Predicting Training Success with General Mental Ability, Specific Ability Tests, and (Un)Structured Interviews: A meta-analysis with unique samples

Matthias Ziegler*, Erik Dietl, Erik Danay*, Markus Vogel*** and Markus Bühner******

*Psychological Institute, Humboldt-Universität zu Berlin, Berlin, Germany. matthias.ziegler@psychologie.hu-berlin.de
**Psychological Institute, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany
***Provadis, Frankfurt am Main, Germany
****Psychological Institute, Karl-Franzens Universität, Graz, Austria

Several meta-analyses combine an extensive amount of research concerned with predicting training success. General mental ability is regarded as the best predictor with specific abilities or tests explaining little additional variance. However, only few studies measured all predictors within one sample. Thus, intercorrelations were often estimated based on other studies. Moreover, new methods for correcting range restriction are now available. The present meta-analyses used samples derived from a German company in which applicants for different apprenticeships were tested with an intelligence structure test, specific ability tests as well as a structured and an unstructured interview. Therefore, intercorrelations between different assessment tools did not have to be estimated from other data. Results in the final examination, taking place at least 2 years after the original assessment, served as criterion variable. The dominant role of general mental ability was confirmed. However, specific abilities were identified that can be used as valuable additions. Job complexity moderated some of the relationships. Structured interviews were found to have good incremental validity over and above general mental ability. Unstructured interviews, on the other hand, performed poorly. Practical implications are discussed.

1. Introduction

Several meta-analyses show general mental ability (GMA) as a valid predictor of job performance as well as training success (Hülsheger, Maier, & Stumpp, 2007; Hunter & Hunter, 1984; Kramer, 2009b; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Schmidt, 2002). Recent developments concerning the correction of indirect range restriction within meta-analyses suggest that reported operational validities of GMA have been underestimated by about 25% (Hunter, Schmidt, & Le, 2006). An exception is the recently published meta-analysis by Kramer (2009b). Thus, the first objective of the present study was to replicate Kramer's findings by conducting a meta-analysis on the operational validity of GMA for predicting training success in new (unpublished) samples of German applicants including an adequate treatment of indirect range restriction. Based on such meta-analyses, GMA is often portrayed as one of the strongest predictors with other predictors having little or no incremental validity. However, in order to estimate the combined effect of different personnel selection tools, their correlation is needed. Unfortunately, not all studies include the tools in question, thus, the correlation often has to be estimated from other studies. The current study combined only primary studies, which included results of specific cognitive ability tests, supplementary tests as well as structured and unstructured interviews (SI and UI) along with a GMA test. Therefore, the correlations were directly observed and had not to be estimated from other studies. This also allows a closer look at the actual intercorrelations of different training success predictors without having to use other studies. Based on the meta-analytically integrated correlations, regression analyses were conducted to find

International Journal of Selection and Assessment out whether specific tests have incremental validity above and beyond GMA. A special interest here lay within the results for SI and UI, because they belong to the most used selection tools in Germany (Schuler, Hell, Trapmann, Schaar, & Boramir, 2007). Finally, possible biasing effects due to gender will be investigated.

2. GMA and training success

Gottfredson (1997) defined GMA 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' (p. 13). GMA is also often referred to as general cognitive ability, general intelligence, or *g* factor.

An extensive meta-analysis by Hunter and Hunter (1984) indicated an operational validity of GMA as predictor of training success of .54. Job complexity was identified as moderator variable, which suggests that the operational validity of GMA increases with the complexity of the job. Furthermore, Schmidt (2002) demonstrated that the validity of GMA generalizes across different types of tests, occupations, organizations, and criteria.

A lack of meta-analytical studies exploring the criterion validity of GMA measures in European countries was pointed out by Salgado and Anderson (2002). For this reason, the authors conducted a meta-analysis with data from 10 countries of the European Community, which resulted in findings comparable with those in US studies. The operational validity of GMA was specified as .54 for predicting training success. Moreover, evidence for international validity generalization was found (Salgado *et al*., 2003).

The first meta-analysis in Germany conducted by Hülsheger *et al*. (2007) yielded an operational validity (corrected for range restrictions and criterion reliability) of GMA for training success of .47, a value slightly lower than those in United States or European meta-analyses. Hülsheger and colleagues considered the specifics of the German educational system as a potential reason for this difference because these specifics might have caused range restrictions. In fact, a recent study by Lang, Kersting, and Hülsheger (2010) showed that there is a substantial range restriction in specific applicant pools compared with the general norm. Kramer (2009b) provided an extension and update of Hülsheger's meta-analysis while applying two methodological improvements. By means of a substitution procedure, the unreported values of nonsignificant correlations were included and the range restriction of GMA was indirectly corrected. This resulted in a value of .59 for training success in Germany. According to Kramer (2009b), the above-mentioned results for job complexity in the United States and Europe could not be replicated, because for the most complex (ISCO-88 skill level 4, International Labour Office, 1990) and the simplest occupations (skill level 1) only few studies were at hand.

Adequate methods for correction of indirect range restriction in personnel selection studies were developed only recently (Hunter & Schmidt, 2004; Hunter *et al*., 2006; Le & Schmidt, 2006). So far, the majority of meta-analyses applied correction for direct range restriction although range restriction in most data sets is indirect. A reanalysis of four job data sets by Hunter *et al*. (2006) indicates that the inappropriate use of correction for direct range restriction leads to an underestimation of operational validities for both job performance measures (21%) and training performance measures (28%). Moreover, a second investigation of the database used in the meta-analysis of Hunter and Hunter (1984) yielded an underestimation of the correlation between GMA and job performance of about 25% (Hunter *et al*., 2006). The 25% larger operational validity for training success in Kramer's (2009a) meta-analysis compared with Hülsheger *et al*.'s (2007) meta-analysis is in line with these findings. As a consequence, the first objective of the present study was the replication of Kramer's findings for the operational validity of GMA for the prediction of training success in Germany (Kramer, 2009b) in a new set of unpublished studies while adequately treating indirect range restriction. Furthermore, meta-analytical results of specific cognitive ability tests, supplementary tests as well as UI and SI are compared with the GMA results. The present analyses will also try to test the often-reported finding that more specific cognitive ability tests have no incremental validity above and beyond GMA. Recent findings in academic settings provide evidence for the validity of more specific abilities (Ziegler, Danay, Schölmerich, & Bühner, 2010).

3. Operational validities of SI and UI

SI and UI differ in their degree of standardization. Unlike SIs, UIs are conducted less systematically, which means that questions and ratings are not fixed and often vary between different applicants. This usually leads to a lower average validity of UIs compared with SIs (Schmidt & Hunter, 1998). However, when training performance was used as criterion, McDaniel, Whetzel, Schmidt, and Mauer (1994) found comparable values for both types of interviews. Their meta-analysis yielded operational validities of .34 and .36 for SI and SIs, respectively. According to Schmidt and Hunter (1998), this small difference appears to be the result of a second-order sampling error (Hunter & Schmidt, 2004), and for this reason, they used an average value of .35 as validity estimate for SI and UI in predicting training success.

3.1. Incremental validities of SI and UI

Incremental validities of SI and UI above and beyond GMA are highly affected by the correlation between GMA and interview scores. More specifically, the smaller the correlation, the higher the probability that interview scores account for variance in job performance or training success above GMA. To estimate the incremental validity of interviews, Schmidt and Hunter (1998) used the correlation of .32 between interview scores and GMA, an overall figure that originated from an analysis by Huffcutt, Roth, and McDaniel (1996, see also Huffcutt & Arthur, 1994). In a recent new analysis Berry, Sackett, and Landers (2007) reported considerably lower estimates (below. 30). However, a critical issue is that the estimation of the correlation between interview scores and cognitive ability tests often has to use data from many studies. Thus, the correlation is estimated using information from different sources. It is unclear as to how far this might affect the estimation. Therefore, this study only included samples in which all information was available.

Regarding the prediction of training performance, Schmidt and Hunter (1998) reported an operational validity of .56 for GMA tests and an incremental validity of interviews of .03, a 5% increase in the multiple R compared with the GMA results. Furthermore, considering job performance as criterion, their analysis yielded an operational validity of .51 for GMA tests and incremental validities of .12 and .04 for SI and UI, which indicates an increase in validities of 24% and 8%, respectively. However, these numbers are based on the inferences made with regard to the bivariate correlations between interviews and GMA, which might be distorted as explained above. If these correlations were actually smaller than assumed, the incremental interview validity could be underestimated. Moreover, the previously mentioned limitations with regard to correcting for range restriction also distorted the reported findings and may also have affected the estimate for the GMA–interview correlation. Recent research demonstrates the importance of careful consideration of range restriction in the examination of predictor intercorrelations, in particular when both predictors are used in combination (Sackett, Lievens, Berry, & Landers, 2007).

Based on the questions raised above, the present analyses aimed at answering four specific questions for a German context:

1. How large are the correlations between different training success predictors when using only data including both measures and applying recent techniques to correct for range restriction?
2. How large are the operational validities for GMA, specific aptitude tests, and interviews (structured and unstructured) for training success when applying recent techniques to correct for range restriction?

   2.1. Does job complexity act as a moderator as has been reported previously?

3. Do interviews and/or specific cognitive ability tests have incremental validity above and beyond GMA when predicting training success and applying recent techniques to correct for range restriction?
4. Are there any biases due to gender differences?

4. Method

In order to examine the questions mentioned above, separate data sets, all collected in one company over the course of 3 years, were integrated. The company selects apprentices each year out of more than 2000 applicants. For this purpose, a test battery is used that is described in more detail later. The results of the applicants' test battery were related to the results of their final theoretical examination at the end of the vocational training. In the following section, the methodological procedure will be described. First of all, information about the eight primary studies as well as the selection procedure used in the cooperating company is provided. Subsequently, predictors included in the test battery and the criterion are specified followed by a brief section containing information about artifacts analyzed in this study.

4.1. Primary studies

The current study is based on a data set provided by a cooperating company operating in the pharmaceutical industry. All in all, data from eight primary samples were obtained. Applicants can be divided into eight different occupational groups, each of which was treated as a primary study: chemical skilled workers, pharmaceutical technicians, chemical laboratory workers, biology lab assistants, office communication assistants, foreign language correspondence clerks, electronic technicians (specialized in industrial engineering, electronic process control, or mechatronics), and mechanics (specialized in industrial, automotive, or precision mechanics). In general, the applicants' training lasted for 2.5–3.5 years, with high school graduates having the option to shorten their training by a period of 6 months. Table 1 shows the distribution of the applicants into the different occupational groups and provides information about their age and educational achievement.

4.2. Selection procedure of the company

The cooperating company receives about 6,000–7,000 written applications for 400–450 available apprenticeship training positions every year. In order to select the best applicants, a procedure of multiple selection stages is used. In the first stage, the written applications are examined regarding the school grades and results of the final exams. Subsequently, approximately 2,000 applicants enter the second stage, in which they have to pass an employment test that covers GMA and specific abilities considered to be essential for the different

Table 1. Composition of primary studies: age, educational achievement, predictors, and criterion

Occupational group	Age at the start of the training			Educational achievement				GMA			Verbal			Numerical		
	M	*SD*	*n*	HS	RS	FA	Abi	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Chemical skilled workers	18.0	2.8	239	26	188	11	13	101.1	5.4	217	101.1	6.8	217	101.1	7.0	217
Pharmaceutical technicians	17.4	1.5	45	2	33	2	8	100.6	5.5	43	101.8	6.4	43	99.4	7.0	43
Chemical laboratory workers	18.4	1.5	143	0	40	4	98	105.9	4.2	142	106.4	4.8	142	105.4	6.2	142
Biology lab assistants	18.9	1.5	71	0	9	4	57	104.9	4.5	71	106.0	4.5	71	103.8	6.3	71
Office communication assistants	17.4	1.4	40	0	34	3	3	102.3	4.6	40	104.5	6.0	40	100.1	6.2	40
Foreign language correspondence clerks	19.3	.9	72	0	1	0	71	103.7	4.4	71	104.2	4.3	71	103.2	6.3	71
Electronic technician occupations	17.4	1.6	93	2	69	1	8	104.4	4.9	70	104.1	5.9	70	104.8	6.8	70
Mechanic occupations	17.3	1.8	68	12	52	0	0	101.5	5.8	58	101.6	7.6	58	101.4	6.5	58

Occupational group	Dictation			Basic calculations			Structured interview			Unstructured interview			Final exam		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Chemical skilled workers	101.4	8.0	204	95.4	7.1	217	3.4	.6	200	3.6	.5	206	72.7	10.7	239
Pharmaceutical technicians	104.1	6.0	41	97.3	6.3	43	3.4	.8	41	3.6	.4	41	73.8	10.8	45
Chemical laboratory workers	108.6	4.9	140	101.5	7.4	143	3.4	.6	138	3.9	.5	140	79.3	8.4	143
Biology lab assistants	109.2	4.1	66	101.5	7.6	71	3.9	.6	69	4.1	.6	69	83.4	7.6	71
Office communication assistants	106.9	4.4	40	98.2	5.9	40	3.7	.7	40	4.0	.5	40	69.8	8.8	40
Foreign language correspondence clerks	111.6	2.4	63	102.8	7.0	71	3.7	.5	70	3.9	.4	70	82.3	7.7	72
Electronic technician occupations	102.7	6.9	69	99.4	8.2	69	3.5	.6	65	3.8	.5	65	68.0	11.7	93
Mechanic occupations	98.8	8.5	54	96.4	6.8	58	3.3	.6	54	3.6	.5	55	69.6	9.0	68

Note. Abi = Abitur (grammar school), range from lowest to highest level of education in Germany; FA = Fachabitur (specialized grammar school), HS = Hauptschule (secondary general school), RS = Realschule (secondary intermediate school).

occupation groups. In the last stage, about 1,000 applicants are invited for a job interview, which focuses on their idea of their future occupation, their personal interests and motivation as well as on the overall impression an applicant makes. It is important to note here that at this stage, the applicants are preselected based on cognitive ability (and school grades). At the end of this selection procedure, 400–450 applicants remain eligible for the apprenticeship training with the cooperating company.

4.3. Material

GMA was assessed using selected subtests of the Höchster-Intelligenztest (Heß, 1994), a revision of the Intelligenz-Strukturtest 70 (Amthauer, 1973) that was specifically developed for the cooperating company. The combined scores in two specific cognitive abilities, verbal ability (VA) and numerical ability (NA), compose GMA. They are measured in five subtests that contain 20 items each. Results of the subtests are added up to form the overall measure of GMA. VA includes the subtests Completion of Sentences, Selection of Words, and Analogies. NA is assessed in the subtests Arithmetic Problems and Numerical Series. Age-standardized standard values (mean = 100, standard deviation = 10) are available for the overall GMA scores as well as for the results of the subtests of NA and VA. The norm groups were derived by combining several applicant waves and are thus a good estimation for the general applicant population.

In order to analyze specific job-relevant abilities in addition to GMA, supplementary tests are applied that vary from one occupational group to another. Some of these tests were conducted in all occupational groups and therefore included in the meta-analyses: a Dictation Test (D) to measure spelling skills as well as a Basic Calculation Test (BC) to assess the ability of solving basic arithmetic operations. The cooperating company developed both tests. Raw scores are transformed into standard values (mean = 100, standard deviation = 10), but age-standardized values are not available. Depending on the specific job demands, an English test, a wire bending test, and a cutting test were also applied. Because these tests were not part of the general test battery, their results were not integrated into the meta-analyses but investigated separately.

To complete the test battery, two different interviews are used. In a SI, the applicants' idea of their future occupation is asked and compared with real job tasks. Employing interview experts who are responsible for the applicants' training in the company and therefore have an accurate understanding of the reality of the relevant job is supposed to ensure validity. The more accurate the ideas of an applicant are, the better he will be judged. Secondly, the applicants' appearance and behavior irrespective of their application, certificates or test results are examined in a UI. There is a lack of standards for the rating of aspects such as appearance in this rather subjective process. The UI is conducted largely by graduated psychologists or trainers of the relevant occupations, all of them being employees of the company. Interviewees are rated according to a scale from 1 = *very good* to 5 = *not at all*, which allows for a rating in steps of .5 points. To ease interpretation, interview scores are reverse-recoded here.

Additional tests only used in some subgroups (see Table 7) are a test for English language sufficiency, a science-understanding test, wire bending, and cutting. Whereas the first two tests can be considered as knowledge tests, the latter two require applicants to bend a wire according to a model or to cut a figure out of a sheet of paper. Unfortunately, no information regarding reliabilities was available.

Results of the final written exam at the end of the vocational training were used as criterion for training success. The applicants' exam performance was measured in scores ranging from 0 to 100, with 100 being the best result. For each occupational group, a specific standardized exam investigates knowledge questions relevant for the job. The German Chamber of Commerce and Industry administered all exams.

Table 1 displays means and standard deviations of the seven predictors explored in the meta-analysis (GMA, VA, NA, D, BC, SI, and UI) as well as the criterion (Final Exam) regarding the eight primary samples. Because of preselection in previous stages of the selection process, variances of GMA, VA, and NA as well as D are restricted in all groups and the mean values lie above the norm sample.

4.4. Meta-analytical procedure

One specific feature of this study is that each occupational group can be treated as a primary study. Therefore, it is possible to integrate the eight primary studies and the seven predictors of training success meta-analytically. We applied the psychometric meta-analytical formulas suggested by Hunter and Schmidt (2004) and Hunter *et al*. (2006), which take into account the developments in correction for indirect range restriction. In general, psychometric meta-analyses estimate the amount of observed variance of findings across studies that can be explained by artifactual errors (%VE). The present study considers sampling error, criterion and predictor reliability, and range restriction of the predictors as artifacts. As Hunter and Schmidt (2004) point out, if more than 75% of the observed variance of a predictor across several studies is explained by artifactual errors, the remaining variability may well be accounted for by additional artifactual errors not analyzed. This might also suggest that moderators for the respective predictor do not exist. Because of the fact that the operational validity indicates the estimated true validity of a measure for applied purposes, no corrections for predictor reliability were made. Nevertheless, corrections for the reliability of the criterion and range restriction were made in order to calculate the mean operational validity ($\hat{\rho}$), the variance of ($\hat{\rho}$), the 90% credibility values (CV) and the 95% confidence interval. Solely for estimations of the amount of variance explained by artifacts, corrections for the predictor reliability were taken into account additionally. The 90% CV is related to the variability of the single correlations across studies and was used to determine whether validity generalization could be assumed. If it lies above zero, so does the operational validity of 90% of the studies included in the meta-analysis. More specifically, if 100% of the observed variance is explained by artifacts, the 90% CV corresponds with ($\hat{\rho}$). The 95% confidence interval is related to the variability of ($\hat{\rho}$) in a single study. All calculations have been made using MS Excel, SPSS, and MPlus.

4.5. Information about artifacts

Unlike in the majority of meta-analyses, we corrected nearly every correlation individually and did not need to develop a specific distribution for each artifact because most of the necessary information could be derived from the primary samples. All occupational groups pass the same test battery, therefore variation for the artifacts predictor and criterion reliability between the eight groups does not exist. Table 2 displays the available artifact information of the seven predictors and the criterion. Further details are outlined in the next section.

4.6. Range restriction

Because of the availability of age-standardized standard values reflecting the applicant population, the predictors GMA, NA, and VA yielded an $SD_{\text{Applicants}}$ of 10. Scales of D and BC were transformed to equal the GMA scale, which also leads to an $SD_{\text{Applicants}}$ of 10. The $SD_{\text{Applicants}}$ values of .80 and .81 regarding the SI and UI originated from all applicants that were invited to stage three of the selection procedure (SI: $N = 2{,}831$, UI: $N = 2{,}877$). The observed variance in the primary studies used was below the values from the applicant population and were therefore corrected for range restriction.

Table 2. Overview of artifact information

Predictor	Range Restriction ($SD_{\text{Applicants}}$)	Predictor reliability	Criterion reliability
GMA	10	.90	.64
Verbal ability	10	.83	.64
Numerical ability	10	.88	.64
Dictation	10	.90	.64
Basic calculations	10	.90	.64
Structured interview	0.80	.89	.64
Unstructured interview	0.81	.78	.64

Note. For D, the very low SD of the foreign language correspondence clerks caused calculation problems (i.e., corrected correlations > 1), so we used the next higher SD of another occupational group (biology lab assistants), see Schmidt, Oh, and Le (2006).

4.7. Predictor reliability

Estimates for the reliability coefficients of GMA, NA, and VA were calculated based on a formula by Lienert and Raatz (1998, p. 330), a modification of Mosier's (1943) more general formula for equal weighted factors in which intercorrelations of subtests are taken into account. Our calculations yield reliabilities of .90, .83, and .88 for GMA, VA, and NA, respectively. Information about reliability coefficients was available neither for the two supplementary tests D and BC nor for the UI and SI. Correct disattenuation of indirect range restriction requires information about predictor reliabilities (Hunter *et al*., 2006). Therefore, concerning D and BC, the value of .90 for both (we thank an anonymous reviewer for this recommendation) was chosen. We used the same value as for GMA, so that D and BC did not have an advantage over GMA in the meta-analytic disattenuation process. With regard to the UI and SI, reliability coefficients of a meta-analyses investigating the validity of employment interviews (McDaniel *et al*., 1994) were used. The medians of .78 and .89 for the UI and SI were considered to be appropriate values because the reliability distribution is skewed and we therefore decided against using the mean which would lead to less conservative estimates.

4.8. Criterion reliability

The German Chamber of Commerce and Industry provides no information regarding the reliability of their final exam. In that case, Raju, Burke, Normand, and Langlois (1991) recommend the use of the mean value identified by a meta-analysis investigating the same criterion based on a comparable sample. Schuler, Funke, and Baron Boldt (1990) conducted a meta-analysis exploring the predictive validity of school grades for vocational training in a sample of German primary studies, which results in a hypothetical artifact distribution with a mean value of .64. Because of the fact that the current study is based on a sample of German primary studies as well and both meta-analyses investigate corresponding criterions, this mean value was used for our calculations as well.

4.9. Meta-analyses of incremental validities above and beyond GMA

In a first step, all bivariate correlations between the different predictors were meta-analytically derived as described above. We conducted this procedure twice. In one analysis, we corrected for range restriction and unreliability in both predictors. However, when asking for increments we wanted to have operational values, thus we also conducted the same analyses without correcting for the unreliability of the predictors. Based on these correlations, hierarchical regression analyses were conducted. In each of the analyses, the final exam served as criterion and GMA was entered in the first block with one of the remaining predictors being entered in a second block. Because GMA represents the combination of VA and NA, these two tests were not included in the regression analyses to avoid spurious effects due to multicollinearity. An important aspect here is that all necessary intercorrelations were available and did not have to be estimated based on other studies.

This analysis was accompanied by a more fine-grained look at specific abilities that were tested only for some job groups. Because the total number of primary studies was too small here, we did not integrate these findings meta-analytically. Instead hierarchical regression analyses were performed in which the final exam served as the criterion and GMA was entered as predictor in a first block with the specific abilities following in a further block.

5. Results

5.1. Correlation between predictors

Table 3 contains the results for the meta-analytical integration of the correlations between all predictors. It can be seen that GMA and VA as well as NA were closely related. This is not surprising considering that GMA is the sum of both others. The correlations between GMA and BC as well as D were lower but still substantial. However, the relationships between GMA and both kinds of interviews were close to zero. With the exceptions of the correlations between NA and BC as well as SI and UI, all other predictor intercorrelations were small.

5.2. Validity of GMA, specific cognitive abilities, supplementary tests, and interviews for predicting training success

The second research question aimed to replicate findings regarding the predictive validity of different selection tools for training success in Germany using recent meta-analytical techniques. The results of the meta-analyses conducted are summarized in Table 4. Simply looking at the average correlations shows that the specific ability tests and the GMA test did not differ that much. However, the corrected values show the advantage of GMA. GMA emerged as the best predictor for training success with an operational validity of .65, followed by VA ($\hat{\rho} = .45$) and NA ($\hat{\rho} = .41$). Given that none of the 90% CV included a negative value, validity generalization across the eight occupational groups examined in this study could be assumed.

The more specific aptitude tests BC and D performed slightly less well. Validity generalization could be demonstrated for D, whereas the 90% CV could not be calculated for BC because of a negative estimate of the variance of $\hat{\rho}$, which suggests a zero variance for $\hat{\rho}$ (Hunter & Schmidt, 2004). Therefore, the operational validity of BC also generalized across all eight occupations investigated.

With an operational validity of .23, the SI can be considered as a moderately good predictor of training success and its validity generalized across all occupations. The UI yielded the lowest operational validity of .13, which demonstrates that it is a less useful predictor for training success. Furthermore, its validity did not generalize.

More than 75% of the observed variance of the predictors GMA, NA, VA, BC, and SI could be explained by the four artifactual errors included in this study (three for BC). This result suggests that the remaining variability may be accounted for by other artifactual errors not considered here and that there are no moderators for these predictors (Hunter & Schmidt, 2004). For D and UI, the amount of variance accounted for by artifacts was 19% and 26%, respectively, which indicates that moderators probably exist for those predictors.

5.3. Analysis of moderator effects – validity for low and medium job complexity

The personnel manager of the cooperating company assigned the eight occupational groups investigated in this study to two levels of job complexity. Chemical Skilled Workers, Pharmaceutical Technicians, Electronic

Table 3. Meta-analytic correlations between predictors

	1	2	3	4	5	6	7
1. GMA	(0.90)	1.00	1.00	.62	.74	−.01	.19
2. Verbal ability	.88	(0.83)	0.50	.59	.09	.02	.25
3. Numerical ability	.91	0.43	(0.88)	.26	.71	−.03	.05
4. Dictation	.56	0.51	0.23	(.90)	.34	.08	.18
5. Basic calculations	.67	0.08	0.64	.30	(.90)	−.02	.21
6. Structured interview	−.01	0.02	−0.03	.07	−.02	(.89)	.66
7. Unstructured interview	.16	0.20	0.04	.15	.18	.55	(.78)

Note. $N = 1{,}110–1{,}319$. All meta-analytic correlations below the diagonal are corrected for range restriction (only for the more reliable predictor except D and BC). All meta-analytic correlations above the diagonal are corrected for range restriction (only for the more reliable predictor except D and BC) and unreliability in both predictors. Values in diagonal are the reliabilities.

Table 4. Meta-analyses of all predictors for predicting training success

Predictor	k	N	M_r	S^2_r	$\hat{\rho}$	$S^2\ \hat{\rho}$	%VE	90% CV	95% CI	r
GMA	8	712	.22	.02	.65	.01	95	.51–.79	.49–.81	.28
Verbal ability	8	712	.15	.02	.45	.03	82	.20–.70	.26–.64	.24
Numerical ability	8	712	.19	.01	.41	.01	85	.27–.54	.27–.55	.22
Dictation	8	677	.22	.01	.38	.06	19	.01–.75	.19–.57	.35
Basic calculations	8	712	.18	.01	.33	.00	100	.33	.21–.45	.27
Structured interview	8	677	.14	.01	.23	.01	79	.08–.37	.10–.36	.19
Unstructured interview	8	686	.03	.02	.13	.15	26	−.44–.70	−.17–.43	.13

Note. %VE = variance accounted for by artifacts; 90% CV = 90% lower and upper credibility value; 95% CI = 95% confidence interval; k = number of studies; M_r = weighted average of observed validity; N = total number of participants; r = zero-order correlation after combining the samples; S^2_r = observed variance of validity coefficients; $\hat{\rho}$ = mean operational validity (corrected for range restriction and criterion reliability); $S^2\ \hat{\rho}$ = estimate of the variance of $\hat{\rho}$.

Technician Occupations and Mechanic Occupations were assigned to the group of low job complexity. The group of medium job complexity included Chemical Laboratory Workers, Biology Lab Assistants, Office Communication Assistants, and Foreign Language Correspondence Clerks.

Results of the meta-analyses regarding the different levels of job complexity for the prediction of training success are reported in Table 5. Both groups only included four studies each; therefore the results should be interpreted with caution. The total sample size ranged from 309 to 388. The largest operational validity was found for the specific aptitude test D with a value of .85 in the medium complexity group, whereas it resulted in a lower operational validity of .23 in the low complexity group. In both groups, 100% of the variance was accounted for by artifacts and validity generalization could be assumed.

With a value of −.25, the UI yielded the lowest operational validity of all predictors in the medium complexity group while an operational validity of .40 was attained in the low complexity group. In this group, there was also no evidence for the existence of further moderators. This was not the case in the medium complexity group. For reasons of clarity, the analyses were conducted for all predictors.

5.4. Incremental validities above and beyond GMA

The third research question was whether interviews and/or specific cognitive ability tests have an incremental validity above and beyond GMA. To answer this question, a complete meta-analytically derived correlation matrix was first composed containing all bivariate correlations (see Table 3). The results for the regression analyses based on these intercorrelations can be found in Table 6. It can be seen that only SI had a significant and positive regression weight after controlling for GMA. Because of the multicollinearity between GMA and BC, a suppressor effect occurred.

These analyses were accompanied by a closer analysis within specific job groups. The results as shown in Table 7 indicated that the specific ability tests indeed have incremental validity above and beyond GMA. For the subgroup containing office communication assistants and foreign language correspondence clerks, the English test had incremental validity above GMA. For the other subgroup, science understanding and cutting were incremental to GMA and wire bending acted as a suppressor.

5.5. Gender effects

A last analysis was conducted to test whether the results found might be different for men and women. Meade and Fetzer (2009) suggested to test such biases using a hierarchical regression analysis. In the first block, only the predictor variable was entered. The second block contained the dichotomous group variable. A significant change in explained variance can be seen as evidence for different intercepts. The interaction between predictor and group variable (gender) was entered in the third step. A significant result here would indicate differing slopes for the groups. As can be seen in Table 8, neither predictor was affected by differing intercepts or slopes because of gender.

6. Discussion

The present study was conducted to investigate four specific research questions regarding the prediction of training success using different personnel selection tools as well as the relationships between these tools. To this end, recently developed meta-analytical techniques were used to integrate data sets obtained in a large German company. One outstanding feature of this study is that all predictors used were actually assessed in each sample so that correlations between them did not have to be estimated from other data sets. The results replicated prior findings indicating GMA as a strong predictor of training performance. However, the SI as well as some very specific cognitive ability tests were found to predict variance above and beyond GMA. All effects found did not differ by gender.

Table 5. Meta-analyses of all predictors for training success by low and medium job complexity

Predictor	k	N	M_r	S^2_r	$\hat{\rho}$	$S^2\,\hat{\rho}$	%VE	90% CV	95% CI	r
Low job complexity										
GMA	4	388	.24	.02	.62	.01	91	.45–.79	.40–.84	.18
Verbal ability	4	388	.20	.02	.47	−.01	100	.47	.29–.65	.16
Numerical ability	4	388	.18	.01	.35	.01	86	.23–.48	.17–.54	.12
Dictation	4	368	.14	.00	.23	−.01	100	.23	.13–.34	.12
Basic calculations	4	387	.14	.00	.27	−.03	100	.27	.20–.35	.10
Structured interview	4	360	.21	.00	.32	−.02	100	.32	.22–.42	.19
Unstructured interview	4	367	.12	.00	.40	−.04	100	.40	.24–.56	.08
Medium job complexity										
GMA	4	324	.20	.02	.72	−.01	100	.72	.50–.94	.23
Verbal ability	4	324	.09	.01	.35	.16	31	−.26–.95	−.14–.84	.09
Numerical ability	4	324	.22	.01	.49	.00	100	.49	.30–.68	.25
Dictation	4	309	.32	.01	.85	−.03	100	.85	.80–.90	.38
Basic calculations	4	325	.21	.02	.40	.02	74	.18–.62	.18–.62	.26
Structured interview	4	317	.06	.01	.09	.01	84	−.04–.22	−.12–.30	.08
Unstructured interview	4	319	−.07	.02	−.25	.12	27	−.75–.26	−.65–.16	−.05

Note. %VE = variance accounted for by artifacts; 90% CV = 90% lower and upper credibility value; 95% CI = 95% confidence interval; k = number of studies; M_r = weighted average of observed validity; N = total number of participants; r = zero-order correlation after combining the samples; S^2_r = observed variance of validity coefficients; $\hat{\rho}$ = mean operational validity (corrected for range restriction and criterion reliability); $S^2\,\hat{\rho}$ = estimate of the variance of $\hat{\rho}$.

Table 6. Meta-analytically derived incremental validities for specific aptitude tests and interviews

	β_{GMA}	$\beta_{2nd\ predictor}$	R	R^2	ΔR^2	ΔR in %
GMA	.65***	–	.65***	.42***	–	–
Dictation	.64***	.02	.65***	.42***	<.01	<.01
Basic calculations	.78***	−.19***	.67***	.44***	.02***	.02
Structured interview	.65***	.24***	.69***	.48***	.06***	.06
Unstructured interview	.65***	.03	.65***	.42***	<.01	<.01

Note. β_{GMA} = standardized regression weight for GMA in block 2; $\beta_{2nd\ predictor}$ = standardized regression weight for the respective second predictor in block 2; R = multiple correlation coefficient for both predictors; R^2 = amount of criterion variance explained by predictors; ΔR^2 = amount of criterion variance incrementally explained by second predictor above GMA; ΔR in % = increase in multiple R due to adding the second predictor. ***$p < .001$.

Table 7. Regression analyses for specific aptitude tests controlling for GMA

Predictors	Office communication assistants and foreign language correspondence clerks[a]				Chemical skilled workers, pharmaceutical technicians, chemical laboratory workers, biology lab assistants, electronic technicians, and mechanics[b]			
	β	R^2	ΔR^2	r	β	R^2	ΔR^2	r
Block 1								
GMA	.20**			.36***	.18***			.26***
		.13***				.07***		
Block 2								
English	.58***			.63***	–			
Science understanding	–				.22***			.25***
Wire bending	–				−.15**			.02
Cutting	–				.18***			.17***
		.43***	.30***			.13***	.06***	

Note. r = zero order test-criterion correlation. [a]$n_1 = 120$. [b]$n_2 = 597$. **$p < .01$; ***$p < .001$.

6.1. GMA, SI, and UI

In most previous analyses, the correlation between GMA and either SI or UI was estimated based on different data sets. Only a few studies included all measures. The primary studies used here all included these measures, which made it possible to estimate their relationships without data from other studies. Surprisingly, the correlation between GMA and the SI was virtually nonexistent when controlling for range restriction and unreliability.

Table 8. Differential prediction for all predictors of training success for gender

Source	N	Block 1: predictor		Block 2: gender (intercept differences)		Block 3: gender × predictor interaction (slope differences)	
		β	R^2	β	ΔR^2	β	ΔR^2
1. GMA	712	.21**	.05	.01	<.01	−.03	<.01
2. Verbal ability	712	.15**	.02	−.02	<.01	.02	<.01
3. Numerical ability	712	.19**	.04	−.03	<.01	−.01	<.01
4. Dictation	678	.17**	.03	.02	<.01	−.05	<.01
5. Basic calculations	712	.18**	.03	−.02	<.01	<.01	<.01
6. Structured interview	677	.12**	.02	.02	<.01	.08	<.01
7. Unstructured interview	686	.03	<.01	<.01	<.01	.09	<.01

**$p<.01$.

Regarding UI, the correlation was still small but increased considerably. It can be assumed that in an unstructured situation, the impact of individual differences in intelligence is more important. In a SI that aims at specific noncognitive aspects, these differences should not make an impact. The fact that the relationship with VA is also close to zero for the SI but even stronger for the UI confirms this. In that sense, applicants try to deduce what the unstructured questioning is about. Advantages in GMA obviously help and then produce better ratings. Prior analyses had used a value of .32 as an estimate of the correlation between GMA and interviews and might therefore have underestimated the effects for interviews. However, it has to be noted here that the meta-analytic integration conducted here also revealed heterogeneity with regard to the correlations. This means that there are potential moderators influencing the correlation between GMA and interviews. One such moderator could be interview quality itself. It could be assumed that the worse the interview is regarding conceptualization and rating guidelines, the more GMA matters. For the primary studies used here, this would mean that the interviews have a high-quality standard. This is supported by the fact that all interviewers were well trained and experienced in the selection process. Unfortunately, no further information regarding the interview was obtainable. Therefore, future research should try to determine the factors moderating the size of the correlation between GMA and interviews.

6.2. Test criterion correlation of GMA, specific cognitive abilities, supplementary tests, and interviews

The data set referred to in this study contained civilian validity studies of eight occupational groups. The best predictor of training success was GMA with an operational validity of .65, followed by VA and NA, with operational validities of .44 and .41, respectively. With operational validities of .38 (D) and .33 (BC), the supplementary tests can be considered good predictors of training success as well.

The operational validities for each interview form did not differ dramatically from those reported previously. This is especially true for the SI. The main focus of the SI was the applicants' ideas regarding their future job. It is reasonable to assume that applicants motivated to get the apprenticeship informed themselves about the potential job tasks. Thus, high ratings in the SI would reflect motivation to get the job as much as existing knowledge about the job in question. The UI aimed at the applicants' appearance and behavior. Here, no clear rating guidelines exist making this rating rather subjective. This might explain the rather poor performance. The fact that the moderator analysis indicated the existence of further moderatos for the UI test criterion correlation for medium complex jobs further underscores the problematic nature of this interview form. However, it also has to be noted here that UIs do not always perform worse than SIs (Schmidt & Zimmerman, 2004).

6.3. Comparison with international GMA results

The value reported for GMA exceeds the operational validities of GMA found in other meta-analyses conducted by Kramer (2009b) (operational validity of GMA = .59), Schmidt and Hunter (1998) (operational validity of GMA = .56), or Salgado *et al.* (2003) (operational validity of GMA = .54). The increase in operational validity (in comparison with Salgado *et al.*, 2003; Schmidt & Hunter, 1998) is most likely due to the application of corrections for indirect rather than direct range restriction. When Hunter *et al.* (2006) reanalysed a data set of petroleum plant operator and maintenance jobs (Schmidt, Hunter, & Caplan, 1981) using the same algorithm applied here, they discovered that operational validities of GMA measures predicting training performance were underestimated by 28%. However, the 95% confidence interval reported here includes the previously reported values showing that the larger point estimate could be due to chance.

Beyond this, differences in operational validities might result from differences in training success criteria used in primary studies and related validity estimates. Consistent with Kramer (2009b), only results of standardized exams were used as a criterion measure for training success in this study. Consequently, the correction value of .64 for the criterion reliability of training success measures is the same value used by Kramer (2009b), although lower than the value of .80 found in the analysis by Hunter and Hunter (1984). However, Salgado *et al.* (2003) used ratings of trainers in most of their primary samples, which results in a much lower value of .56. As Kramer (2009a) used the same value for criterion reliability and corrected range restriction also indirectly, the lower operational validity compared with our study might originate from lower magnitudes of range restriction in the primary samples in Kramer's meta-analysis. Another reason could be the different starting basis of primary samples. It is important to consider these differences when comparing the present findings with other meta-analyses because lower values for criterion reliability estimates lead to higher operational validities and lower magnitudes of range restriction lead to lower operational validities.

6.4. Moderating effect of job complexity

Only the predictors D and UI yielded a variance accounted for by artifacts of $<75\%$. The negative predictive value of UI within the medium complexity group is an unexpected finding, though. As hypothesized above, it could be assumed that the unclear rating objective and the resulting subjective decisions could be behind this problem. In that sense, UIs would only work if the rating aim of the interviewer is valid. Considering the more complex nature of the medium complexity group, incorrect rating aims might have been more widespread. D showed substantial operational values in the low and medium complexity group. Its operational value differed between the two levels of complexity, indicating a moderating effect. Considering the jobs that were put into the two different categories, this result is further supported. The medium complexity jobs where D was a better predictor have more job tasks related to writing, typing, and comparable activities. Within the low complexity category, these tasks have a smaller impact. Previous meta-analyses demonstrated a positive association between job complexity and the magnitude of GMA measures (Hunter & Hunter, 1984; Salgado *et al.*, 2003). Given that D can be considered as a part of VA, our findings correspond with those of the meta-analyses mentioned above and support the idea of specific abilities as important predictors besides GMA. The results for D can further be explained by the assumption that an increase in intelligence involves better spelling abilities because the person acquires knowledge more quickly.

It is important to note that the classification of occupational groups to low and medium level of job complexity was not made using an occupational classification such as the European version of the International Standard Classification of Occupations ISCO-88 COM (Elias & Birch, 1994). Furthermore, only four primary studies were classified to each level of job complexity, therefore the results of the moderator analysis might be influenced by second-order sampling error.

6.5. Incremental validities

A main objective of this study was the examination of incremental validity of different selection tools over GMA for the prediction of training success in Germany. Because of the fact that all correlations between predictors and GMA were available, we did not have to use an estimation based on other studies. The results showed that within the general test battery, only the SI was able to contribute above GMA. This effect was supported by the nonexistent correlation between SI and GMA. The suppressor effect for BC can be explained by the strong correlation between both predictors. One aspect that has to be considered here is the fact that all applicants were preselected based on cognitive abilities before entering the interview phase. Thus, this finding despite a probably large base rate demonstrates the value of SIs. This is in line with other findings and further supports the role of SIs in personnel selection.

Prior studies using samples from the US military did not find an increment of specific abilities above and beyond GMA (Ree, Earles, & Teachout, 1994). Separate hierarchical regression analyses for two job groups were conducted with the German data used in this study. The results supported the idea that some very specific aptitude tests actually capture variance that is predictive above and beyond GMA. In light of Brunswik's lens model (Brunswik, 1955), this result does not come as a surprise. Within the job group of office communication assistants and foreign language clerks who regularly have to deal with written English, English knowledge proved to be important. For the other job group whose job tasks are more hands-on, such specific tasks that actually test practical knowledge or workmanship proved to be incremental. However, the suppression effect that occurred for wire bending indicates that not motoric abilities but actually the underlying cognitive abilities are responsible for the effect. Wire bending acts as a suppressor for cutting. Both tasks require the applicants to work with their hands, thus, here the common variance originates. However, this common variance seems to be unrelated with the criterion, which is not surprising considering its written nature. Yet, cutting also puts a demand on the applicants' figural reasoning abilities, which might explain its test criterion correlation. Thus, putting aside specific aptitudes as predictors of training success based on large studies mixing different job groups with different demands might be premature.

6.6. Gender as a moderating influence

The analyses conducted here did not reveal any differential effects based on gender. Of course, this has to be interpreted with care and does not automatically mean that the selection tools investigated are immune to gender biases. One limiting factor is that all studies were conducted within the same company using the same tests. Thus, generalizations across other tests or companies require replications. Moreover, gender equality is a goal hard striven for in Germany. Therefore, it could be assumed that the results might differ in countries where less value is put on gender equality.

6.7. Practical and theoretical implications

The results generated in this meta-analysis have important implications for the theory and practice of personnel selection and support the idea that GMA validities are generalizable on an international level. From a practical point of view, our findings support the use of GMA measures and other generalizable predictors, in particular SI, for selection purposes of the examined occupations in Germany. By reasons of predictive power, we recommend a combination of GMA and SI when hiring decisions are made. Moreover, depending on the specific job requirements specific tasks should be considered as additional selection tools. Finally, practitioners should abstain from using UIs.

6.8. Limitations and strengths

The findings of this study should be interpreted in consideration of its limitations and strengths. Unlike other United States and European large-scale meta-analyses, our analyses included only eight unpublished primary samples, thereby enhancing the occurrence of second-order sampling error (Hunter & Schmidt, 2004). Furthermore, not all necessary artifact information was available. For example, all applicants were first selected on basis of their school grades. Thus, this already might have caused range restrictions in the predictors. However, because information on all grades was not available, a correction was not possible. Thus, the estimates reported here could be even conservative. Moreover, training success was measured by a single criterion – an objective knowledge test – although it embraces facets such as supervisor ratings or practice examinations as well. Because of the fact that the cooperating company did not use personality questionnaires in the years sampled here, no inferences regarding the predictive power of constructs such as the Big 5 could be drawn.

Nevertheless, this meta-analysis exhibits a number of strengths. It represents the first German meta-analysis of GMA, specific cognitive abilities, supplementary tests, and interviews for the prediction of training success that applies recent developments of correction for indirect range restriction in personnel selection. Moreover, correlations between the predictors and GMA are integrated meta-analytically, using the correlations of all primary studies, thereby also delivering accurate values for the incremental validities of the predictors in Germany. Last but not least, all findings are based on a wide range of occupations.

References

Amthauer, R. (1973). *Intelligenz-Struktur-Test IST 70 [Intelligence-Structure Test. IST 70]* (4th ed.). Göttingen: Hogrefe.

Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisting interview-cognitve ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology*, *60*, 837–874.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Elias, P., & Birch, M. (1994). *Establishment of community-wide occupational statistics. ISCO-88 (COM). A guide for users*. Warwick Institute for Employment Research. University of Warwick, Coventry.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.

Heß, M. (1994). *Der Hoechster Intelligenztest. Revision auf der Basis des I-S-T 70 von R. Amthauer [The Hoechster Intelligence Test. Revision on the basis of the I-S-T 70 by R. Amthauer.]*. Frankfurt: Interner Bericht der Hoechst AG.

Huffcutt, A., & Arthur, W., Jr (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*, 184–190.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, *81*, 459–473.

Hülsheger, 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*, 3–18.

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. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology*, *91*, 594–612.

International Labour Office. (Ed.) (1990). *International standard classification of occupations: ISCO-88*. Geneva: International Labour Office.

Kramer, J. (2009a). *Metaanalytische Studien zu Intelligenz und Berufsleistung in Deutschland [Meta-analytical studies of intelligence and work performance in Germany]*. Unpublished dissertation, Universität Bonn, Bonn.

Kramer, J. (2009b). General mental ability and occupational success in Germany: Further metaanalytic elaborations and amplifications. *Psychologische Rundschau*, *60*, 82–98.

Lang, J., Kersting, M., & Hülsheger, U. (2010). Range shrinkage of cognitive ability test scores in applicant pools for German Governmental Jobs: Implications for range restriction corrections. *International Journal of Selection and Assessment*, *18*, 321–328.

Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new meta-analytic procedure. *Psychological Methods*, *11*, 416–438.

Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse (Test-building and test-analysis)*. Weinheim: Beltz.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Mauer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*, 599–616.

Meade, A., & Fetzer, M. (2009). Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods*, *12*, 738–761.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, *8*, 161–168.

Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology*, *76*, 432–446.

Ree, M. J., Earles, J. A., & Teachout, M. (1994). Predicting job performance: Not much for than g. *Journal of Applied Psychology*, *79*, 518–524.

Sackett, P. R., Lievens, F., Berry, C. M., & Landers, R. N. (2007). A cautionary note on the effects of range restriction on predictor intercorrelations. *Journal of Applied Psychology*, *92*, 538–544.

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., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *6*, 573–605.

Schmidt, F., & Zimmerman, R. (2004). A counterintuitive hypothesis about employment interview validity and some supporting evidence. *Journal of Applied Psychology*, *89*, 553–561.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187–210.

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., & Caplan, J. R. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, *66*, 261–273.

Schmidt, F. L., Oh, I. S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology*, *59*, 281–305.

Schuler, H., Funke, U., & Baron Boldt, J. (1990). Predictive validity of school grades – A meta-analysis. *Applied Psychology: An International Review*, *39*, 89–103.

Schuler, H., Hell, B., Trapmann, S., Schaar, H., & Boramir, I. (2007). Die Nutzung psychologischer Verfahren der externen Personalauswahl in deutschen Unternehmen. Ein Vergleich ueber 20 Jahre [Use of personnel selection instruments in German organizations during the last 20 years]. *Zeitschrift fuer Personalpsychologie*, *6*, 60–70.

Ziegler, M., Danay, E., Schölmerich, F., & Bühner, M. (2010). Predicting academic success with the Big 5 rated from different points of view: Self-rated, other rated and faked. *European Journal of Personality*, *24*, 341–355.