

Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis.

Citation for published version (APA):

Hülshager, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment*, 15(1), 3-18. <https://doi.org/10.1111/j.1468-2389.2007.00363.x>

Document status and date:

Published: 01/01/2007

DOI:

[10.1111/j.1468-2389.2007.00363.x](https://doi.org/10.1111/j.1468-2389.2007.00363.x)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Validity of General Mental Ability for the Prediction of Job Performance and Training Success in Germany: A meta-analysis¹

Ute R. Hülshager*, Günter W. Maier** and Thorsten Stupp**

*University of Amsterdam Business School, Amsterdam, The Netherlands. u.r.hulshager@uva.nl

**Department of Psychology, University of Bielefeld, Bielefeld, Germany

A meta-analysis into the operational validity of general mental ability (GMA) measures in Germany is presented. The meta-analysis addresses the question whether findings of US and European meta-analyses are generalizable to Germany given the differences in the education systems of these countries. The high level of differentiation in the German educational system is expected to enhance the homogeneity of applicant pools resulting in a low level of variability in predictor scores which reduces the observed GMA-performance relationships. Our analysis is based on 54 independent German articles and unpublished reports. Results indicated an operational validity of $\rho = .467$ for training success ($k = 90$; $N = 11,969$) and $\rho = .534$ for job performance ($k = 9$; $N = 746$). Moderator analyses showed that job complexity and the year of publication are relevant moderator variables, with lower job complexity levels and older studies being associated with higher operational validities. Findings suggest that overall German operational validities are comparable with findings in the United States or other European countries. However, for training success operational GMA validities are slightly lower in Germany compared with US or European meta-analyses.

1. Introduction

Evidence of the validity of general mental ability (GMA) tests for the prediction of vocational criteria has been extensively presented and discussed in the literature. GMA which is also referred to as general cognitive ability or general intelligence has been defined by a group of 52 experts as 'a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience' (Gottfredson, 1997, p. 13). Thus, GMA can be considered to be equivalent to the *g* factor which

underlies all specific cognitive abilities and can be assessed by any test battery that combines tests measuring different specific cognitive abilities (e.g., IST-70; Amthauer, 1973) or by tests especially designed to measure *g* or fluid intelligence (e.g., Raven's Progressive Matrices; Kratzmeier, 1979; see Salgado, Anderson, Moscoso, Bertua, De Fruyt, & Rolland, 2003b for a related definition).

Since Schmidt and Hunter (1977) countered the situational specificity hypothesis with the validity generalization hypothesis many researchers have conducted meta-analyses and estimated the average operational validity of GMA measures for the predic-

tion of job performance and training success (Hunter & Hunter, 1984; Levine, Spector, Menon, Narayanan, & Cannon-bowers, 1996; Pearlman, Schmidt, & Hunter, 1980; Schmitt, Gooding, Noe, & Kirsch, 1984). Referring to US primary studies, all these meta-analyses unanimously confirmed that validity generalization exists, i.e., the observed variability of validity estimates across studies is due to artifactual errors (e.g., small sample size, criterion unreliability and range restriction). After correction of these methodological artifacts little variation in validity estimates across personnel selection settings remains (see Schmidt, 2002; Schmidt & Hunter, 2004, for a detailed review).

However, the validity of GMA measures varies between jobs. Hunter and Hunter (1984) conducted an extensive and frequently cited meta-analysis on the predictive validity of GMA measures for training success and job performance. They examined the operational validity of GMA tests for different civilian occupations considering job complexity level as a moderator variable. The authors reported that the validity of GMA measures varies systematically with the complexity level of a given job: The more complex the job the higher the operational validity of GMA tests. For jobs of medium complexity they reported an average operational validity of $\rho = .51$ for job performance and $\rho = .57$ for training success.

In European countries, the encouraging results for the criterion-related validity of GMA measures found in US meta-analytic studies are frequently cited as references for the validity of GMA in European countries without reservation (Salgado & Anderson, 2002). Thus, German psychology students as well as psychologists working in organizations rely on German personnel selection textbooks (e.g., Kanning & Holling, 2002; Schuler & Funke, 1993) which indicate the validity of GMA measures for personnel selection purposes referring to US meta-analyses (Hunter & Hunter, 1984; Schmidt & Hunter, 1998; Schmitt *et al.*, 1984). Yet, US meta-analytic findings should not be transferred to European countries without explicitly testing whether they are in fact generalizable to non-US contexts, especially to countries which differ significantly from the United States with respect to their educational systems. Therefore, we conducted a country-level meta-analysis examining the validity of GMA measures for training success and job performance in Germany.

2. GMA validities in the European community

In 2002 Salgado and Anderson alluded to a lack of meta-analytical studies examining the criterion validity of GMA measures in European countries. As a conse-

quence, Salgado, Anderson, Moscoso, Bertua, and De Fruyt (2003a) conducted a meta-analysis across several European countries and found evidence for international validity generalization. For job performance, the mean corrected operational validity of GMA measures was .62 and .54 for training success. Yet, for training success, only 47% of the variance was accounted for by statistical artifacts, indicating that a large amount of the variability in observed validities remained even after sampling error, predictor and criterion reliability and range restriction had been corrected for. These findings suggest that moderators of the relationship between GMA measures and training success are likely to exist. As studies from 10 different European countries were included in the cross-country meta-analysis one such moderator might be the country which primary studies originated from. Therefore, Salgado and Anderson (2003) tested whether differences in the magnitude of validities between European countries existed (Salgado *et al.*, 2003a). Analyzing validities of GMA tests for six different European countries (France, Germany, Belgium, the Netherlands, Spain and the United Kingdom) evidence for differences in GMA validity magnitudes between European countries was presented: Whereas operational validities were .68 (job performance) and .63 (training success) in Germany, they were .56 (job performance) and .58 (training success) in the United Kingdom.

This extensive European-wide meta-analysis was the first to give an insight into the validity of GMA measures in Germany. However, several questions have yet to be addressed, making a comprehensive country-specific German meta-analysis necessary: First, only a small percentage of variance was accounted for by artifactual errors (job performance: 49%; training success: 51%) indicating the need to search for potential moderator variables. Second, the meta-analysis for the German subgroup was based on a comparatively small sample of nine German papers reporting validity coefficients for only eight independent samples for job performance and 26 independent samples for training success, resulting in a lack of power and the problem of second-order sampling errors. Therefore, it is necessary to establish a more extensive sample of German primary studies, including published and unpublished papers, in order to base the country-specific meta-analysis on a representative sample of German studies. Such a comprehensive sample of studies is an indispensable precondition for the realization of moderator analyses. Third, GMA validities can be expected to be lower in Germany than in other European countries or the United States because of peculiarities of the German educational system. Educational systems can be classified with respect to the extent of their stratification which is characterized by the degree of differentiation and tracking at secondary school level (Buchmann &

Dalton, 2002; Müller & Shavit, 1998). The German educational system is characterized by a strong institutional differentiation (Buchmann & Dalton, 2002). Students are sorted into different types of schools preparing them for different educational and occupational trajectories according to their level of achievement. Once they attend one type of school their opportunities to change to another type are limited. In contrast, school systems in the United States, Norway and Spain display a low level of stratification: Comprehensive secondary schools are prevalent, students are not differentiated by their level of achievement, ability-based tracking is uncommon and students can choose different educational and occupational trajectories throughout secondary school. Because of the extreme level of stratification in the German school system, qualifying students for a specific and limited set of jobs, more homogeneous applicant pools can be expected in Germany. This in turn may lead to a high level of range restriction in intelligence test scores and therefore to a reduced observed validity of GMA measures. For this reason we expect the observed validity of GMA for job performance and training success to be lower in Germany than in the United States or other European countries. The educational preconditions, that is the type of school attended and the educational level achieved, and consequently the educational preselection are much stronger for complex than for less complex jobs. Therefore, we specifically expect observed validities to be lower for jobs of high than for jobs of low complexity. In addition to cross-national meta-analyses (Salgado *et al.*, 2003a) country-specific meta-analyses complement our understanding of the role GMA plays in the prediction of job performance in different countries as indicated by a recent UK meta-analysis (Bertua, Anderson, & Salgado, 2005).

In sum, the goal of the present study is threefold: First, we want to determine the mean operational validity of GMA measures in Germany based on a comprehensive and representative database of German primary studies. Second, we want to examine the role of job complexity and the year of publication of primary studies as possible moderator variables in order to explain the variance that remained after the correction for methodological artifacts. Job complexity is of interest in order to test our hypothesis that, contrary to findings in the United States (Hunter & Hunter, 1984) or Europe (Salgado *et al.*, 2003b), there is a negative association between job complexity and GMA validity. The year of publication is considered because primary studies cover a time frame of seven decades. Changes in public life and work-related factors like employment policies, selection and assessment practices, the nature of jobs or type of applicants might affect the validity of GMA measures (Bertua *et al.*, 2005). Third, results of the present

German meta-analysis will be compared with findings of the United States or other European countries.

3. Method

3.1. Compilation of database

An extensive literature search was conducted in order to establish a database of German validity studies. Studies included in the database met the following criteria: (1) they were conducted using a German sample (Austrian or Swiss studies were not included), (2) samples were applicants, employees or trainees, (3) studies reported effect size indices (mainly correlation coefficients) of the association between job or training performance and GMA tests (concentration tests and professional aptitude tests were excluded), (4) student samples as well as studies reporting academic performance of university students were excluded.

To find relevant research, we adopted the following strategies: The databases PsycInfo (1872–2004), PSYINDEX (1977–2004) and PsyDoc, a database of German psychological online research literature and dissertations, were searched. Test manuals of German intelligence tests were examined and the central employment office as well as the human resources department of the German armed forces was contacted. Moreover, the reference list of a European meta-analysis including German studies (Salgado & Anderson, 2003) as well as citation lists of other relevant books and articles (e.g., Brambring, 1983; Funke, Krauss, Schuler, & Stapf, 1987; Sarges & Wottawa, 2001) were scrutinized to identify additional studies. In order to find unpublished studies, all DAX-100 companies (the 100 strongest companies listed on the German stock index) were contacted by mail and asked for any unpublished internal papers reporting validity coefficients of GMA tests. Looking through non-scientific journals written by and for practitioners we found various articles reporting validity studies conducted in German companies. The authors were contacted in order to get full information on these studies.

Our final database comprised 54 individual papers and books (including eight unpublished papers) reporting 90 independent samples with training success as the criterion ($N = 11,969$), and nine independent samples with overall job performance as the criterion ($N = 746$). With respect to training success, three studies were included in the database, which referred to military occupations reporting validities for six independent samples. In accordance with Hunter and Schmidt (1990, 2004), each sample contributed only one correlation to the meta-analysis. If studies reported conceptual replications (e.g., coefficients were indicated for different GMA measures or for various criteria),² composite correlations were computed by using the formulae provided by Hunter and Schmidt (2004).

Reliabilities of these composite measures were computed using Mosier's (1943) formula.

3.2. Coding of the studies

Two independent judges coded every study regarding the following information: validity coefficient, criterion type (training success vs job performance), sample size, predictor reliability, information on range restriction and skill level. Inter-rater agreement was assessed by intra-class-coefficients for validity coefficients (.993), sample size (.985), predictor reliability (1.0), range restriction (.981) and skill level (.910). For criterion type Cohen's κ was computed (.771). Consensus concerning all prior disagreements was reached by discussion.

In order to classify the jobs analyzed in the primary studies into homogeneous categories and assign them to job complexity levels, we used the European variant of the International Standard Classification of Occupations (ISCO-88 COM; Elias & Birch, 1994). The ISCO-88 (International Labour Office, 1990) as well as its European Adaptation (ISCO-88 COM) is a hierarchical framework which assigns jobs to occupational groups according to similarities in duties and tasks. The framework assigns major occupational groups to four different skill levels which are defined as '... the degree of complexity of constituent tasks and skill specialisation ...' (Elias & Birch, 1994, p. 1).

The assignment of skill levels to major occupational groups as well as examples of jobs belonging to these groups are depicted in the appendix. Because the classification system does not provide a skill level for managerial and military occupations, we assigned the occupations of the first major occupational group (legislators, senior officials and managers) to the highest skill level following the classification of Salgado *et al.* (2003b). Military occupations were assigned to skill levels according to their professional education and the skill specialization required by their job. None of the occupations analyzed in the primary studies fell into the lowest skill level group (cf. appendix A).

For two reasons we preferred the ISCO-88 COM to the DOT (US Department of Labor, 1991), which is commonly used for similar purposes in US meta-analytic studies (e.g., Schmidt, Hunter, & Pearlman, 1981). First, we wanted to use an approved, standardized classification system which represents the characteristics of the German occupational system and which allows international comparisons. The ISCO-88 COM is a European Union Standard and therefore represents German occupational titles better than the DOT. Second, by providing skill levels it allows occupational groups to be assigned directly to different job complexity levels. In order to allow comparisons between studies which used the DOT and our study we referred to the list of representative jobs of

occupational categories and the assigned job complexity levels described by Salgado *et al.* (2003b). We classified 46 of the 49 listed jobs according to the ISCO-88 COM occupational groups and skill levels. The correlation between the three job complexity levels assigned by Salgado *et al.* (2003b) and the four skill levels assigned according to the ISCO-88 COM was .80 ($p < .001$) which underlines the comparability of the European and the US classification system. Therefore, the terms job complexity and skill level will be used interchangeably.

3.3. Procedure

To integrate the results of the primary studies, the meta-analytic approach of Raju, Burke, Normand, and Langlois (1991) and the corresponding computer program (Raju & Fleer, 2003) was applied. When only partial sample-based information on artifacts (predictor reliability, criterion reliability and range restriction) is available, the program uses the weighted average of all available range restriction or reliability values in a given set of studies as an estimate for missing coefficients. If no sample-based information on artifacts is available at all, it is possible to use the mean of a hypothetical distribution of other meta-analyses as a default value. This procedure does not rely on the assumption that population correlation, range restriction, predictor and criterion reliability are uncorrelated across populations. Thus, we used this meta-analytic approach to calculate how much of the variance of validities between studies is accounted for by artifacts. For this purpose, sampling error, unreliability of predictor and criterion and range restriction were considered. In addition, sampling error, unreliability of the criterion and range restriction were corrected in order to estimate the operational validity of GMA measures. When the operational validity is estimated, no corrections for predictor reliability are made because the operational validity indicates the estimated true validity of a measure for applied purposes. As test reliabilities are never perfect, this operational value is of greater interest than the theoretical validity under the assumption of perfect test reliability (Kuncel, Hezlett, & Ones, 2004; Salgado *et al.*, 2003b). Therefore, in order to calculate the mean operational validity, variance of ρ , 90% credibility values and 95% confidence interval corrections for the reliability of the criterion and range restriction are made. Only for the estimation of the percentage of variance accounted for by artifacts are corrections for the reliability of predictor scores taken into account additionally. Following the recommendation of Hunter and Schmidt (2000), results are based on a random effects model which yields conservative and more reliable estimates than the fixed effects model.

3.4. Artifact corrections

3.4.1. Predictor reliability

As the FORTRAN program of Raju and Fler (2003) allows the correction formulas to be conducted even when only partial sample-based artifact data are available, sample-based information about predictor reliability and range restriction were considered if reported in the studies. Hunter and Schmidt (2004) recommend the reliability coefficient of equivalence and stability (CES) for corrections of measurement error as it covers measurement error most comprehensively. If CES estimates (which are assessed by the correlation between parallel forms administered on different occasions) are not available, coefficients of stability (CS; test-retest reliability) or coefficients of equivalence (CE; Cronbach's α) can be considered instead. However using the latter two types of reliability, full correction for measurement error is not possible.³ As CES estimates were not reported in primary studies, test-retest reliabilities (CS) or internal consistency reliabilities (CE) were coded if reported in the studies. Test-retest reliabilities were preferred to internal consistency reliabilities as recommended by Salgado *et al.* (2003a; see also Schmidt & Hunter, 1999). In the cases where primary studies did not indicate the predictor reliability, the coefficients were looked up in the corresponding test manual. In total, 74 predictor reliability coefficients were available for studies examining training success and six coefficients for job performance. If no information about predictor reliability was accessible on the basis of individual studies or test manuals, the program used the weighted average of the available predictor reliabilities. For the meta-analyses of training success and job performance, the weighted mean reliabilities were .906 for training success and .922 for job performance (see Table 1). These weighted mean reliabilities were taken to replace missing values occurring in subgroups when the moderator's job complexity and the year of publication were analyzed.

3.4.2. Range restriction

Direct range restriction can be assessed by using information on selection ratios reported in primary

studies. The selection ratio describes the proportion of applicants selected by the test. Using Schmidt and Hunter's (2004, p. 111–112) formulae, we derived estimates of the standard deviations among the selected applicants (u) from the selection ratios obtained from primary studies. Table 1 indicates that for training success 37 range restriction values with a sample size weighted mean of .687 were available. This figure is comparable with the range restriction ratios indicated in other meta-analyses on the validity of GMA measures (.67, Salgado *et al.*, 2003a). For job performance, six range restriction values with a weighted mean of .846 were indicated in primary studies. This figure is higher than range restriction ratios utilized in large-scale US (.67, Hunter & Hunter, 1984) or European meta-analyses (.62, Salgado *et al.*, 2003a). As the value of .846 is based on only six studies, this estimate should not be interpreted as being representative of German job performance studies because it could be biased by second-order sampling error. However, it should be considered that with the meta-analytic approach of Raju *et al.* (1991) the validity of every primary study is corrected by its own range restriction value and that out of nine primary studies range restriction values were indicated for six primary studies. Therefore, only three studies were corrected by the weighted mean range restriction value of .846 which could be upwardly biased. Again, weighted mean range restriction values were taken to replace missing range restriction values in all subgroups when the moderator's job complexity and the year of publication were analyzed. Table 2 indicates the range restriction values and the weighted mean per skill level group for training success. No range restriction value was accessible in the fourth skill level group. Therefore, the sample size weighted mean range restriction value of studies falling into the next skill level (skill level 3; .670) was taken as the default value.

3.4.3. Criterion reliability

In the case of criterion reliability nearly no information was provided within single studies. However, Schuler, Funke, and Baron-Boldt (1990) conducted a meta-analysis examining the predictive validity of school

Table 1. Information on predictor reliability and range restriction for training success and job performance

Artifact	Criterion	Number of available coefficients	$M_{r_{xx}}$	$S^2_{r_{xx}}$	M_u	S^2_u	Max	Min
Predictor reliability	Training success	74	.906	.003	–	–	.972	.695
	Job performance	6	.922	.001	–	–	.962	.880
Range restriction	Training success	37	–	–	.687	.046	1.000	.411
	Job performance	6	–	–	.846	.064	1.000	.412

Note: $M_{r_{xx}}$, weighted mean of reliability coefficients; $S^2_{r_{xx}}$, weighted variance of reliability coefficients; M_u , weighted mean of range restriction indices; S^2_u , weighted variance of range restriction indices.

Table 2. Training success – range restriction indices for GMA measures for different skill levels

Skill level 4	Skill level 3		Skill level 2	
	<i>u</i>	<i>F</i>	<i>u</i>	<i>F</i>
No information on range restriction was available:	.415	1	.411	2
.670 was set as the default value	.435	1	.447	2
	.452	1	.455	2
	.466	1	.482	1
	.471	2	.536	1
	.511	1	.659	1
	.517	1	.726	1
	.532	2	.733	1
	.534	1	1.000	5
	.550	1		
	.570	1		
	.760	1		
	1.000	4		
	.687 ^a	29	.687 ^a	17
	M_u : .670		M_u : .682	
	S_u^2 : .017		S_u^2 : .020	

Note: ^aIf no information on range restriction was available in primary studies within this subgroup .687 was set as the default value; *u*, range restriction ratio; M_u , weighted mean; S_u^2 , weighted variance; *F*, frequency.

grades for vocational training success based on German primary studies. Within this meta-analysis they established a hypothetical artifact distribution of training success reliabilities. The mean (.801) of their hypothetical distribution of training success reliabilities was set as the default value for criterion reliability in our meta-analysis. In accordance with the meta-analysis presented by Schuler *et al.* (1990) the vast majority of primary studies in our meta-analysis (i.e., 88.8%) assessed training success by (standardized) examination grades (e.g., final grades of the German Chamber of Industry and Commerce). As it is important to use country-specific correction values in a country-specific meta-analysis, the estimate of .801 is considered an accurate estimate for the reliability of training success measures in Germany because it was retrieved from a meta-analysis examining the same kind of training success measures in Germany. Note that this estimate equals the value of .80 used by Hunter and Hunter (1984) and Bertua *et al.* (2005).

In all of the primary studies measuring job performance the criterion was measured by supervisor ratings. As no meta-analysis was found that was based on German studies and that presented a distribution of criterion reliabilities for job performance ratings, we relied on the mean interrater reliability of .52 which was recommended by Rothstein (1990) and confirmed by a large-scale meta-analysis on the reliability of supervisor ratings (Viswesvaran, Ones, & Schmidt, 1996). Furthermore, this was the value used in a recent European meta-analysis (Salgado & Anderson, 2003).

3.5. Fail-safe *N* analysis

A concern closely linked to the method of meta-analysis is the so-called file-drawer problem (Rosenthal, 1979), describing a systematic bias of the studies included in meta-analyses. As studies reporting significant results tend to be published more often than studies reporting insignificant findings, the latter may be underrepresented in meta-analytic investigations. Therefore, Ashworth, Osburn, Callender, and Boyle (1992) suggested calculating the effect of unrepresented or missing studies with a mean validity and standard deviation of zero while fixing the proportion of unrepresented studies at 10% of the studies included in a given meta-analysis. Hence, we additionally calculated the 90% lower credibility value, assuming that in addition to the studies already included in our meta-analysis, there would be another 10% unrepresented studies with a validity of zero. This estimate is indicated in the tables as LCV, lowest hypothetical lower credibility value.

4. Results

4.1. Operational validities of GMA measures

Results of the meta-analysis examining the validities of GMA measures for training success and job performance are shown in Table 3. As column one shows, many more studies have been conducted reporting validity coefficients for training ($k = 90$) than for job performance criteria ($k = 9$). As a consequence, the overall sample sizes ranged between 746 and 11,969. Table 3 shows an operational validity for training success of .467.⁴ The fact that the 90% credibility value (.272) was clearly greater than zero indicates that GMA tests are valid predictors of training success, and that validity generalization across different German organizations exists. However, the remaining variance of the operational validity ($S_\rho^2 = .023$), as well as the relatively small percentage of variance accounted for by artifacts (34.3%), indicates that after the correction for artifacts variations in the correlations still existed. Hence, the actual magnitude may vary across different settings and a moderator analysis seems justified.

The meta-analysis for job performance is based on a very small sample of nine independent studies and a total sample size of 746. Consequently, it should be considered that the results may be influenced by second-order sampling error. As can be seen in Table 3, the operational validity of GMA measures for job performance amounted to .534. The 90% credibility value of .296 was substantial indicating validity generalization of GMA measures across samples. Again, the variance of the operational validity ($S_\rho^2 = .034$) and the small percentage (34%) of variance accounted for by

Table 3. Meta-analysis of GMA measures for training success and job performance

Criteria	k	N	Mean r	S_r^2	ρ	S_ρ^2	% VE	90% CV	95% CI	LCV
Training success	90	11,969	.312	.022	.467	.023	34.3	.272–.661	.428–.505	.172
Job performance	9	746	.333	.020	.534	.034	34.0	.296–.770	.386–.682	.187

Note: k, number of studies; N, total number of participants; Mean r, weighted average of observed validity; S_r^2 , observed variance of validity coefficients; ρ , mean operational validity (corrected for range restriction and criterion reliability); S_ρ^2 , estimate of the variance of ρ ; % VE, variance accounted for by artifacts; 90% CV, 90% lower and upper credibility value; 95% CI, 95% confidence interval; LCV, lowest hypothetical 90% lower credibility value.

Table 4. Meta-analysis of GMA measures for training success by different skill level groups

Skill level	Job complexity	k	N	Mean r	S_r^2	ρ	S_ρ^2	% VE	90% CV	95% CI	LCV
2	Low	45	4,931	.351	.021	.520	.015	45.4	.362–.677	.471–.569	.230
3	Medium	35	5,510	.293	.019	.452	.025	32.9	.250–.654	.388–.516	.156
4	High	6	1,089	.187	.009	.299	.008	57.5	.181–.413	.186–.412	.117

Note: k, number of studies; N, total number of participants; Mean r, weighted average of observed validity; S_r^2 , observed variance of validity coefficients; ρ , mean operational validity (corrected for range restriction and criterion reliability); S_ρ^2 , estimate of the variance of ρ ; % VE, variance accounted for by artifacts; 90% CV, 90% lower and upper credibility value; 95% CI, 95% confidence interval; LCV, lowest hypothetical 90% lower credibility value.

artifacts show that moderators of the overall relationship probably exist.

4.2. Analysis of moderator effects – validity for different skill level groups (job complexity)

Results of the meta-analyses for different GMA-skill level combinations for the prediction of training success are shown in Table 4. As only six studies fell into the highest job complexity group, related findings should be interpreted with caution.

The largest operational validity (.520) was found for the second skill level group. This group was characterized by low job complexity and comprised occupations like office clerks, machinery mechanics, electrical mechanics and police officers. Only 45.4% of the variance was accounted for by artifacts. However, the 90% CV was .362, indicating a substantial validity for this group. For the third skill level group with medium job complexity (e.g., chemical science technicians, administrative associate professionals, police inspectors), the validity was lower (.452) but reached a satisfactory 90% CV of .250. Again, the amount of variance accounted for by artifacts was rather small (32.9%) indicating that further moderators seemed to play a major role. The lowest validity was found for the highest job complexity (skill level 4; e.g., chief executives, science professionals, commissioned officers) with an operational validity of .299 and a 90% CV of .181. Artifacts accounted for approximately half (57.5%) of the variance of observed validities. The results suggest that the skill level of different occupations has a moderating effect on the validity of GMA measures for training success: The validity was lower for higher skill level than for lower skill level occupations.

4.3. Analysis of moderator effects – validity for recent and older studies

In addition to job complexity, the year of publication was considered as a potential moderator. Therefore, the year of publication and skill level were included simultaneously in a multiple regression analysis assessing the specific contribution of each moderator. The year of publication and skill level were entered in the regression as continuous variables with skill level ranging from 2 (low) to 4 (high). Taken together the two moderator variables accounted for 41% of the variance of operational validities ($R^2 = .411$; $p < .01$) for training success. Both the year of publication ($\beta = -.534$; $p < .001$) and skill level ($\beta = -.262$, $p < .01$) were negatively related to operational validities. In the case of job performance, the year of publication and skill level accounted for 72% of the differences in operational validities ($R^2 = .718$, NS). The year of publication was significantly related to operational validities ($\beta = -.901$; $p < .05$), whereas skill level was not ($\beta = .186$; NS). Owing to the very small sample size ($k = 7$), results for job performance should be interpreted with great caution.

In order to visualize the moderator effect of year of publication, subgroup analyses were conducted. This involved splitting the sample into two parts, leaving a comparable number of studies (k) as well as total sample size (N) in each subgroup. Accordingly, the year 1990 was chosen as a cut-off point, because it left a sufficient number of studies in each subgroup for training success as well as for job performance. The results presented in Table 5 show that the validities of GMA measures for the prediction of training success diverged between 'old' and 'more recent' validation studies, with validation studies carried out before 1990 providing higher validity

Table 5. Meta-analysis of GMA measures for training success and job performance by year of publication

Year of publication	<i>k</i>	<i>N</i>	Mean <i>r</i>	S_r^2	ρ	S_ρ^2	% VE	90% CV	95% CI	LCV
Training success										
< 1990	58	5,902	.387	.023	.555	.019	43.3	.378–.731	.509–.601	.240
≥ 1990	32	6,067	.239	.010	.381	.012	44.7	.241–.521	.328–.433	.153
Job performance										
< 1990	5	360	.424	.022	.739	.004	85.7	.660–.820	.615–.863	.389
≥ 1990	4	386	.247	.002	.342	.000	100	.342–.342	.275–.410	.185

Note: *k*, number of studies; *N*, total number of participants; Mean *r*, weighted average of observed validity; ρ , mean operational validity (corrected for range restriction and criterion reliability); S_r^2 , observed variance of validity coefficients; S_ρ^2 , estimate of the variance of ρ ; % VE, variance accounted for by artifacts; 90% CV, 90% lower and upper credibility value; 95% CI, 95% confidence interval; LCV, lowest hypothetical 90% lower credibility value.

estimates (.555) than studies carried out in or after 1990 (.381). As the 95% confidence intervals did not overlap it can be concluded that the year of publication was a valid moderator variable.

For job performance, the results were similar: The five studies carried out before 1990 had a mean operational validity of .739, whereas later studies had a mean operational validity of .342. The 95% confidence intervals did not overlap and in both subgroups artifacts accounted for nearly 100% of variance. This underlines that the year of publication was a strong moderator for validity estimates of job performance ratings. However, as these analyses were based on a very small sample of studies, any conclusions drawn should be regarded as tentative.

5. Discussion

Many US meta-analytic studies have provided evidence that intelligence is the best stand alone predictor of job and training performance (Schmidt, 2002). They have shown that validity generalization exists across different test types, facets of intelligence, different occupations and organizations. Only recently was the adaptability of these results, based on studies conducted in the US, to the European Community tested explicitly (Salgado et al., 2003a). Salgado and colleagues showed that validity generalization exists in Europe, but the criterion validities that they found across European countries were slightly higher than those in US meta-analytic studies. Furthermore, variations in the magnitude of validities were found between European countries and the need to search for possible moderator variables in the German subsample was expressed (Salgado & Anderson, 2003). Therefore, the aim of the present study was to determine the mean operational validity of GMA measures for training success and job performance criteria in Germany and to compare them to findings from the United States or other European countries. In addition, job complexity and the year of publication were considered as possible moderator variables.

5.1. Mean operational validity of GMA measures in Germany

Concerning our first aim, findings suggest that GMA measures are valid predictors of training success and job performance ratings in Germany. With operational validities of .467 for training success and .534 for job performance and the respective 90% credibility values lying clearly above zero, the conclusion that validity generalization exists across different samples and situations in Germany is justified. The robustness of our findings is underlined by the fact that, even when additional unrepresented studies with a validity of zero are included in our analyses, the 90% credibility values are still greater than zero for training success, job performance and all subgroup analyses (see LCV). Yet, training success validities reported in our study are slightly lower than those typically found in the United States whereas job performance validities are not. In their comprehensive overview of US meta-analytic studies Schmidt and Hunter (1998) indicated a validity of .56 for training success and .51 for job performance as being representative of other meta-analytic findings for general cognitive abilities. Moreover, our validities are lower than those reported in a European meta-analysis, which examined the mean validity for Europe as a whole. For GMA tests, Salgado et al. (2003a) indicated operational validities of .54 for training success and .62 for job performance in a European sample of studies including studies from Spain, Portugal, Scandinavian countries, the United Kingdom, France, the Netherlands, Belgium and Germany.

In sum, German mean operational validities are comparable to those found in large-scale US or European meta-analyses. Only with respect to training success did the German operational validity differ slightly from values reported in US or European meta-analyses. Hence, although the relationship between GMA and performance is likely to be the same across countries on the construct level (assuming that predictor and criterion measures were perfectly reliable and that we had a representative sample from the general popula-

tion), operational validities might differ slightly due to country-specific factors. The identity and nature of these country-specific factors needs to be examined.

5.2. Why are operational validities in Germany slightly lower than validities found in the United States and other European countries?

5.2.1. Specifics of the German educational system

Specifics of the German educational system might explain why GMA validities for training success are slightly lower in Germany compared with other countries. The major difference between the German and other educational systems concerns the level of stratification, hence the degree of structuring and differentiation (Buchmann & Dalton, 2002; Müller & Shavit, 1998). Whereas the US educational system is characterized by a weak level of stratification where students are not sorted into different kinds of schools but rather attend comprehensive secondary schools (Buchmann & Dalton, 2002), students are assigned to one of three types of secondary schools according to their level of academic achievement in Germany after only four years of education: *Hauptschule*, comprising basic secondary education, *Realschule*, offering a more demanding curriculum and *Gymnasium*, offering direct university entrance qualifications and preparing students for academic careers (Baumert, Cortina, & Leschinsky, 2003; Postlethwaite, 1988). Theoretically, it is possible to change from the other two school types to the *Gymnasium*, but only a minority of students manage to do so (Baumert *et al.*, 2003). Only recently (02/06) the Special Rapporteur on the Right to Education Vernor Muñoz Villalobos (Muñoz Villalobos, 2006) criticized the German educational system as being separative rather than integrative and called the classification of children at the age of 10 'internacionalmente anormal' (internationally abnormal, p. 4). The type of school that pupils attend influences the career they pursue afterwards and the professional opportunities open to them because students from a certain kind of school are qualified for a specific and limited set of jobs. In fact, students from different types of schools do not only differ with respect to the qualifications they acquire but also with regard to their actual level of cognitive ability (Horn, 1983). Consequently, students will not only be selected for specific jobs but they will also be attracted to certain jobs according to their educational status. This results in more homogeneous applicant groups in Germany than in the United States. In 1966, Wesley and Murch juxtaposed the United States and the German educational systems and pointed out that in the US 50% of a student age-group obtained the high school diploma compared with only 6.6% in Germany. As a consequence, far fewer German than US academic students attained a university degree. In 2002, it remained the case that fewer German (50%) than US students (64%) or students from the

United Kingdom (74%) entered tertiary education (OECD, 2004). Therefore, taking into account that intelligence measures and educational achievement are linked (Gustafsson & Undheim, 1996; Neisser, Boodoo, Bouchard, Boykin, Brody, & Ceci, 1996), the vertical structuring of the German educational system may lead to a restricted variability of predictor scores in job-specific applicant groups. This, in turn, reduces the association between GMA tests and training success in Germany compared with other countries.

Our assumptions about the early differentiation of the German school system based on cognitive ability are corroborated by empirical findings on mean differences in GMA test scores between students attending different kinds of schools (Amthauer, Brocke, Liepmann, & Beauducel, 2001; Horn, 1983; Kratzmeier, 1979; Schmidt-Atzert, Deter, & Jaekel, 2004; Tent, 1969). Mean GMA test scores measured with the LPS (Horn, 1983) were considerably higher in samples of pupils attending the *Gymnasium* (227.3) or *Realschule* (196.2) than in a sample of students attending the *Hauptschule* (161.4; Tent, 1969). In order to assess the effect of the German educational system on the variability of GMA test scores in job-specific applicant groups directly, GMA variances in applicant samples should be compared with general population norms. Yet, despite a thorough literature review, we could not retrieve GMA variances of German applicant samples from any primary study. Thus, it seems necessary for future research to compare the means and variances of GMA test scores in German applicant samples with general population norms as well as to applicant samples from the United States or other European countries in order to test our proposition that the stratification in the German school system leads to variance restriction among applicant samples.

5.2.2. Influence of indirect range restriction on operational validities

The stratification of the German school system leads to indirect range restriction. This has to be distinguished from direct range restriction which is commonly corrected in meta-analyses. Direct range restriction describes the problem that in personnel selection research the correlation between predictor and criterion is not based on the full applicant population, but only on those applicants who have been selected because of their scores in the predictor variable. Thus, results of primary studies are usually based on incumbent populations which display a lower variance in predictor scores than applicant populations because those applicants with very low predictor scores have not been hired (Schmidt & Hunter, 2004). In contrast, the specifics of the German school system do not lead to direct but to indirect range restriction because the selection is not based on the predictor itself but on other predictor-related characteristics. The German

educational system which differentiates students and prepares them for specific occupational trajectories leads to constricted applicant groups for certain jobs because students from different kinds of schools are qualified for different kinds of jobs. Hence, certain jobs tend to attract students from a certain kind of school. By considering the selection ratio in a given study, indirect range restriction is not accounted for because only a restricted sample of students applied for a specific job. In contrast, concerning cognitive ability, job-specific applicant pools were not considerably less variable than broad workforce norms in the United States (Sackett & Ostgaard, 1994). This could be attributed to the relatively undifferentiated secondary schooling in the United States (Buchmann & Dalton, 2002). As indirect range restriction has a repressing effect on operational validities, correcting for indirect range restriction would certainly raise operational validities to the level found in US or European meta-analyses. In sum, it can be assumed that the actual validity of GMA measures for the prediction of training success might be as high in Germany as it is in other countries, if the actual magnitude of operational validities was not decreased by an indirect effect of range restriction due to the German educational system.

The effect of range restriction in predictor scores caused by the differentiation of the German school system might be more pronounced in the case of training success than in the case of job performance. This might be the case because the type of school that pupils attend has a stronger effect on the vocational training they pursue following school than on the job they hold years later. While the pathway for German apprentices to participate in a specific training program is predefined and strongly depends on the type of school they attended, there are several ways individuals with different educational backgrounds can end up in the same job and consequently in the same job incumbent sample. Therefore, the applicant pool may be more homogeneous if training success is examined instead of job performance.

Besides differences in school systems, cultural differences between the United States and Europe as well as between European countries could be considered as potential causes of validity differences of personnel selection measures, as Salgado and Anderson speculated (2003, cf. Bertua *et al.*, 2005 for related arguments): Differences in work-related values (Hofstede, 1980), different leadership prototypes (Brodbeck, Frese, Akerblom, Audia, Bakacsi, & Bandova, 2000), as well as different staffing practices (Ryan, McFarland, Baron, & Page, 1999) could possibly exert an impact on validities of any personnel selection measure. However, we do not yet know if and how these factors have an effect on GMA validities. A fruitful pathway for future research would be to investigate directly whether cultural dimensions influence validities of personnel selection measures. This

would help us to clarify how differences in validity magnitudes between countries could be explained.

5.2.3. Differences in performance criteria and reliability estimates

Apart from educational differences between Germany, the United States and other European countries, disparities in the kinds of criteria used in primary studies and related reliability estimates have to be considered as a potential cause of differences in operational validities. In $k = 87$ independent studies (88.8%) standardized examination grades (e.g., final grades of the German Chamber of Industry and Commerce) were used as the criterion measure for training success whereas only 11 studies assessed training success by supervisor ratings. In contrast, in the European meta-analysis ratings by trainers or supervisors were used in the majority of primary studies (Salgado *et al.*, 2003a). Consequently, the correction value for criterion reliability of training success measures was considerably higher in our meta-analysis (.801) than in the European meta-analysis (.56) and higher than the coefficient reported by Viswesvaran *et al.* (1996) in their meta-analysis of the interrater reliability of supervisor ratings (.52). However, it was equal to the value used in meta-analyses by Bertua *et al.* (2005) and Hunter and Hunter (1984). As the magnitude of the correction value for criterion reliability affects the magnitude of operational validities (with lower criterion reliability estimates leading to higher operational validities than lower estimates for criterion reliability) these differences have to be considered when interpreting minor differences in operational validities between the present findings and findings from other US or European meta-analyses.

5.3. Moderating effect of job complexity

For training success, the complexity of occupations was analyzed as a possible moderating variable. In all three skill level groups, the operational validities and respective 90% credibility values were substantial, confirming the validity of GMA measures for different skill levels. Yet, operational validities differed substantially between the three different skill levels, suggesting a strong moderating effect: The lower the skill level the higher the operational validity.

The finding that higher skill levels are associated with lower operational validities does not correspond to previous meta-analyses which indicate a positive association between job complexity and the magnitude of operational validities (Hunter & Hunter, 1984; Salgado *et al.*, 2003b). However, a negative association between job complexity and GMA validity was expected for our German sample of primary studies because of the aforementioned level of stratification of the German educational system. The group of people receiving

higher education and thereby the qualification for highly complex jobs is especially a strongly preselected group. The German school system has been criticized for its imperviousness: In the 1960s only 10% of German students changed the type of school they attended, and even in the year 2000 only 16.3% of 15-year old students changed school, whereof only 30% changed to a higher and 70% to a lower level type of school (Baumert et al., 2003). The system can be criticized as being permeable top down but not bottom up. For students having attended schools giving primarily basic education (*Hauptschule*, *Realschule*), it is very difficult to change to the *Gymnasium* and to gain the high school diploma. This, however, is a precondition for applying for certain jobs and for entering university. As a consequence, the variance of intelligence values may be restricted particularly in the group of students attending the *Gymnasium*. This was shown in a study reported in the Manual of the German Adaptation (Kratzmeier, 1979) of Raven's Standard Progressive Matrices: the group of students attending the *Gymnasium* differed from the group of students attending the *Hauptschule* not only with respect to the mean value of GMA but also with respect to the standard deviation. The standard deviation was lower in the *Gymnasium* ($SD = 5.63$) than in the *Hauptschule* ($SD = 8.16$). As the *Gymnasium* is the type of school preparing students for higher skill level jobs, particularly for jobs which require a high school diploma or a university degree, the range restriction in the predictor is emphasized and causes a reduction in the predictor-criterion correlation. Simply put, having obtained the university-entrance diploma at the *Gymnasium* one can apply for the job of a machine operator or attend university to become a mechanical engineer, but without having attended the *Gymnasium* it is almost impossible to become a mechanical engineer. Hence, range restriction is stronger and validities are lower for jobs of higher than for jobs of lower job complexity. Because this kind of range restriction is not reflected in the selection ratios considered for the corrections of direct range restriction, it could be argued that predictor-criterion correlations might have been undercorrected, especially for high skill level occupations.

5.4. Moderating effect of year of publication

As the studies included in the database cover a wide time frame (1938–2004), the year of publication was analyzed as a possible moderator variable. Findings of the regression analysis as well as the subgroup analyses show that the year of publication moderates the magnitude of GMA validities for training success and job performance with older studies displaying higher validities than more recent studies. The percentage of variance accounted for by artifacts is higher and the remaining variance of operational validities is lower in

the subgroups than in the total group indicating that the year of publication is a valid moderator variable.

The results are in accordance with findings of a meta-analysis about the predictive validities of school grades for academic and vocational training success in Germany (Schuler et al., 1990) where the year of publication was found to have a strong moderating effect. Like in the present study, older studies reported higher validities than more recent studies. In contrast, in their examination of the General Aptitude Test Battery (GATB), Hartigan and Wigdor (1989) did not find evidence for a decline in validity over time. Although prior meta-analyses about the validity of cognitive abilities included studies with a wide range of publication years (Hunter & Hunter, 1984; Levine et al., 1996; Salgado et al., 2003b), the moderating influence of the year of publication has not been analyzed in these studies. Yet, Bertua et al. (2005) concede that possible changes in work-related factors like the nature of jobs or the type of applicants might have an effect on test validities. It would be beneficial if future studies examined the year of publication as a possible moderator variable in order to analyze whether this moderator is general or specific to Germany. If the finding can be replicated in other countries, it may be questioned whether it is appropriate to compute mean operational validities on the basis of a wide time range of studies. Especially if one is interested in an estimate of the current mean true validity, it should be discussed whether older studies should be included in meta-analyses.

The literature search associated with the present meta-analysis disclosed that relatively few validation studies have been conducted in Germany in the last years, especially for job performance. This is in accordance with the observation that intelligence tests are not applied as frequently in Germany as in other countries like the United Kingdom or Spain (Ryan et al., 1999). Hence, more studies are needed examining the current validity of GMA measures in Germany. This is important in order to test the trend of decreasing GMA validities found in our study.

5.5. Limitations and strengths

The results of the present study should be interpreted with respect to its limitations and strengths. First, in comparison with large-scale US meta-analytic studies, our study is limited by its comparatively small number of 54 articles and unpublished reports, which introduces second-order sampling error. As a consequence, we had to limit the moderator analysis of skill level to training success because for job performance only two to three studies fell into each skill level group. Second, only eight unpublished reports could be included in the database despite our efforts to detect unpublished data by contacting the DAX-100 companies and examining journals directed at practitioners. Third, another pro-

blem the present study had to overcome concerns the availability of artifact information: There were no studies providing information on the reliability of criteria. Yet, by choosing the mean of artifact distributions of a German or European meta-analysis, recently published empirical correction values for the reliability of training success and job performance could be chosen. A final point concerns the comparatively small amount of variance accounted for by artifacts which points to potential moderator effects. The moderators investigated in the present study are not the only variables which might moderate the GMA–performance relationship because even in the subgroups the amount of variance explained by artifacts did not exceed 50% in most of the cases. Future moderator analyses might focus on differences in specific GMA measures (see Hülshager, Maier, Stumpp, & Muck, 2006), specific cognitive abilities or the distinction between concurrent and predictive validation studies.

5.6. Conclusion and implications

The present findings of a German meta-analysis underline that GMA validities are in fact generalizable internationally although small differences in operational validities might exist between countries. The actual level of validity in a specific country is important information, not only from a theoretical but also from an applied point of view: For personnel selection practitioners who have to decide which selection method to use, the validity of GMA measures in their country is valuable information. Moreover, for globally active organizations, which assign workers overseas, knowledge about even slight differences between countries might be of interest.

The present research shows that differences between countries may be reflected in different moderating influences. Interestingly, the established moderator effect of job complexity pointed in a different direction than in US and European meta-analyses. Furthermore, it could be shown that the year of publication is an important moderator variable which should be considered in future studies. Hence, stimulated by the work of Salgado, Anderson, and colleagues, our study contributes to cultural comparisons in personnel selection research and thereby helps in understanding the commonalities and differences between European countries.

Acknowledgements

We are indebted to Nambury Raju^{*} for his helpful statistical advice concerning the method of meta-analysis as well as to Peter Muck, Rainer Hülshager and Chris Collier for their careful review of the manuscript. Additionally, we would like to thank Martin

Kersting for providing us access to studies of the Deutsche Gesellschaft für Personalwesen e.V. (DGP).

Notes

1. Parts of this study were presented at the 44th meeting of the German Society of Psychology, Göttingen, Germany, September 26–30, 2004.
2. In the case of training success, criteria were grades from vocational training or supervisor ratings, in the case of job performance, studies only reported supervisor ratings. Other criteria like objective performance measures were not reported in any primary studies included in the database.
3. Yet this underestimation of measurement error in the predictor will not affect estimates of the operational validity because predictor reliability was only considered to estimate the percentage of variance accounted for by artifactual errors.
4. Six of the studies reporting validity coefficients for training success were based on military occupations. In order to allow comparisons with other meta-analyses excluding military occupations (Salgado et al., 2003b) the operational validity for training success was re-analyzed excluding military occupations ($k = 84$; $N = 10,839$). The operational validity (.484) did not differ significantly from the operational validity of the whole sample (.467).

References

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

- *Althoff, K. (1968) Bericht über Bewährungskontrollen der Eignungsuntersuchungen im Land Niedersachsen [Report about evaluations of aptitude tests in Lower Saxony]. *DGP Informationen*, 7/1968, 1–36.
- *Althoff, K. (1971) Die Voraussage des Berufserfolges von Regierungsinspektoren und Revierförstern – zwei Bewährungskontrollen [The prediction of professional success for government- and forest inspectors: two tests of validity]. *Diagnostica*, 17, 145–180.
- *Althoff, K. (1974) Kurzbericht über eine Bewährungskontrolle der Eignungsprognosen für Bewerber der allgemeinen Verwaltung in Nordrhein-Westfalen [Report about the predictive validity of aptitude tests for applicants in the administration of North Rhine-Westphalia]. *DGP Informationen*, 12/1974, 12–21.
- *Althoff, K. (1977) Zusammenhänge zwischen Ergebnissen von Eignungstests und beruflicher Bewährung: Dargestellt am Beispiel der Auswahl von Bewerbern für den gehobenen und höheren Polizeivollzugsdienst [Correlation between aptitude tests and occupational performance: an example of applicants in the upper and higher police service]. *Schriftenreihe der Polizei-Führungsakademie*, 1, 6–26.
- *Althoff, K. (1982) Bewährungskontrolle IV – Die Vorhersage des Ausbildungserfolges von Assistentenanwärtern verschiedener niedersächsischer Kommunen [Evaluation

- IV – prediction of training success of applicants in different communes of Lower Saxony]. *DGP Informationen*, 12/1982, 58–66.
- *Althoff, K. (1984) Zur prognostischen Validität von Intelligenz- und Leistungstests im Rahmen der Eignungsdiagnostik [Predictive validity of intelligence and aptitude tests in aptitude testing]. *Psychologie und Praxis – Zeitschrift für Arbeits- und Organisationspsychologie*, 28, 144–148.
- *Amthauer, R. (1973) *I-S-T 70: Intelligenz-Struktur-Test [I-S-T 70: Intelligence structure test]*, (4th edn). Göttingen, Germany: Hogrefe.
- Amthauer, R., Brocke, B., Liepmann, D. and Beauducel, A. (2001) *I-S-T 2000 R: Intelligenz-Struktur-Test 2000 R [I-S-T 200 R: Intelligence structure test 2000 R]*. Göttingen, Germany: Hogrefe.
- Ashworth, S.D., Osburn, H.G., Callender, J.C. and Boyle, K.A. (1992) The effect of unrepresented studies on the robustness of validity generalization results. *Personnel Psychology*, 45, 341–361.
- *Awiszus, M. (1967) Über den Voraussagewert des Intelligenzstrukturtests (I-S-T) für die Leistung in einer Berufsaufbauschule [On the predictive validity of the Intelligence Structure Test (I-S-T) for performance at a vocational school]. *Der Schulpsychologe*, 50–61.
- *Backhaus, J., Müller, S. and Steinmann, V. (1997) Zur Prognose des Berufserfolgs durch den Eignungstest “Bankkaufmann” [Prediction of career success with the aptitude test “Bankkaufmann”]. *Sparkasse*, 7, 338–343.
- Baumert, J., Cortina, K.S. and Leschinsky, A. (2003) Grundlegende Entwicklungen und Strukturprobleme im allgemein bildenden Schulwesen [Developments and structural problems in the educational system]. In Cortina, K.S., Baumert, J., Leschinsky, A., Mayer, K.U. and Trommer, L. (eds), *Das Bildungswesen in der Bundesrepublik Deutschland [The educational system in the Federal Republic of Germany]*. Hamburg, Germany: Rowohlt Taschenbuch, pp. 52–147.
- *Benedikt, H.-P. and Mittmann S. (2004) Evaluierung des Auswahlverfahrens für Auszubildende in einer ostdeutschen Sparkasse [Evaluation of a selection procedure for applicants of a savings bank in East Germany]. *Zeitschrift für Personalpsychologie*, 3, 128–134.
- Bertua, C., Anderson, N. and Salgado, J.F. (2005) The predictive validity of cognitive ability tests: a UK meta-analysis. *Journal of Occupational and Organizational Psychology*, 78, 387–409.
- Brambring, M. (1983) Spezielle Eignungsdiagnostik [Special measurement of aptitude]. In Groffmann, K.-J. and Michel, L. (eds), *Enzyklopädie der Psychologie: Intelligenz- und Leistungsdiagnostik [Encyclopedia of psychology: intelligence and performance measurement]*. Göttingen: Hogrefe, pp. 414–481.
- *Brandstätter, H. (1970) *Leistungsprognose und Erfolgskontrolle [Performance prediction and evaluation]*. Bern: Huber.
- *Bretz, E., and Oldendörp, H. (1992) Bewährungskontrolle: Vorhersage des Ausbildungserfolgs im Angestelltenlehrgang I [Evaluation: prediction of training success in training course I]. *DGP Informationen*, 03/1992, 75–84.
- Brodbeck, F.C., Frese, M., Akerblom, S., Audia, G., Bakacsi, G., Bendova, H., et al. (2000) Cultural variation of leadership prototypes across 22 European countries. *Journal of Occupational and Organizational Psychology*, 73, 1–29.
- Buchmann, C. and Dalton, B. (2002) Interpersonal influences and educational aspirations in 12 countries: the importance of institutional context. *Sociology of Education*, 75, 99–122.
- *Deutsche Lufthansa AG (1980) *Validierung der Auswahl von auszubildenden Flugzeugmechanikern [Validation of the selection of aircraft mechanic apprentices]*. Frankfurt, Germany: Deutsche Lufthansa AG.
- *Diepenbrock, W. and Thielepape, M. (1973) Untersuchung über die Vorhersage des Ausbildungserfolgs von Meisteranwärtern in der Automobilproduktion: Eine Bewährungskontrolle [Examination of the prediction of vocational performance in the automotive industry: an evaluation]. *DGP Informationen*, 09/1973, 1–14.
- Elias, P. and Birch, M. (1994) *Establishment of community-wide occupational statistics. ISCO 88 (COM). A guide for users*. UK: University of Warwick.
- *Funke, U. (1986) Die Validität verschiedener eignungsdiagnostischer Verfahren bei Lehrstellenbewerbern [The validity of various vocational aptitude tests for training applicants]. *Psychologie und Praxis – Zeitschrift für Arbeits- und Organisationspsychologie*, 30, 92–97.
- Funke, U., Krauss, J., Schuler, H. and Stapf, K.H. (1987) Zur Prognostizierbarkeit wissenschaftlich-technischer Leistungen mittels Personvariablen: Eine Metaanalyse der Validität diagnostischer Verfahren im Bereich Forschung und Entwicklung [Predicting scientific-technical achievement using personality variables: a metaanalysis of the validity of diagnostic instruments in the field of research and development]. *Gruppendynamik*, 18, 407–428.
- Gottfredson, L.S. (1997) Editorial: mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13–23.
- *Graudenz, H. (1982) Bewährungskontrolle 2 – Vorhersage des Ausbildungserfolgs von Beamtenanwärtern des gehobenen Dienstes beim RP Darmstadt [Evaluation 2 – prediction of training success of applicants for the upper grade of the civil service in Darmstadt]. *DGP Informationen*, 12/1982, 32–50.
- *Graudenz, H. (1984) Personalauswahl für den mittleren Justizvollzugsdienst in Hessen [Personnel selection for prison warders in Hesse]. *DGP Informationen*, 12/1984, 1–17.
- *Greif, S. (1972) *Gruppenintelligenztests: Untersuchungen am WIT, IST, LPS und AIT [Group intelligence tests: an examination of the WIT, IST, LPS, and AIT]*. Bern: Herbert Lang.
- Gustafsson, J.E. and Undheim, J.O. (1996) Individual differences in cognitive functions. In Calfee, R.C. and Berliner, D.C. (eds), *Handbook of educational psychology*. New York: Prentice Hall, pp. 186–242.
- Hartigan, J.A. and Wigdor, A.K. (1989) *Fairness in employment testing: Validity generalization minority and the General Aptitude Test Battery*. Washington, DC: National Research Council (as cited in Bertua, Anderson, & Salgado, 2005).
- *Haustedt, T. (2001) *Prognostische Validität eines Auswahlverfahrens in einem Transportunternehmen [Predictive validity of a selection procedure in a transport company]*. Unpublished diploma thesis, Technical University (TU) Berlin, Berlin, Germany.
- Hofstede, G. (1980) *Culture's consequences: international differences in work-related values*. Beverly Hills, CA: Sage.
- *Holstegge, S. (1994) *Bewerbersauswahl und Ausbildungserfolg im öffentlichen Dienst [Applicant selection and vocational suc-*

- cess in the civil service]. Unpublished diploma thesis, University of Bielefeld, Bielefeld, Germany.
- Horn, W. (1983) *Leistungsprüfsystem: L-P-S [Performance test: LPS]*, 2nd edn. Göttingen, Germany: Hogrefe.
- Hülshager, U.R., Maier, G.W., Stumpp, T. and Muck, P.M. (2006) Vergleich kriteriumsbezogener Validitäten verschiedener Intelligenztests zur Vorhersage von Ausbildungserfolg in Deutschland: Ergebnisse einer Metaanalyse [Comparing criterion-related validities of different intelligence tests for the prediction of training success in Germany: a meta analysis]. *Zeitschrift für Personalpsychologie*, **5**, 145–162.
- Hunter, J.E. and Hunter, R.F. (1984) Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, **96**, 72–98.
- Hunter, J.E. and Schmidt, F.L. (1990) *Methods of meta-analysis: correcting error and bias in research findings*. Thousand Oaks, CA: Sage.
- Hunter, J.E. and 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.
- Hunter, J.E. and Schmidt, F.L. (2004) *Methods of meta-analysis. Correcting error and bias in research findings*, 2nd edn. Thousand Oaks, CA: Sage.
- International Labour Office (1990) *ISCO-88. International Standard Classification of Occupations*. Geneva, Switzerland.
- *Jäger, A.O. (1960) Zum prognostischen Wert psychologischer Eignungsuntersuchungen [On the predictive validity of psychological aptitude testing]. *Psychologische Rundschau*, **11**, 160–178.
- Jäger, O. and Althoff, K. (1994) *Der Wilde-Intelligenz-Test (WIT): Ein Strukturdiagnostikum [Wilde Intelligence Test: A structure oriented diagnostic instrument]*, 2nd edn. Göttingen, Germany: Hogrefe.
- *Jungkunz, D. (1995) *Berufsausbildungserfolg in ausgewählten Ausbildungsberufen des Handwerks. Theoretische Klärung und empirische Analyse [Vocational education in selected craft professions. A theoretical and empirical analysis]*. Weinheim: Deutscher Studien Verlag.
- *Jungkunz, D. and Bodinet, K. (1989) Korrelative Bedeutung von Testergebnissen, schulischer Vorbildung, Berufsschulnoten und Fähigkeitseinschätzungen für Berufsabschlussnoten [Correlative significance of test results, educational background, grades obtained in vocational school and ability ratings for grades on final examinations in vocational training]. *Zeitschrift für Pädagogik*, **35**, 711–730.
- *Jungkunz, D. and John, W. (1991) Bedingungsfaktoren bei der Vergabe von Zeugnisnoten an Auszubildende handwerklicher Ausbildungsberufe [Factors influencing grades of technical apprentices]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, **87**, 481–494.
- Kanning, U.P. and Holling, H. (2002) *Handbuch personaldiagnostischer Instrumente. [Handbook of personnel assessment instruments]*. Göttingen, Germany: Hogrefe.
- *Kersting, M. (2001) Zur Konstrukt- und Kriteriumsvalidität von Problemlösezenarien anhand der Vorhersage von Vorgesetztenurteilen über die berufliche Bewährung [On the construct and criterion validity of problem-solving scenarios based on the prediction of supervisor assessment of job performance]. *Diagnostica*, **47**, 67–76.
- *Kettel, K.J. and Simmat, W. (1969) Beitrag zu einem psychologischen Berufsprofil für Fernmeldehandwerker [On the psychological job profile of telecommunication technicians]. *Psychologie und Praxis*, **13**, 163–174.
- *Kleinevoss, R. (1983) Untersuchung zur Vorhersage des Ausbildungserfolges von Anwärtern des gehobenen Dienstes einer Bundesbehörde [Examination of the prediction of training success of applicants for the upper civil service of a federal authority]. *DGP Informationen*, **12/1983**, 41–72.
- *Kober, U. (1978) Eignungsuntersuchungs-Ergebnisse bei mehreren Ausbildungsberufen in einer Bewährungskontrolle durch verschiedene Methoden [Results of aptitude tests for several professions: an evaluation involving various methods]. *Psychologie und Praxis*, **22**, 168–177.
- Kratzmeier, H. (1979) *RAVEN-Matrizen-Test – standard progressive matrices – manual*. Weinheim, Germany: Beltz.
- Kuncel, N.R., Hezlett, S.A. and Ones, D.S. (2004) Academic performance, career potential, creativity, and job performance: can one construct predict them all? *Journal of Personality and Social Psychology*, **86**, 148–161.
- Levine, E.L., Spector, P.E., Menon, S., Narayanan, L. and Cannon-Bowers, J.A. (1996) Validity generalization for cognitive, psychomotor, and perceptual tests for craft jobs in the utility industry. *Human Performance*, **9**, 1–22.
- *Lometsch, D. (1975) *Untersuchungen zur Vorhersage des Berufserfolgs von Führungsnachwuchskräften [Analysis of the prediction of career success of management trainees]*. Unpublished report Deutsche Unilever GmbH (as cited in Jäger & Althoff, 1994).
- *Mack, W. (1996) Intelligenz und Kompetenzentwicklung bei KFZ-Mechanikern [Intelligence and competence development of car mechanics]. In Witruk, E. and Friedrich, G. (eds), *Pädagogische Psychologie im Streit um ein neues Selbstverständnis: Bericht über die 5. Tagung der Fachgruppe Pädagogische Psychologie in der Deutschen Gesellschaft für Psychologie e.V. in Leipzig 1995 [Pedagogical Psychology arguing about a new self-image. Report about the 5th Conference of the section of Pedagogical Psychology of the German Association of Psychology e.V. in Leipzig 1995]*. Landau, Germany: Verlag Empirische Pädagogik, pp. 132–139.
- *Marcus, B., Schuler, H., Quell, P., and Hümpfer, G. (2002) Measuring counter productivity: development and initial validation of a German self-report questionnaire. *International Journal of Selection and Assessment*, **10**, 18–35.
- *Marschner, G. (1971) Der Revisions-Test als allgemeiner Leistungstest [The Revisions Test as a general achievement test]. *Diagnostica*, **17**, 14–26.
- *Marschner, G. (1984) Reliabilitäts- und Validitätsuntersuchungen mit dem Prüfsystem für Schul- und Bildungsberatung (PSB) [Reliability and validity studies on the "Prüfsystem fuer Schul- und Bildungsberatung" (PSB)]. *Psychologie und Praxis – Zeitschrift für Arbeits- und Organisationspsychologie*, **28**, 32–38.
- Mosier, C.I. (1943) On the reliability of a weighted composite. *Psychometrika*, **8**, 161–168.
- Müller, W. and Shavit, Y. (1998) The institutional embeddedness of the stratification process. In Shavit, Y. and Müller, W. (eds), *From school to work*. Oxford, UK: Clarendon, pp. 1–48.
- Muñoz Villalobos, V. (2006) *Conferencia de prensa [Press conference]*. Unpublished manuscript, Bundesministerium für Bildung und Forschung, Berlin, Germany.

- Neisser, U., Boodoo, G., Bouchard, T.J. Jr., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Petloff, R., Sternberg, R.J. and Urbina, S. (1996) Intelligence: knowns and unknowns. *American Psychologist*, **51**, 77–101.
- *Neumann, E. (1938) Psychotechnische Eignungsprüfung und Anlernung im Flugmotorenbau [Psychotechnical aptitude testing and instruction in airplane motor construction]. *Industrielle Psychotechnik*, **15**, 111–162.
- OECD. (2004) *Education at a glance. OECD indicators 2004*. Paris: Organization for Economic Co-operation and Development.
- Pearlman, K., Schmidt, F.L. and 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.
- Postlethwaite, T.N. (1988) *The encyclopedia of comparative education and national systems of education*. Oxford, UK: Pergamon.
- *Psychologischer Dienst der Bundeswehr (2004) *Unveröffentlichter Arbeitsbericht [Unpublished report]*. Bonn, Germany: Streitkräfteamt der Bundeswehr.
- Raju, N.S., Burke, M.J., Normand, J. and Langlois, G.M. (1991) A new meta-analytic approach. *Journal of Applied Psychology*, **76**, 432–446.
- Raju, N.S. and Fleer, P.F. (2003) *VG2M: a computer program for conducting validity generalization analysis [Computer software]*. Chicago, IL: Illinois Institute of Technology.
- Rosenthal, R. (1979) The “File Drawer Problem” and tolerance for null results. *Psychological Bulletin*, **86**, 638–641.
- 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.
- Ryan, A.M., McFarland, L., Baron, H. and Page, R. (1999) An international look at selection practices: nation and culture as explanations for variability in practice. *Personnel Psychology*, **52**, 359–391.
- Sackett, P.R. and Ostgaard, D.J. (1994) Job-specific applicant pools and national norms for cognitive ability tests: implications for range restriction corrections in validation research. *Journal of Applied Psychology*, **79**, 680–684.
- Salgado, J.F. and Anderson, N. (2002) Cognitive and GMA testing in the European Community: issues and evidence. *Human Performance*, **15**, 75–96.
- Salgado, J.F. and Anderson, N. (2003) Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology*, **12**, 1–17.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C. and de Fruyt, F. (2003a) International validity generalization of GMA and cognitive abilities: a European community meta-analysis. *Personnel Psychology*, **56**, 573–605.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F. and Rolland, J.P. (2003b) A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, **88**, 1068–1081.
- Sarges, W. and Wottawa, H. (2001) *Handbuch wirtschaftspsychologischer Testverfahren [Handbook of economic-psychological tests]*. Lengerich, Germany: Pabst Science Publishers.
- *Schäfer, N. (1986) *Eignung und Arbeitszufriedenheit: Beiträge zur Eignungsdiagnostik [Occupational skills and job satisfaction. Contributions to aptitude measurement]*. Frankfurt am Main, Germany: Peter Lang.
- 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. and Hunter, J.E. (1977) Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, **62**, 529–540.
- Schmidt, F.L. and 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. and Hunter, J.E. (1999) Theory testing and measurement error. *Intelligence*, **27**, 183–198.
- Schmidt, F.L. and Hunter, J. (2004) General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, **86**, 162–173.
- Schmidt, F.L., Hunter, E. and Pearlman, K. (1981) Task differences as moderators of aptitude test validity in selection: a red herring. *Journal of Applied Psychology*, **66**, 166–185.
- *Schmidt, J.U. (1987) Zur Vorhersagbarkeit des Ausbildungserfolgs im gehobenen Verwaltungsdienst durch einen auf der Grundlage des Berliner Intelligenzstrukturmodells entwickelten Test [On the predictability of training success in the upper grade of the civil service by a test based on the Berlin Intelligence Structure Model]. *DGP Informationen*, **12/1987**, 29–45.
- *Schmidt-Atzert, L., and Deter, B. (1993a) Die Vorhersage des Ausbildungserfolgs bei verschiedenen Berufsgruppen durch Leistungstests [The prediction of training success in different jobs by achievement tests]. *Zeitschrift für Arbeits- und Organisationspsychologie*, **37**, 191–196.
- *Schmidt-Atzert, L., and Deter, B. (1993b) Intelligenz und Ausbildungserfolg: Eine Untersuchung zur prognostischen Validität des I-S-T 70 [Intelligence and training success: a study of the predictive validity of the German Intelligence Test I-S-T 70]. *Zeitschrift für Arbeits- und Organisationspsychologie*, **37**, 52–63.
- *Schmidt-Atzert, L., Deter, B., and Jaeckel, S. (2004) Prädiktion von Ausbildungserfolg: allgemeine Intelligenz (g) oder spezifische kognitive Fähigkeiten? [The prediction of training success: General intelligence (g) or specific cognitive abilities?]. *Zeitschrift für Personalpsychologie*, **3**, 147–158.
- Schmitt, N.W., Gooding, R.Z., Noe, R.A. and Kirsch, M. (1984) Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, **37**, 407–422.
- *Scholz, B., and Biernat, J. (1995) Eine Bewährungskontrolle des DGP-Auswahlverfahrens für den mittleren feuerwehrtechnischen Dienst [An evaluation of the DGP selection process for the middle grade of a fire department]. *DGP Informationen*, **12/1995**, 21–30.
- *Schuler, H. (1994) Communication rather than inspiration and perspiration? On performance requirements in highly qualified occupations. In Heller, K.A. and Hany, E.A. (eds), *Competence and responsibility*, Vol. 2 (pp. 112–116). Kirkland, WA: Hogrefe & Huber.
- *Schuler, H., Barthel, E., and Fünfgelt, V. (1984) Erfolg von Mädchen in gewerblich-technischen Ausbildungsberufen: Ein Modellversuch [Occupational success of girls in technical professions]. *Psychologie und Praxis – Zeitschrift für Arbeits- und Organisationspsychologie*, **28**, 67–78.

- Schuler, H. and Funke, U. (1993) Diagnose beruflicher Eignung und Leistung [The diagnosis of occupational aptitude and performance]. In Schuler, H. (ed.) *Lehrbuch Organisationspsychologie*. [Textbook of organizational psychology]. Bern, Switzerland: Huber, pp. 235–283.
- Schuler, H., Funke, U. and Baron-Boldt, J. (1990) Predictive validity of school grades: a meta-analysis. *Applied Psychology: An International Review*, **39**, 89–103.
- *Schuler, H., Moser, K., Diemand, A. and Funke, U. (1995) Validität eines Einstellungsinterviews zur Prognose des Ausbildungserfolgs [Validity of an employment interview for the prediction of training success]. *Zeitschrift für Pädagogische Psychologie*, **9**, 45–54.
- *Schwadorf, H. (2003) *Berufliche Handlungskompetenz [Professional competence]*. Hohenheim, Germany: ibw.
- *Seggebruch, G. (1982) Zur Vorhersage des Berufserfolgs von Bewerbern für den mittleren Verwaltungsdienst in einer Kommunalverwaltung [On the prediction of training success of applicants for the middle grade of the civil service]. *DGP Informationen*, *12/1982*, 6–31.
- *Seggebruch, G. (1984) Vergleich der Ergebnisse aus Eignungsuntersuchungen und Zwischenprüfungen im gehobenen allgemeinen Verwaltungsdienst – eine Bewährungskontrolle [Comparison of aptitude tests and intermediate examinations of the upper grade of the civil service – an evaluation]. *DGP Informationen*, *12/1984*, 56–74.
- *Streitkräfteamt (1996a) *Unveröffentlichter Arbeitsbericht [Unpublished report]*. Bonn, Germany: Streitkräfteamt der Bundeswehr.
- *Streitkräfteamt (1996b) *Unveröffentlichter Arbeitsbericht [Unpublished report]*. Bonn, Germany: Streitkräfteamt der Bundeswehr.
- *Streitkräfteamt (2001) *Unveröffentlichter Arbeitsbericht [Unpublished report]*. Bonn, Germany: Streitkräfteamt der Bundeswehr.
- Tent (1969) *Die Auslese von Schülern für weiterführende Schulen: Möglichkeiten und Grenzen [Selecting pupils for secondary schools: chances and limits]*, 2nd edn. Göttingen, Germany: Hogrefe.
- *Thielepape, M. (1980) Untersuchung über die Vorhersage des Ausbildungserfolgs von Krankenschwestern und –pflegern – eine Bewährungskontrolle [Evaluation of the prediction of training success of hospital nurses]. *DGP Informationen*, *07/1980*, 21–43.
- US Department of Labor (1991) *Dictionary of occupational titles*, 4th edn. Washington, DC: US Department of Labor.
- Viswesvaran, C., Ones, D.S. and Schmidt, F.L. (1996) Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, **81**, 557–574.
- *Weiß, R.H. (1971) *Grundintelligenztest Skala 3 – CFT 3 [The Basic Intelligence Test Scale 3 – CFT3]*. Braunschweig, Germany: Westermann.
- Wesley, F. and Murch, G.M. (1966) Die Ausbildung und akademische Laufbahn eines Psychologen in den USA [Education and academic career of a psychologist in the USA]. *Psychologische Rundschau*, **17**, 109–128.
- *Wolf, B. (1990) Untersuchungen zur Vorhersage des Ausbildungserfolgs von Aufstiegsbewerbern für den gehobenen Polizeivollzugsdienst [Examination of the prediction of training success of applicants for the upper grade of the police service]. *DGP Informationen*, *12/1990*, 51–66.
- *Wolff, P., and Voullaire, C. (1968) Eignungsbegutachtung von Körperbehinderten für einen Verwaltungsberuf – eine Bewährungskontrolle [Judgement of qualification of physically handicapped for an administrative occupation – an evaluation]. *Diagnostica*, **14**, 70–87.

Appendix A

Table A1. ISCO-88 COM major groups, classification codes, examples of occupations and related skill levels

Major group	Classification code	Examples of occupations belonging to this group	Skill level
Legislators, senior officials and managers	1	– ^a	4
Professionals	2	Science professionals (different careers)	4
Technicians and associate professionals	3	Police inspectors; chemical science technicians; administrative associate professionals; finance associate professionals; legal associate professionals	3
Clerks	4	Secretaries; numerical clerks; office clerks	2
Service workers and shop and market workers	5	Police officers; hairdressers; flight attendants	2
Skilled agricultural and fishery workers	6	– ^a	2
Craft and related workers	7	Miners; machinery mechanics; electrical mechanics; telegraph and telephone installers and service	2
Plant and machine operators and assemblers	8	– ^a	2
Elementary occupations	9	– ^a	1
Armed forces	0	Non-commissioned marine officer	2
		Commissioned marine officer	4
		Commissioned air force officer	4
		Commissioned army officer	4

Note: ^aNo occupations included in the database fell into this group.