

Criterion-Related Validity of Dutch Police-Selection Measures and Differences Between Ethnic Groups

Lonneke A. L. de Meijer*, Marise Ph. Born***, Gert Terlouw** and Henk T. van der Molen***

*Erasmus University Rotterdam, Rotterdam, The Netherlands. demeijer@fsw.eur.nl

**The Police Academy of The Netherlands, The Netherlands. gert.terlouw@politieacademie.nl

***Erasmus University Rotterdam, Rotterdam, The Netherlands

This study investigated the criterion-related validity of cognitive ability as well as non-cognitive ability measures and differences between ethnic majority ($N = 2365$) and minority applicants ($N = 682$) in Dutch police officer selection. Findings confirmed the relatively low predictive validity of cognitive ability generally found for police jobs. Previous research reported no differential prediction. The present study, however, found small but systematic evidence for differences in validity for the ethnic majority and minority group of both cognitive and non-cognitive measures. For the minority group, training performance appeared to be mainly predicted by the cognitive ability test. For the majority group, cognitive ability showed very little predictive power. Non-cognitive ability variables appeared to be somewhat more predictive in this group.

1. Introduction

In the domain of personnel selection, differences on psychological measures between ethnic majority and ethnic minority groups have been extensively investigated. Many of these studies focused on cognitive ability (or g), which has been found to be a consistently good predictor of job performance across a variety of occupations (Schmidt & Hunter, 1998, 2004). In particular for more complex job levels, the predictive validity of g is high (Hunter, 1986). Both Kanfer and Ackerman (1989), and Salas and Cannon-Bowers (2001) have shown that cognitive ability also is essential in the training context with respect to workplace

learning. Other researchers have reported a strong effect of g in several large-scale studies in military settings on training performance (Olea & Ree, 1994; Ree, Carretta, & Teachout, 1995; Ree & Earles, 1991). At the same time, several studies (e.g., Goldstein, Zedeck, & Goldstein, 2002; Murphy, 2002; Outtz, 2002) have shown that cognitive ability tests represent the predictor that most likely will have substantial adverse impact on employment opportunities for most ethnic minority groups. Yet, evidence has been found that differences between the ethnic majority and the ethnic minority group in cognitive ability test scores are considerably larger than the differences in measures of job performance (e.g., Roth, Huffcutt, & Bobko, 2003; Waldman & Avolio, 1991).

When employers want to maximize the skill level of their employees on the one hand and diversify their workforce on the other hand, both goals cannot be

An earlier version of this article was presented at the 21st annual conference of the Society for Industrial and Organizational Psychology (SIOP), May 2006, Dallas (TX).

achieved at the same time because of existing subgroup differences on the cognitive ability test. A possible solution for this dilemma has been sought in the use of non-cognitive ability predictors, e.g., non-cognitive dimensions measured with the assessment center (AC) and the employment interview. The AC and the employment interview are instruments that have shown smaller score differences between ethnic groups and, consequently, a lower adverse impact on employment opportunities than the cognitive ability test (De Meijer, Born, Terlouw, and Van der Molen, 2006; Murphy, 2002). This finding has been explained by the non-cognitive dimensions measured with these devices. The aim of the present study is to investigate the predictive power of cognitive and non-cognitive ability dimensions as well as their differential predictive validity in a multi-cultural setting in the context of police training at the Police Academy of the Netherlands.

Non-cognitive ability constructs may especially be useful in predicting police officer job performance. A meta-analysis of European validity studies by Salgado, Anderson, Moscoso, Bertua, De Fruyt, and Rolland (2003) showed several remarkable findings. First, cognitive ability did not predict job performance in police occupations as well as in other occupations. Salgado *et al.* reported a large (corrected) predictive validity of cognitive ability for managerial occupations ($r=.67$; number of studies $k=6$). Nevertheless, for police occupations the (corrected) predictive validity was quite low ($r=.24$; $k=5$) and even lower than for all other occupations in the meta-analysis. Second, for training success the authors also reported the lowest predictive validity of cognitive ability for the police ($r=.25$; $k=3$). Finally, and more in general across jobs, they showed that the predictive validity of cognitive ability was smallest for low complex jobs ($r=.51$) as well as for low complex training ($r=.36$). Other studies (Dayan, Kasten, & Fox, 2002; Hirsh, Northrop, & Schmidt, 1986; Pynes & Bernardin, 1989), not included in Salgado *et al.*'s meta-analysis, found that the (corrected) predictive validity of cognitive ability for law enforcement occupations was relatively low, namely between .10 and .31. Although cognitive ability is likely to be correlated with performance in virtually any job or training, in part because all jobs and trainings for these jobs call for some learning, judgment, and active information processing (Murphy, 2002), Hirsh *et al.* (1986) argued that non-cognitive, behavioral, dimensions, such as interpersonal skills, play a major role in the determination of police officer success. In support of this explanation, Dayan *et al.* (2002) reported that over 50% of the calls to police departments are about dealing with emotional situations, dealing with threatening and violent people, and settling family disputes. In addition, they found that for police performance, cognitive and non-cognitive factors had a comparable amount of predictive power.

In personnel selection, non-cognitive constructs generally are measured by means of an AC exercise, an employment interview, or a personality questionnaire. Although ACs and interviews are measurement methods that in principle can be developed to measure virtually any construct (both cognitive and non-cognitive), in the present study, the focus is on an AC and an employment interview that predominantly measure non-cognitive constructs. Therefore, they will be labeled non-cognitive measures in the remainder of this paper.

Both the AC and the employment interview have shown to have acceptable predictive validity for (police) job performance as well as (police) training success (Dayan *et al.*, 2002; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Pynes & Bernardin, 1989; Robertson & Smith, 2001). However, there is ongoing debate about the predictive power of the personality questionnaire. On the one hand, a large meta-analysis by Barrick, Mount, and Judge (2001) showed that especially Conscientiousness is a valid predictor across jobs. On the other hand, Murphy and Dziewieczynski (2005) and, more recently, Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007a,b) argued that personality inventories almost always turn out to be fairly poor predictors of performance. Murphy and Dziewieczynski (2005) discuss three reasons why the Big Five dimensions of personality seem to have little to do with performance in most jobs. One reason is the often vague theoretical linking between personality constructs and job dimensions. Second, little is known about how to match personality constructs to jobs. Job-analysis methods have, to a large extent, focused on determining abilities and skills that are necessary for successful job performance. It is, however, not clear whether the same methods can be applied to determine which personality constructs make a difference in performing one's job. The third reason they mention for the low predictive validity of personality is that personality-related measures used in organizations have included measures of poorly defined constructs. It is likely that these three reasons apply to training performance as well, as Cortina, Doherty, Schmitt, Kaufman, and Smith (1992) found poor predictive validities of personality inventories for police training success.

In the present study, two goals are pursued. The first goal is to investigate the predictive validity of a cognitive ability test and of several non-cognitive ability selection measures (i.e., a personality questionnaire, an AC, an employment interview, and a final employment recommendation). The strength of the relationship between the cognitive ability test scores and training results will be compared with the relationship between non-cognitive ability measures and training results. The second goal is to examine potential differences in predictive validity of selection measures between the ethnic majority and the ethnic minority group.

As for the first goal, the following three Hypotheses (1a–c) are tested. First, with regard to the personality questionnaire, it is expected – in line with the results of a meta-analysis by Barrick *et al.* (2001) – that only the Big Five factor Conscientiousness will have a comparable predictive power to the cognitive ability test (Hypothesis 1a). It is expected that the other four Big Five factors – in line with Barrick *et al.* (2001), Murphy and Dziewieczynski (2005), and Cortina *et al.* (1992) – will show less predictive validity than the cognitive ability test (Hypothesis 1b). In correspondence with Salgado *et al.*'s (2003) findings on police occupations, it further is expected that the AC, the employment interview, and the final employment recommendation each will have a predictive power that is comparable to that of the cognitive ability test (Hypothesis 1c). Hypotheses 1a through 1c were examined for ethnic majority and ethnic minority trainees, separately.

As for the second goal, we investigate whether the various selection measures will show differential validity. Most research in this area has been conducted in North America (e.g., Hunter, Schmidt, & Hunter, 1979; Rotundo & Sackett, 1999) and has used cognitive ability tests as predictors. The general conclusion from this body of research has been that there is no differential validity. To our knowledge, however, little attention has been given to possible differential prediction of non-cognitive ability measures. North American studies on differential prediction typically concern cognitive test differences between native-born English-speaking ethnic minorities and Whites. While little evidence exists for test bias against US ethnic minorities, Te Nijenhuis and Van der Flier (2000) argued that the US differential-prediction findings cannot be directly generalized to non-native-born, non-native-language-speaking minorities in the Netherlands. For these people, who have a limited knowledge of the language and culture, as is the case for first- or even second-generation ethnic minorities in the Netherlands (Te Nijenhuis & Van der Flier, 2000) and more generally in Europe, these tests may be of limited use and therefore may show limited predictive validity. Te Nijenhuis and Van der Flier (2000) investigated the differential validity of cognitive as well as non-cognitive tests in the Netherlands. On several occasions, they indeed found evidence for differential prediction, especially with performance criteria that had lower cognitive loadings. A possible explanation was sought in the fact that these criteria were subjective evaluations containing potential criterion bias. Criterion bias implies that for ethnic minority members the focus may be on different aspects of performance than for ethnic majority members. A hypothetical example of a situation in which criterion bias could occur is when training performance of ethnic minority trainees is attributed to their decision-making skills while training performance

of ethnic majority trainees is attributed to their social skills. In the present study at the Dutch police, supervisors' subjective ratings are used as training evaluations. Therefore, it is possible that criterion bias plays a role at the Dutch police as well. In correspondence with findings of Te Nijenhuis and Van der Flier (2000), it is therefore expected that differences in predictive validity between the ethnic majority and the ethnic minority group will exist both on cognitive ability and on non-cognitive ability tests (Hypothesis 2).

2. Method

2.1. Participants

Data came from a sample of trainees ($N = 3117$; 66% male; $M_{\text{age}} = 23.75$, standard deviation [SD] = 5.97), who had been admitted to the police officer training. Data were gathered from September 2001 to January 2006. The sample contains ethnic majority group members ($N = 2365$; 65% male; $M_{\text{age}} = 23.68$, $SD = 6.10$), and first- as well as second-generation ethnic minority group members in the Netherlands ($N = 682$; 67% male; $M_{\text{age}} = 24.05$, $SD = 5.44$). First-generation ethnic minority members are born outside the Netherlands. Second-generation minority group members, in contrast to first-generation minorities, are born in the Netherlands while at least one of their parents is born outside the Netherlands. The largest ethnic minority groups in the Netherlands are the Antillean, Moroccan, Surinamese, and Turkish groups, which are equally represented in our minority sample. Of 70 trainees (2%), it was not known if they belonged to the ethnic majority or ethnic minority group. They were excluded from further analyses. The study had a longitudinal design covering about 12 months. Criteria were gathered from the police officer training about 1 year after the selection procedure and include evaluations of workplace performance on typical police tasks, namely: Maintaining Order and Helping Victims.

2.2. Selection and training at the police academy of the Netherlands

Applicants who are interested in a job as police officer first apply to the local police force where they want to work after they will complete their training. For the selection procedure, the local police forces routinely send all applicants to the national police Center for Competence Measurement and Monitoring (CCM). During a requirement check at the CCM, the following minimal criteria are checked on the basis of an application form: Minimal age (16 years), Dutch nationality, possession of a swimming diploma, no criminal record, and possession of a school diploma (minimal level is

preparatory vocational education level B [VBO-B]). Applicants in the selection process go through two stages. During the first stage a Dutch language-proficiency test is filled in. During the second phase a physical exercise, a cognitive ability test, a personality questionnaire, an AC assignment and an employment interview are executed. The psychologist who conducts the interview is also the one who writes the final employment recommendation to the police force. For the employment recommendation, the test results of the personality questionnaire, the AC ratings, and the employment interview ratings are used. Next to the final recommendation, the final dossier to the local police forces consists of test scores of the physical exercise, the cognitive ability test, and the language-proficiency test. On the basis of the information from the CCM, the local police force decides whether to accept or reject.

The professions for which accepted students are to be trained for are assistant police employee (2-year training), police employee (3-year training), or all-round police employee (4-year training). The training on these three levels is organized in the same way, i.e., 3 months of theoretical training is alternated with 3 months of on-the-job training. The theoretical knowledge gained during the first 3 months has to be put into practice during the later 3 months. Each 6 months are rounded off with an examination of on-the-job performance. The three training levels differ in responsibility: The more advanced a trainee is, the more responsibility (s)he will have. All trainees who finish the training will get a job as (assistant/all-round) police employee. We will now present a more detailed description of the selection measures and the criteria used.

2.2.1. Cognitive ability test

The Police Intelligence Test (PIT; Rijks Psychologische Dienst, 1975) is a cognitive ability test and consists of 107 items divided over six subtests: Verbal Comprehension, Picture Arrangement, Numerical Reasoning, Word Fluency, Spatial Ability, and Inductive Reasoning. The time limit is 51 min. Applicants completed the PIT in Dutch. Prior research by Lem and Van Doorn (2000) indicated α reliabilities varying from .69 to .87. The correlations between the subscales varied from .32 to .57. A study by Van der Maesen (1992) showed corrected predictive validity coefficients of .39 and .46 ($N = 162$).

2.2.2. Personality questionnaire

To measure the Big Five factors Extraversion, Altruism, Conscientiousness, Emotional Stability, and Intellect, the Police Personality Questionnaire (PPV; Van Leeuwen, 2000) was used. The applicants completed the PPV in Dutch. A recent progress report by Klinkenberg and Van Leeuwen (2003) indicated α reliabilities varying from .72 to .78. Correlations between the scales are all lower than .60. Comparison with NEO-PI-R showed

observed construct validity coefficients between .17 and .58 ($N = 160$). A study by Lem and Van Doorn (2000) showed observed predictive validity coefficients between .15 and .43 ($N = 61$).

2.2.3. AC

A role-play exercise is utilized, in which an assessor and an actor independently make ratings on a seven-point Likert-scale ranging from 1 (extremely weak) to 7 (excellent), on each of the following seven dimensions: Communication Skills, Social Skills, Empathy, Initiative, Stress Tolerance, Authority, and Decisiveness. Interrater reliabilities ranged from .82 to .88 ($N = 198$). Principal component analysis with varimax rotation yielded two factors, Agency and Communion (in accordance with Wiggins and Trapnell, 1996), which together explained 77% of the variance. As a measure of Agency, the average rating across the dimensions of Authority, Decisiveness, Initiative, Communication Skills, and Stress Tolerance was used ($\bar{r} = .59$; $\alpha = .87$). As a measure of Communion, the average rating of the dimensions Social Skills and Empathy was used ($\bar{r} = .77$; $\alpha = .87$). The reliability of the difference (r_{diff}) between scores on Agency and Communion was .78.

2.2.4. Employment interview

The interview questions are focused on evaluating behavior on the following eight dimensions: Communication Skills, Social Skills, Flexibility, Stress Tolerance, Emotional Stability, Tolerance Towards Others, Integrity, and Self-Understanding. A single interviewer conducts the interview. The interviews are semi-structured and behaviorally based, with one behaviorally anchored seven-point Likert scale ranging from 1 (extremely weak) to 7 (excellent) for each of the eight dimensions. The average rating across the eight dimensions was used as the dependent variable because the ratings were substantially correlated ($\bar{r} = .42$; $\alpha = .85$). Moreover, principal component analysis with varimax rotation yielded one interview factor that explained 50% of the variance.

2.2.5. Final employment recommendation

The final recommendation as to whether an applicant is fit for a job as police officer is based on results from the personality questionnaire (PPV), the AC, and the employment interview. These scores are integrated into an employment recommendation. The dimensions in the final recommendation are: Communication Skills, Social Skills, Empathy, Initiative, Flexibility, Stress Tolerance, Authority, Decisiveness, Tolerance Towards Others, Integrity, and Self-Understanding (for definitions, see De Meijer, Van Zielst, and Van der Molen, 2007). A seven-point Likert scale ranging from 1 (extremely weak) to 7 (excellent) is used to evaluate the behavior on the 11 dimensions. Principal component analysis

with varimax rotation yielded three employment-recommendation factors, Agency, Communion, and Socio-Cultural Awareness, which altogether explained 67% of the variance. As a measure of Agency, the average rating across the dimensions Authority, Decisiveness, Initiative, Communication Skills, Stress Tolerance, and Flexibility was used ($\bar{r} = .48$; $\alpha = .85$). As a measure of Communion, the dimensions Social Skills and Empathy, were used ($\bar{r} = .66$; $\alpha = .79$) and for Socio-Cultural Awareness, the dimensions Tolerance Towards Others, Integrity, and Self-Understanding ($\bar{r} = .39$; $\alpha = .65$). The reliability of the difference (r_{diff}) between scores on Agency and Communion is .51, r_{diff} between scores on Agency and Socio-Cultural Awareness is .58, and r_{diff} between scores on Communion and Socio-Cultural Awareness is .57.

2.2.6. Criteria: Training results

Supervisors were asked to rate trainees as satisfactory (1) or unsatisfactory (0) on a number of items per examination, which measured actual police work concerning 'Maintaining Order' (i.e., providing for public safety by Maintaining Order, responding to emergencies, protecting people and property, enforcing criminal laws, and identifying, pursuing, and arresting suspects and perpetrators of criminal acts [O*Net Online, 2007, January 31]) and 'Helping Victims' (i.e., rendering aid to accident victims and other persons requiring first aid for physical injuries [O*Net Online, January 31, 2007]). Per examination, one single supervisor observed and, subsequently, evaluated each trainee. Supervisors rated trainees' practical skills in actual police situations with actual civilians. Supervisors were trained to evaluate police trainees. All supervisors belonged to the ethnic majority group.

Each examination involved an evaluation on a number of items, among which a subset of so-called critical items. The critical items each had to be rated as being satisfactory in order to pass the examination and are descriptions of most effective behavior in a given situation. Next to the critical items, a number of remaining items as a whole had to be satisfactory scored in order to pass the examination. These focused on required daily routines. Maintaining Order (13 items) had six critical items on each of which the trainee should receive a satisfactory score (examples are: 'works safely', 'gives information correctly', and 'displays authority appropriately'). Of the remaining seven items, a minimum of five items had to be satisfactory scored (an example is: 'checks a person's identity'). For Helping Victims (13 items), three items were critical (examples are: 'finds out what someone's problem is' and 'gives emotional support'). Of the remaining 10 items, a number of seven items had to be rated as being satisfactory (an example is: 'ends the

conversation properly'). If these requirements were not met, the trainee had to sit a re-examination.

Maintaining Order (13 items; $\alpha = .47^1$) and Helping Victims (13 items; $\alpha = .58^1$) were chosen among a series of examinations because they are two of the most important aspects of police work (cf. O*Net Online, January 31, 2007). The items of the two examinations were averaged for each separate examination. The correlation between the average scores on Maintaining Order and Helping Victims was .04 (NS). The 26 item-ratings were also combined into an overall training score ($\alpha = .54^1$).

2.3. Analyses

In order to conduct correlational analysis, Structural Equation Modeling (SEM) with Amos 6.0 (Arbuckle, 2005) was used to investigate the relationships between selection measures and training criteria. Differences in correlations with regard to the ethnic majority versus ethnic minority group were tested via multi-group analyses. Furthermore, hierarchical linear regression analysis was conducted, in which scores on a certain selection measure and ethnic group membership were entered, as variables, in the first step and the interaction term in the second step. In this manner, differences between the ethnic majority and minority group in regression equations are examined. One important problem of taking ethnic group membership as part of an interaction term (group membership then becomes a moderator) into a regression equation, is that group sizes should be about the same in order to have adequate statistical power (Aguinis & Stone-Romero, 1997). In our sample, the ethnic majority group ($N = 2365$) was much larger than the ethnic minority group ($N = 682$). Therefore, we decided to conduct the regression analyses with roughly the same group sizes. A random sample of 700 ethnic majority trainees was drawn from our original sample (SPSS 14.0, 2005), which we then compared to the 682 ethnic minority trainees.

3. Results

3.1. Preliminary results

Table 1 reports the α reliabilities, means, and SDs of the selection measures and the criteria and the correlations among the selection measures for the ethnic majority and minority groups, separately. All selection measures had good reliabilities. Therefore, the correlations between the selection measures were not corrected for attenuation (the correlations between the selection measures and the training criteria and among the training criteria were).

Table 1. Descriptives of selection measures and training criteria

Selection measures and criteria	Cronbach's α		Mean (SD)		Correlation coefficients																				
	Ethnic majority	Ethnic minority	Ethnic majority	Ethnic minority	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
<i>Cognitive ability measure:</i>																									
PIT																									
1. Verbal Comprehension	.62	.68	10.25 (2.25)	9.53 (2.45)	—	.49**	.45**	.48**	.39**	.41**	.14**	.04	-.13*	.02	.20**	.26**	.17**	.23**	.26**	.17**	.18**	.28**	.20**	.24**	.10
2. Inductive Reasoning	.70	.75	11.03 (2.30)	10.41 (2.45)	.43**	—	.59**	.37**	.58**	.49**	.11*	.06	-.06	.09*	.16**	.27**	.18**	.18**	.26**	.16**	.11*	.21*	.22*	.22*	.10
3. Numerical Reasoning	.82	.80	8.29 (3.28)	7.39 (3.33)	.43**	.52**	—	.39**	.46**	.38**	.06	.04	-.05	.08*	.11*	.18**	.12*	.12*	.17**	.12*	.09*	.12*	.11	.09	.09
4. Word Fluency	.81	.79	9.74 (3.54)	8.75 (3.44)	.40**	.34**	.41**	—	.34**	.37**	.17**	.05	-.11*	.11*	.22**	.18**	.09*	.15**	.18**	.09*	.10*	.23**	.18*	.20**	.08
5. Spatial Ability	.81	.83	2.87 (4.08)	19.68 (4.22)	.34**	.53**	.38**	.28**	—	.39**	.11*	.10*	.03	.09*	.17**	.13*	.10*	.06	.11*	.08*	.03	.05	.04	.05	.05
6. Picture Arrangement	.64	.75	8.32 (1.60)	8.05 (1.74)	.28**	.36**	.25**	.26**	.29**	—	.09*	.05	-.09*	.07	.19**	.20**	.13*	.19**	.21**	.13*	.16**	.06	-.05	.19**	.05
<i>Non-cognitive ability measures:</i>																									
PPV																									
7. Extraversion	.76	.71	6.30 (2.67)	6.30 (2.76)	.02	-.01	-.02	.01	-.01	.01	—	.58**	.33**	.38**	.53**	.17**	.13*	.24**	.22*	.19**	.12*	.05	.12	-.06	-.06
8. Altruism	.75	.75	5.84 (2.74)	5.98 (2.85)	-.03	-.02	-.05*	-.03	.01	-.01	.56**	—	.44**	.35**	.54**	.05	.06	.13*	.08*	.13*	.12*	.05	.13	.03	.03
9. Conscientiousness	.70	.71	5.56 (2.91)	5.90 (2.85)	-.09**	-.07*	-.07*	-.08*	-.02	-.06*	.26**	.37**	—	.29**	.40**	-.08*	-.04	-.07	-.08*	-.04	-.06	.08	.10	.00	.00
10. Emotional Stability	.74	.74	5.91 (2.74)	6.09 (2.82)	-.03	.00	-.01	.02	.02	.02	.34**	.34**	.29**	—	.40**	.07	.05	.13*	.11*	.09**	.01	-.02	.04	-.12	-.12
11. Intellect	.77	.75	6.07 (2.75)	5.96 (2.89)	.08*	.05*	.01	.07*	.06*	.06*	.49**	.51**	.42**	.41**	—	.21**	.11*	.24**	.27**	.14**	.13*	-.05	.11	-.13	-.13
AC																									
12. Agency	.87	.86	4.69 (.83)	4.51 (.84)	.13**	.10**	.08*	.12**	.05*	.14**	.12**	.03	.01	.08*	.17**	—	.62**	.58**	.91**	.59**	.42**	-.01	-.08	.07	.07
13. Communion	.87	.87	4.52 (1.03)	4.36 (1.06)	.09**	.08*	.04*	.09**	.05*	.10**	.09**	.04*	.00	.05*	.09**	.63**	—	.41**	.56**	.89**	.33**	.02	-.08	.08	.08
14. Employment interview	.85	.85	4.65 (.52)	4.56 (.55)	.13**	.08*	.05*	.09**	.03	.12**	.24**	.14**	.04*	.18**	.24**	.50**	.38**	—	.75**	.56**	.82**	.02	-.06	.09	.09
<i>Final recommendation</i>																									
15. Agency	.85	.84	4.66 (.67)	4.49 (.69)	.14**	.11**	.08*	.12**	.05*	.14**	.20**	.08*	.04*	.16**	.25**	.90**	.56**	.72**	—	.60**	.52**	.02	-.06	.09	.09
16. Communion	.78	.79	4.60 (.83)	4.48 (.82)	.09**	.06*	.03	.07*	.03	.10**	.18**	.11**	.02	.08*	.14**	.60**	.89**	.56**	.62**	—	.47**	-.04	-.15	.09	.09
17. Socio-Cultural Awareness	.64	.67	4.52 (.51)	4.48 (.56)	.11**	.05*	.02	.07*	.01	.10**	.13**	.12**	.03	.08*	.16**	.36**	.32**	.80**	.50**	.48**	—	.01	-.05	.02	.02
<i>Training criteria:</i>																									
18. Overall training score	.54	.53	.91 (.13)	.91 (.12)	.04	.04	.05	.06	.00	.11*	.05	.03	.03	-.08	.03	.13**	.11*	.16**	.17**	.17**	.17**	—	1.00**	1.00**	1.00**
19. Maintaining Order	.47	.53	.91 (.14)	.91 (.15)	.02	.02	.02	.00	-.04	.07	.09*	.10*	.03	-.04	.05	-.02	.06	.04	.02	.06	.02	1.00**	—	.16*	.16*
20. Helping Victims	.58	.52	.90 (.15)	.91 (.14)	-.04	.03	.06	.06	.00	.07	.02	-.03	-.05	-.04	.03	.21**	.12*	.17**	.21**	.18**	.19**	1.00**	.05	—	—

Note. PIT scores varied between 1 and 15 on Verbal Comprehension and Inductive Reasoning, between 1 and 13 on Numerical Reasoning, between 1 and 25 on Word Fluency, between 1 and 29 on Spatial Ability, and between 1 and 10 on Picture Arrangement; PPV scores varied between 1 and 10; AC, interview, and final-recommendation scores varied between 1 and 7; and training scores varied between 0 and 1. Positive d values indicate the ethnic majority group scoring higher. The correlations of the ethnic minority group are above the diagonal. Correlations between selection measures are corrected for range restriction. Correlations between selection measures and criteria are corrected for range restriction and attenuation. Significant differences between ethnic groups of correlations between selection measures and criteria are in *italic* ($p < .05$). Correlations between training criteria are corrected for attenuation. * $p < .05$, ** $p < .001$.

3.2. Main results

In investigating the predictive validities of all selection measures, the predictive power of the non-cognitive ability measures were compared with the predictive power of the cognitive ability test (Hypotheses 1a–c). Simultaneously, it was examined whether differences existed between predictive validities for the ethnic majority versus ethnic minority group (Hypothesis 2). Correlations between predictor scores and criterion scores were generated by means of SEM (see Figure 1). SEM enabled the investigation of differential prediction by means of multi-group analysis. The fit indices of the models for the three criteria are shown in Table 2. The models showed a good fit.

Table 1 shows SEM results and presents the correlations between predictors and criteria. Only the correlations corrected for direct range restriction and attenuation (for the formulae see Bobko, Roth, & Bobko, 2001) are shown (for the uncorrected correlations, the first author may be contacted). Furthermore, significant correlation differences between the ethnic majority and minority group are marked. For reasons of clarity and conciseness, an overview of the most remarkable results will be described here. First, a comparison in predictive validity will be made between the cognitive ability test and the personality questionnaire. Second, the predictive validity of the cognitive

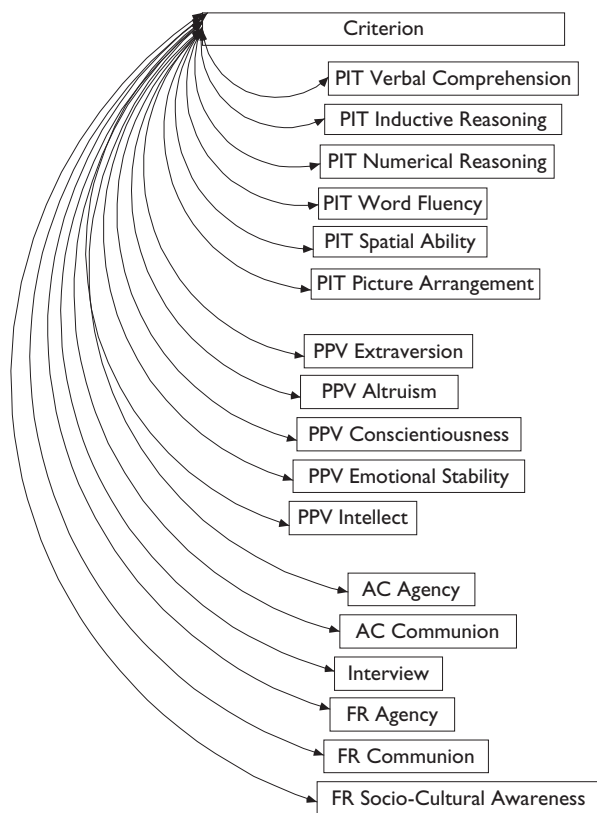


Figure 1. Model for correlational analysis.

Table 2. Correlational analysis: fit indices for training criteria

Criteria	χ^2	df	χ^2/df	TLI	CFI	RMSEA
Overall training score	764.32**	192	3.98	.96	.98	.03
Maintaining Order	764.51**	192	3.98	.96	.98	.03
Helping Victims	764.22**	192	3.98	.96	.98	.03

Note. TLI, Tucker–Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation. ** $p < .001$ (one-tailed).

ability test will be compared with the predictive validity of the AC, the employment interview, and the final recommendation, separately.

Regarding the personality questionnaire, we expected that Conscientiousness would have a predictive power comparable with the cognitive ability test (Hypothesis 1a). Furthermore, the other Big Five factors, namely Extraversion, Altruism, Emotional Stability, and Intellect, were expected to show less predictive power than the cognitive ability test (Hypothesis 1b). Hypothesis 2 predicted that differences in predictive validities between the ethnic majority and the ethnic minority group would exist on the cognitive ability test and on the personality questionnaire. The results in Table 1 show support for Hypothesis 1b, but not for Hypothesis 1a. No support was found for Hypothesis 2 on the personality questionnaire, but support for Hypothesis 2 was found on the cognitive ability test.

All five personality factors showed very little predictive validity. Conducting multi-group analyses, a significant difference in predictive validity between the ethnic majority and minority group was found only for Intellect predicting the training score of Helping Victims ($r_{maj.} = .03$, NS and $r_{min.} = -.13$, NS, respectively). No evidence for differential prediction was found for the other Big Five factors on any of the criteria.

The predictive validity of the cognitive ability test was higher than the above-mentioned predictive validity of the personality questionnaire, especially for the ethnic minority group. More specifically, the verbal subtests of the cognitive ability test (i.e., Verbal Comprehension and Word Fluency) were most predictive of training success for the ethnic minority group compared with the ethnic majority group. Significant differences in validity between the ethnic groups were found for several cognitive ability subtests, namely Verbal Comprehension, Inductive Reasoning, and Word Fluency for the prediction of the overall training score. For training results on Maintaining Order, differences in validity were found for the sub-tests Verbal Comprehension, Inductive Reasoning, Word Fluency, and Picture Arrangement. Finally, for the training scores on Helping Victims, different validity coefficients were found for the sub-tests Verbal Comprehension, Word Fluency, and Picture Arrangement. No differences in prediction were found for the sub-tests Numerical Reasoning and Spatial Ability.

Table 3. Hierarchical regression analyses of the selection measures predicting the training scores

Criteria	Selection measures									
	Cognitive ability test		Personality questionnaire		AC		Employment interview		Final recommendation	
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2
<i>Overall training score:</i>										
Selection measure	.07*	.19*	.00	.02	.08*	-.11	.08*	-.06	.09*	-.12
Ethnic group membership	-.01	.23	.01	.03	.00	-.34*	.00	-.38	.00	-.54*
Interaction		-.28		-.03		.41*		.42†		.60*
ΔR^2		.00		.00		.01*		.00†		.00*
<i>Maintaining Order:</i>										
Selection measure	.05	.15†	.04	.13	.01	-.11	.01	-.04	.01	-.11
Ethnic group membership	.02	.23	.03	.12	.03	-.20	.03	-.13	.03	-.31
Interaction		-.25		-.13		.28		.18		.37
ΔR^2		.00		.00		.00		.00		.00
<i>Helping Victims:</i>										
Selection measure	.06*	.23*	-.03	-.09	.12**	-.06	.10*	.00	.11**	-.05
Ethnic group membership	-.03	.30†	-.02	-.08	-.04	-.37*	-.03	-.32	-.03	-.48†
Interaction		-.40†		.09		.40*		.32		.49†
ΔR^2		.00†		.00		.00*		.00		.00†

Note. For these analyses, roughly equal group sizes were used (i.e., $N = 700$ for the ethnic majority group; $N = 682$ for the ethnic minority group). Standardized regression weights are presented. The crosses and asterisks correspond to the unstandardized regression weights. Ethnic group membership is coded as follows: 1 = 'ethnic minority group'; 2 = 'ethnic majority group.' † $p < .10$, * $p < .05$, ** $p < .001$.

Hypothesis 1c predicted that the AC, the employment interview, and the final employment recommendation would have a predictive power comparable to the cognitive ability test. Hypothesis 2 predicted that differences in predictive validity between the ethnic majority and the ethnic minority group would exist on the AC, the interview, and the employment recommendation. The results in Table 1 show partial support for Hypothesis 1c and support for Hypothesis 2. The predictive power of the AC, the interview, and the final employment recommendation was larger than the predictive power of the cognitive ability test, but only for the ethnic majority group. For the ethnic minority group, the non-cognitive ability tests showed very small predictive validities for the overall training score as well as for training scores on Maintaining Order and Helping Victims.

Regarding the differences in predictive-validity coefficients between the ethnic groups (Hypothesis 2), the AC, the interview, and the final recommendation showed larger predictive validities for the ethnic majority group than for the ethnic minority group. The results in Table 1 showed differential validity for the overall training score of both AC-factors, the interview, and all final-recommendation factors. For training results on Maintaining Order, differences in validity coefficients were found of the AC-factor Communion, the interview, and the final-recommendation factor Communion. The other AC and final-recommendation factors showed no differential validity for Maintaining Order. For the training results on Helping Victims, again, several selection factors showed differences in validity. Differential validity was found of the AC-factor

Agency and on the final-recommendation factors Agency and Socio-Cultural Awareness. The interview and the AC- and final-recommendation factor Communion showed no differences for Helping Victims.

Concerning Hypothesis 2, hierarchical regression analysis was conducted in addition to correlational analyses. Hierarchical regression is an often-used technique to examine differential validity. Scores on a certain selection measure and group membership were entered, as variables, in the first step of the regression. The interaction between both was entered in the second step. A significant interaction effects shows evidence for differential validity. The results are shown in Table 3 (results on sub-test or sub-dimension level are not shown in Table 3, but are only described in the text). Although the incrementally explained variances of the interaction terms are very small or close to zero, significant interaction effects were found for the cognitive ability test, the AC, the employment interview, and the final recommendation. These results, thus, point to the existence of differential validity of both the cognitive ability test and the non-cognitive ability measures (Hypothesis 2).

The regression of the overall training score on the AC (for an illustration, see Figure 2), the employment interview, and the final recommendation differed for the two ethnic groups, with an only marginal difference for the employment interview. Although the regression of the overall training score on the cognitive ability test in general did not show differences, the regression on the sub-tests Verbal Comprehension [$\beta = -.25$, $p = .10$ (marginally)] and Word Fluency ($\beta = -.27$, $p < .05$) did.

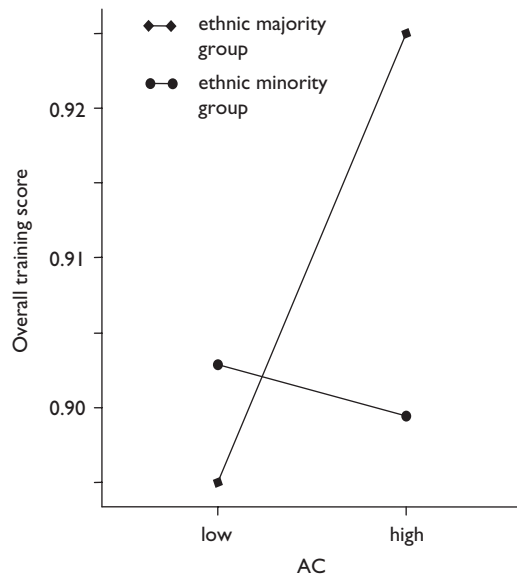


Figure 2. Illustration of ethnic majority and minority AC scores predicting the overall training score.

The regression of Maintaining Order in general did not show differences between the ethnic majority and minority group. However, the regression on the cognitive ability sub-test Word Fluency ($\beta = -.35, p < .05$) did, as well as on the AC- and final-recommendation factor Communion [$\beta = .29, p < .10$ (marginally) and $\beta = .43, p < .05$, respectively]. The regression of Helping Victims on the cognitive ability test (marginally), the AC, and the final recommendation (marginally) differed for the two ethnic groups. Especially, the regression on the cognitive ability sub-test Verbal Comprehension ($\beta = -.34, p < .05$) appeared to be different for the two groups.

Although the effect sizes of differential validity generally are small, the following trend is discernible: The cognitive ability test, especially the verbal sub-tests, appears to show more predictive power for the ethnic minority group than for the ethnic majority group. Contrarily, the AC, the employment interview, and the final employment recommendation appear to show more predictive power for the ethnic majority group than for the ethnic minority group. The personality questionnaire showed very little predictive power for either group.

4. Discussion

As a first goal, the criterion-related validity of both cognitive and non-cognitive ability selection measures was investigated for training performance of police trainees. Second, differential prediction between ethnic groups of both cognitive and non-cognitive ability measures was examined.

When score differences between ethnic groups on a cognitive ability test are larger than score differences in

job or training performance, potentially good employees or trainees could be rejected during selection. A potential problem is a lack of ethnic diversity or heterogeneity in one's workforce. Especially the latter issue is of concern for organizations in the public domain such as the police, since contact with different ethnic groups in society forms an important aspect of the job as police officer. When non-cognitive ability measures are available that show less score differences between ethnic groups than on the cognitive ability test, as in the present study, and that show comparable predictive validities to the cognitive ability test, this could be a solution. The predictive validities of the non-cognitive ability measures and the differential prediction that were found will be discussed below.

Confirming the viewpoint of Murphy and Dziewieczynski (2005) and Morgeson *et al.* (2007a, b), the Big Five personality questionnaire showed almost no predictive power. Cortina *et al.* (1992) found similar results. They used a sample of police recruits and found poor predictive validities of two personality inventories for police training performance. Cortina *et al.* argued that the questionnaires they used were not useful for the police selection, because the tests were not developed specifically for the police. Although the personality questionnaire in the present study was adapted for the Dutch police, it is recommended to further investigate whether the constructs that are measured with the PPV indeed are important for police training performance. In line with suggestions by Hatstrup, Rock, and Scalia (1997), it can also be argued that the Big Five personality constructs do not predict the scores on the specific police-relevant criteria that were used in the present study. They may, however, be useful to predict more general training performance, e.g., teamwork, friendliness, and punctuality. Unfortunately, these criteria could not be directly investigated in the present study.

The corrected predictive validity of cognitive ability for training performance as reported in the literature by Schmidt and Hunter (1998, 2004) is high ($r = .59$). However, Salgado *et al.* (2003) reported much lower (corrected) predictive power for cognitive ability tests in police occupations, namely .24. The present study found even lower (corrected) predictive validities of cognitive ability for training performance than Salgado *et al.* (2003) did, namely .04 ($-.04 < r_{\text{maj.}} < .11$) for the ethnic majority group and .14 ($-.05 < r_{\text{min.}} < .28$) for the ethnic minority group. One possible explanation for the relatively low validities of cognitive ability tests lies in the potential role of non-cognitive factors in the determination of performance in police work as stated by, e.g., Hirsh *et al.* (1986). Interestingly, however, in the present study this explanation will pertain more to the ethnic majority group than to the ethnic minority group. For the majority group various factors measured

during the AC, the interview, and the final employment recommendation, i.e., Agency, Communion, and Socio-Cultural Awareness, were more predictive than cognitive ability for several training criteria. Especially the Agency factor of the AC and the final recommendation appeared to be predictive for the ethnic majority group. For the minority group, the cognitive ability test was most predictive, especially the verbal cognitive ability subtests, i.e., Verbal Comprehension and Word Fluency. The non-cognitive ability tests showed very little predictive power for the minority group.

The results of the hierarchical regression analyses also point to differential validity for all selection measures except for the personality questionnaire. Training performance appeared to be somewhat better predicted by several cognitive ability subtests for the ethnic minority group, and somewhat better predicted by the non-cognitive ability tests for the ethnic majority trainees. Where differences in predictive validity were found, these might have been caused by ethnic bias of ethnic majority supervisors' subjective evaluations (Te Nijenhuis & Van der Flier, 2000), even though evaluations of trainees during the Dutch police training were structured according to evaluation forms. For ethnic majority trainees, relatively more attention may have been given to the non-cognitive ability aspects of performance, i.e., social skills, decisiveness, and authority, measured with the AC, the interview, and the final recommendation. While for ethnic minority trainees, relatively more attention may have been given to the verbal cognitive ability aspects of performance. The question remains whether supervisors' evaluations of ethnic minority trainees are predominantly susceptible to these quite basic language skills to the extent that these skills will overshadow other important non-cognitive factors, such as social skills and decisiveness. To better understand potential supervisors' susceptibility to ethnic bias, research using ethnic majority as well as ethnic minority supervisors should get more attention in the future.

4.1. Limitations

Although the total sample of ethnic minority trainees was very acceptable ($N = 682$), a first limitation of the present study was that this sample was too small to differentiate among ethnic minority groups. Treating ethnic minorities as a homogeneous group that merely contrasts with the ethnic majority group ignores the many visible and cultural differences among ethnic minority groups that may affect score differences, predictive validity coefficients, and differential prediction. In the present study, we extended previous studies by examining the predictive validity of a cognitive ability test as well as several non-cognitive selection measures

in a multi-cultural setting. Since we found differential prediction on all measures, future research should investigate this differential prediction for the various ethnic groups that exist in the Netherlands and, more broadly, in multicultural societies, also for other sets of tests.

Second, although correlations between the cognitive ability test on the one hand and the AC, the employment interview, and the final recommendation on the other hand were all below .14 ($\bar{r} = .08$) for the ethnic majority group and were all below .27 ($\bar{r} = .15$) for the ethnic minority group, there appears to be some overlap between cognitive ability and the non-cognitive constructs. This might slightly contaminate the predictive validities found in the present study. In general, it is to be expected that scores on a cognitive ability test are correlated with AC and interview scores, in general, because performance on an AC or an interview to some extent requires cognitive skills such as active information processing and adequate responding (cf. Murphy, 2002).

A third limitation of the present study was that the sizes of the predictive validities were quite small. An explanation for this finding may be found in the low variance in training scores (see criteria-SDs in Table 1). As a result of this low variance, the correlations and regression weights presented in this study may be somewhat underestimated. On the one hand, low criterion variance may be a valid explanation for possibly underestimated predictive validities. On the other hand, previous research has also found relatively small (corrected) predictive validities of the cognitive ability test and the personality questionnaire for low-level police training and work performance (Cortina et al., 1992; Salgado et al., 2003). As the issue of relatively low criterion variance is a general problem encountered in operational criteria (Cascio & Aguinis, 2005), we believe that the predictive validities of the cognitive and non-cognitive selection measures for low-level police training as found in the present study are not less accurate estimates than other reported findings in the literature. Moreover, in our opinion, the relatively small validities are informative, firstly, because they are systematic. The predictive validities to our view also are informative, since the goals of the present study were aimed at investigating the *differences* in predictive power of various selection measures as well as the *differential* prediction of these measures for different ethnic groups.

A final limitation is the use of training performance as a criterion instead of job performance. The question rises whether training performance can be generalized to work performance, as predicting work performance is the ultimate goal of personnel selection. Using training performance as a performance criterion may be deficient because the goal of personnel selection is to select potentially good employees, not necessarily

good trainees. In a study conducted by Salgado *et al.* (2003), however, the validity of cognitive ability when predicting police training performance was almost equal to the validity when predicting police job performance. The issue of generalizability then concerns the non-cognitive predictors and the potential difference in predicting training performance versus job performance. As the present study has used an on-the-job performance criterion, it is expected that the validity coefficients for the training-performance criterion can largely be generalized to job performance.

5. Conclusion

The predictive validities of the various selection measures are roughly in line with previous research. Regarding differential prediction between ethnic groups we found somewhat different results compared with previous, mostly North American, studies. The effect sizes with regard to differential validity are small but systematic. The cognitive ability test appeared to show somewhat larger predictive-validity coefficients for the ethnic minority group than for the ethnic majority group. The non-cognitive ability measures appeared to show somewhat larger predictive validities for the ethnic majority group than for the minority group. These results may imply that it is important to use both cognitive ability and non-cognitive ability tests for the selection of police officers in order to obtain a diverse ethnic work environment.

Note

1. We acknowledge that the internal consistency (Cronbach's α 's) of the criteria is quite low. However, this is a common phenomenon when different items of a certain measure are behaviorally based and do not measure an underlying construct (e.g., Motowidlo, Dunnette, & Carter, 1990). The criteria used in the present study are multi-dimensional and they measure behaviors that are related to a certain field of police work (e.g., 'maintaining order' or 'helping victims'). Test-retest estimates might be more appropriate, but they were not available.

References

- Aguinis, H. and Stone-Romero, E.F. (1997) Methodological Artifacts in Moderated Multiple Regression and their Effects on Statistical Power. *Journal of Applied Psychology*, **82**, 1, 192–206.
- Arbuckle, J.L. (2005) *Amos 6.0*. [Computer Software]. Chicago: SPSS Inc.
- Barrick, M.R., Mount, M.K. and Judge, T. (2001) Personality and Performance at the Beginning of the New Millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, **9**, 1/2, 9–30.
- Bobko, P., Roth, P.L. and Bobko, P. (2001) Correcting the Effect Size of d for Range Restriction and Unreliability. *Organizational Research Methods*, **4**, 1, 46–61.
- Cascio, W.F. and Aguinis, H. (2005) *Applied Psychology in Human Resource Management* (6th edn). Upper Saddle River, NJ: Prentice Hall.
- Cortina, J.M., Doherty, M.L., Schmitt, N., Kaufman, G. and Smith, R.G. (1992) The "Big Five" Personality Factors in the IPI and MMPI: Predictors of police performance. *Personnel Psychology*, **45**, 1, 119–140.
- Dayan, K., Kasten, R. and Fox, S. (2002) Entry-Level Police Candidate Assessment Center: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, **55**, 4, 827–849.
- De Meijer, L.A.L., Born, M.Ph., Terlouw, G. and Van der Molen, H.T. (2006) Applicant and Method Factors Related to Ethnic Score Differences in Personnel Selection: A study at the Dutch police. *Human Performance*, **19**, 3, 219–251.
- De Meijer, L.A.L., Born, M.Ph., Van Zielst, J. and Van der Molen, H.T. (2007) Analyzing Judgments of Ethnically Diverse Applicants during Personnel Selection: A study at the Dutch police. *International Journal of Selection and Assessment*, **15**, 2, 139–152.
- Goldstein, H.W., Zedeck, S. and Goldstein, I.L. (2002) G: Is this your final answer? *Human Performance*, **15**, 1/2, 123–142.
- Hattrup, K., Rock, J. and Scalia, Ch. (1997) The Effects of Varying Conceptualizations of Job Performance on Adverse Impact, Minority Hiring, and Predicted Performance. *Journal of Applied Psychology*, **82**, 5, 656–664.
- Hirsh, H.R., Northrop, L.C. and Schmidt, F.L. (1986) Validity Generalization Results for Law Enforcement Occupations. *Personnel Psychology*, **39**, 2, 399–420.
- Hunter, J.E. (1986) Cognitive Ability, Cognitive Aptitude, Job Knowledge, and Job Performance. *Journal of Vocational Behavior*, **29**, 3, 340–362.
- Hunter, J.E., Schmidt, F.L. and Hunter, R. (1979) Differential Validity of Employment Tests by Race: A comprehensive review and analysis. *Psychological Bulletin*, **86**, 4, 721–736.
- Kanfer, R. and Ackerman, P.L. (1989) Motivation and Cognitive Abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, **74**, 4, 657–690.
- Klinkenberg, E.L. and Van Leeuwen, A.E. (2003) *Voortgangsverslag Ontwikkeling M5Q-IWSP* ([Progress report development M5Q-IWSP]. Culemborg, the Netherlands: Meurs Personeelsadvies.
- Lem, J. and Van Doorn, E. (2000) *Voortgangsrapportage. Onderzoek 'kenmerkende voorspellers' politie* ([Progress report. Study 'noticeable predictors' police]. Culemborg, the Netherlands: Meurs Personeelsadvies.
- McDaniel, M.A., Whetzel, D.L., Schmidt, F.L. and Maurer, S.D. (1994) The Validity of Employment Interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, **79**, 4, 599–616.
- Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K. and Schmitt, N. (2007a) Reconsidering the use of Personality Tests in Personnel Selection Contexts. *Personnel Psychology*, **60**, 3, 683–729.
- Morgeson, F.P., Campion, M.A., Dipboye, R.L., Hollenbeck, J.R., Murphy, K. and Schmitt, N. (2007b) Are we Getting

- Fooled Again? Coming to Terms with Limitations in the use of Personality Tests for Personnel Selection. *Personnel Psychology*, **60**, 4, 1029–1049.
- Motowidlo, S.J., Dunnette, M.D. and Carter, G.W. (1990) An Alternative Selection Procedure: The low-fidelity simulation. *Journal of Applied Psychology*, **75**, 6, 640–647.
- Murphy, K.R. (2002) Can Conflicting Perspectives on the Role of *g* in Personnel Selection be Resolved? *Human Performance*, **15**, 1/2, 173–186.
- Murphy, K.R. and Dzieweczynski, J.L. (2005) Why Don't Measures of Broad Dimensions of Personality Perform Better as Predictors of Job Performance? *Human Performance*, **18**, 4, 343–357.
- Olea, M.M. and Ree, M.J. (1994) Predicting Pilot and Navigator Criteria: Not much more than *g*. *Journal of Applied Psychology*, **79**, 6, 845–851.
- Occupational Information Network (O*NET) OnLine developed for the US Department of Labor by the National O*NET Consortium. (2007, January 31). (accessed 31 January 2007), from <http://online.onetcenter.org>
- Outtz, J.L. (2002) The Role of Cognitive Ability tests in Employment Selection. *Human Performance*, **15**, 1/2, 161–171.
- Pynes, J.E. and Bernardin, H.J. (1989) Predictive Validity of an Entry-Level Police Officer Assessment Center. *Journal of Applied Psychology*, **74**, 5, 831–833.
- Ree, M.J., Carretta, T.R. and Teachout, M.S. (1995) Role of Ability and Prior Knowledge in Complex Training Performance. *Journal of Applied Psychology*, **80**, 6, 721–730.
- Ree, M.J. and Earles, J.A. (1991) Predicting Training Success: Not much more than *g*. *Personnel Psychology*, **44**, 2, 321–332.
- Rijks Psychologische Dienst. (1975) *Politie Intelligentie Test* ([Police Intelligence Test]. The Hague, the Netherlands: Rijks Psychologische Dienst (RPD).
- Robertson, I.T. and Smith, M. (2001) Personnel Selection. *Journal of Occupational and Organizational Psychology*, **74**, 4, 441–472.
- Roth, P.L., Huffcutt, A.I. and Bobko, Ph. (2003) Ethnic Group Differences in Measures of Job Performance: A new meta-analysis. *Journal of Applied Psychology*, **88**, 4, 694–706.
- Rotundo, M. and Sackett, P.R. (1999) Effect of Rater Race on Conclusions Regarding Differential Prediction in Cognitive Ability Tests. *Journal of Applied Psychology*, **84**, 5, 815–822.
- Salas, E. and Cannon-Bowers, J.A. (2001) The Science of Training: A decade of progress. *Annual Review of Psychology*, **52**, 471–499.
- Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F. and Rolland, J.P. (2003) A Meta-Analytic Study of General Mental Ability Validity for Different Occupations in the European Community. *Journal of Applied Psychology*, **88**, 6, 1068–1081.
- 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**, 1, 162–173.
- 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**, 2, 262–274.
- SPSS. (2005) *SPSS 14.0 [Computer Software]*. Chicago: SPSS Inc.
- Te Nijenhuis, J. and Van der Flier, H. (2000) Differential Prediction of Immigrant Versus Majority Group Training Performance Using Cognitive Ability and Personality Measures. *International Journal of Selection and Assessment*, **8**, 2, 54–60.
- Van der Maesen, P.E.A.M. (1992). *Het Rendement van Personeelsselectie*. [The Efficiency of Personnel Selection]. Unpublished doctoral dissertation, Rijksuniversiteit Groningen.
- Van Leeuwen, A.E. (2000) *Constructie van de M5Q voor IWSP* [Development of the M5Q for IWSP]. Culemborg, the Netherlands: Meurs Personeelsadvies.
- Waldman, D.A. and Avolio, B.J. (1991) Race Effects in Performance Evaluations Controlling for Ability, Education, and Experience. *Journal of Applied Psychology*, **76**, 6, 897–901.
- Wiggins, J.S. and Trapnell, P.D. (1996) A Dyadic-Interactional Perspective on the Five-Factor Model. In: Wiggins, J.S. (ed.), *The Five-Factor Model of Personality*. New York: The Guilford Press, pp. 88–162.