skill in engineering apprentices. *Occupational Psychologist, 36,* 69–74.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings:

Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75,* 322–327.

*Salgado, J. F. (1993). Validez sintética de pruebas de habilidades cognitivas por ordenador [Synthetic validity of the cognitive ability tests by computer]. *Revista de Psicología del Trabajo y las Organizaciones, 11,* 79–92.

Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and evidence. *Human Performance, 15,* 75–96.

*Salgado, J. F., & Blanco, M. J. (1990). Validez de las pruebas de aptitudes cognitivas en la selección de oficiales de mantenimiento en la Universidad de Santiago [Validity of cognitive ability tests to select maintenance workers in the University of Santiago]. *Libro de comunicaciones, II,* 124–130. (III Congreso de Psicología Social. Santiago de Compostela: Tórculo ediciones)

Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *International handbook of industrial, work and organizational psychology, Vol. 1* (pp. 165–199). London: Sage.

*Samuel, J. A. (1970). The effectiveness of aptitude tests in the selection of postmen. *Studies in Personnel Psychology, 2,* 65–73.

*Sánchez, J. (1969). Validación de tests con monitores de formación [Validation of tests with training instructors]. *Revista de Psicología General y Aplicada, 24,* 1015–1020.

*Schmidt-Atzert, L., & Deter, B. (1993). Die vorhersage des ausbildungserfolgs bei verscheidenen berufsgruppen durch leistungstets [The prediction of training success in different jobs by achievement tests]. *Zeitschrift für Arbeits und Organisationspsychologie, 37,* 191–196.

*Schmidt-Atzert, L., & Deter, B. (1993). Intelligenz un ausbildungserfolg: Eine untersuchung Zur prognostischen validität des I-S-T- 70 [Intelligence and training success: A study of the predictive validity of the I-S-T 70]. *Zeitschrift für Arbeits und Organisationspsychologie, 37,* 52–63.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance, 15,* 187–211.

Schmidt, F. L., Gast-Rosenberg, I., & Hunter, J. E. (1980). Validity generalization for computer programmers. *Journal of Applied Psychology, 65,* 643–661.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1,* 199–223.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., & Hunter, J. E. (1999). Theory testing and measurement error. *Intelligence, 27,* 183–198

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology, 66,* 166–185.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61,* 473–485.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627–670.

*Schuler, H., Moser, K., Diemand, A., & Funke, U. (1995). Validität eines eisntellungsinterviews zur prognose des ausbildungserfolgs [Validity of an employment interview for the prediction of training success]. *Zeitschrift für Pädagogische Psychologie, 9,* 45–54.

*Schuler, H., Moser, K., & Funke, U. (1994, July). *The moderating effect of rater-ratee acquaintance on the validity of an assessment center.* Paper presented at the 23th Congress of the International Association of Applied Psychology, Madrid, Spain.

*Serrano, P., García-Sevilla, L., Pérez, J. L., Pina, M., & Ruiz, J. R. (1986). Municipal police evaluation: Psychometric versus behavioral assessment. In J. C. Yuille (Ed.), *Police selection and training: The role of Psychology* (pp. 257–265). Dordrecht, the Netherlands: Martinus Nijhoff Publishers.

*SHL. (1989). *Validation review.* Surrey, United Kingdom: Saville, & Holdsworth.

*SHL. (1996). *Validation review II.* Surrey, United Kingdom: Saville & Holdsworth.

*Smith, M. C. (1976). A comparison of the trainability assessments and other tests for predicting the practical performance of dental students. *International Review of Applied Psychology, 25,* 125–130.

*Smith-Voskuyl, O. F. (1980). *De constructie van een testbatteryj voor de selectie van leerling-verplegenden aan de 'De Tjonsgerschans'* [The construction of a test battery for the selection of student nurses in De Tjonsgerschans]. Unpublished report, University of Amsterdam, Department of Psychology, Amsterdam, the Netherlands.

*Sneath, F., Thakor, M., & Medjuck, B. (1976). *Testing of people at work.* London: Institute of Personnel Management London.

*Spengler, G. (1971, July). *Die praxis der auswahl von führungskräften in der Glanzstoff AG* [The practice in executive selection in Glanzstoff A. G.]. Proceedings of the 17th Congress of the International Association of Applied Psychology, Brussels, Belgium.

*Spielman, W. (1923). Vocational tests for dressmakers' apprentices. *Journal of the National Institute of Industrial Psychology, 1,* 277–282.

*Spielman, W. (1924). The vocational selection of weavers. *Journal of the National Institute of Industrial Psychology, 2,* 256–261.

*Spielman, W. (1924). Vocational tests for selecting packers and pipers. *Journal of the National Institute of Industrial Psychology, 2,* 365–373.

*Starren, A. M. L. (1996). *De selectie van instroomkandidaten binnen het HR-beleid van een onderneming in de informatie Technologie. Een evaluatie onderzoek* [Applicant selection and the HR-policy of an IT company: An evaluation study]. Paper developed for the major Work and Organizational Psychology, University of Amsterdam, the Netherlands.

*Tagg, M. (1924). Vocational tests in the engineering trade. *Journal of the National Institute of Industrial Psychology, 2,* 313–323.

*te Nijenhuis, J. (1997). *Comparability of test scores for immigrants and majority group members in The Netherlands.* Unpublished doctoral dissertation, University of Amsterdam, the Netherlands

*Thurstone, L. L. (1976) *Manual del test PMA.* [Manual for the PMA test]. Madrid, Spain: Tea.

*Timpany, N. (1947). Assessment for foremanship. *British Journal of Psychology, 38,* 23–28.

*Trost, G., & Kirchenkamp, T. (1993). Predictive validity of cognitive and noncognitive variables with respect to choice of occupation and job success. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel selection and assessment* (pp. 303–314). Hillsdale, NJ: Erlbaum.

U.S. Department of Labor. (1991). *Dictionary of occupational title*s (4th ed.) Washingnton, DC: U.S. Government Printing Office.

*van Amstel, B.. & Roe, R. A. (1974). *Verslag van twee studies m. b. t. de part-time opleiding kinderberscherming B aan de Sociale Academie te Twente.* [Report on two studies of the part-time major "Kinderbescherming B" of the Social Academy of Twente]. Unpublished report, University of Amsterdam, Department of Psychology, the Netherlands.

*van der Maesen de Sombreff, P. E. A. M., & Zaal, J. N. (1986). *Incremental utility of personality questionnaires in selection by the Dutch government.* Unpublished manuscript.

*van Leest, P. F., & Olsder, H. E. (1991). *Psychologisch selectieonderzoek en praktijdeoordelingen avn agenten bij de gemeentepolitie Amsterdam.* Amsterdam: RPD Advies.

*Vernon, P. E. (1950). The validation of Civil Service selection board procedures. *Occupational Psychology, 24,* 75–95.

Vinchur, A. J., Schippman, J. S., Switzer F. A., III, & Roth, P. L. (1998).

A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83,* 586–597.

Vineberg, R., & Joyner, J. N. (1989). Evaluation of individual enlisted performance. In M. F. Wiskoff & G. M. Rampton (Eds.), *Military personnel measurement: Testing, assignment, evaluation* (pp. 169–200). New York: Praeger.

Viswesvaran, C., & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (GMA) in industrial, work, and organizational psychology. *Human Performance, 15,* 211–231.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81,* 557–574.

*Vogelaar, F. J. (1972). *Een validatie onderzoek van een selectieprocedure voor politiefunctionarissen.* [A validation study of the selection procedure for policemen]. Technical report, University of Amsterdam, Department of Psychology, the Netherlands.

*Vogue, F., & Darmon, E. G. (1966). Validation d'examens psychotechniques de conducteurs de chariots automoteurs de manutention [Validity of psychotechnic examinations of bus drivers]. *Bulletin du CERP, 15,* 183–187.

*Weiss, R. H. (1980). *Grundintelligenztest skala 3–CFT 3* [Group intelligence test scale 3–CFT 3]. Braunschweig, Germany: Georg Westermann.

*Wickham, M. (1949). Follow-up of personnel selection in the A. T. S. *Occupational Psychology, 23,* 153–168.

*Yela, M. (1956). Selección profesional de especialistas mecánicos [Professional selection of mechanic workers]. *Revista de Psicología General y Aplicada, 13,* 717–719.

*Yela, M. (1968). *Manual del test de Rotación de figuras macizas* [Manual for the complete figure rotated]. Madrid, Spain: Tea.

Received June 7, 2002
Revision received March 28, 2003
Accepted April 8, 2003 ■

Call for Nominations: *Rehabilitation Psychology*

The APA Publications and Communications (P&C) Board has opened nominations for the editorship of ***Rehabilitation Psychology*** for the years 2006–2011. Bruce Caplan, PhD, is the incumbent editor.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2005 to prepare for issues published in 2006. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Rehabilitation Psychology will transition from a division publication to an "all APA" journal in 2006, and the successful candidate will be involved in making suggestions to the P&C Board and APA Journals staff about the transition process.

Gary R. VandenBos, PhD, and Mark Appelbaum, PhD, have been appointed as cochairs for this search.

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to

Rehabilitation Psychology Search Committee
Karen Sellman, Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

The first review of nominations will begin December 8, 2003. The deadline for accepting nominations is **December 15, 2003**.

Copyright of Journal of Applied Psychology is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Journal of Applied Psychology is the property of American Psychological Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.