Predicting Cross-Cultural Training Performance: The Validity of Personality, Cognitive Ability, and Dimensions Measured by an Assessment Center and a Behavior Description Interview

Filip Lievens  
Ghent University

Michael M. Harris  
University of Missouri—St. Louis

Etienne Van Keer  
TMP Worldwide

Claire Bisqueret  
SHL Belgium

This study examined the validity of a broad set of predictors for selecting European managers for a cross-cultural training program in Japan. The selection procedure assessed cognitive ability, personality, and dimensions measured by assessment center exercises and a behavior description interview. Results show that the factor Openness was significantly related to cross-cultural training performance, whereas cognitive ability was significantly correlated with language acquisition. The dimensions of adaptability, teamwork, and communication as measured by a group discussion exercise provided incremental variance in both criteria, beyond cognitive ability and personality. In general, these results are consistent with the literature on domestic selection, although there are some important differences.

The cost of managers on foreign assignments is extremely high for organizations. When such benefits as housing, airfare, and special premiums are included, the typical compensation package is estimated to cost at least 2.5 times an expatriate’s salary, and a significant number of companies indicate that the cost may be 4 to 5 times the base salary (Dowling, Welch, & Schuler, 1999). Moreover, the costs (e.g., in lost business, preparation, and so forth) associated with the failure of an expatriate employee are extremely high and have been estimated to range between $200,000 and $1.2 million (Duane, 2001). Because of the high stakes involved, organizations have been looking for ways to improve the adjustment of managers to foreign countries. One approach has been to focus on selection, which has begun to receive some research attention (Caligiuri, 2000; Deller, 1997; Ones & Viswesvaran, 1997; Spreitzer, McCall, & Mahoney, 1997). A second approach has been to invest in cross-cultural training, which has been identified as another major technique for improving managers’ cross-cultural effectiveness and for reducing failure rates (Bhagat & Prien, 1996; Bhawuk & Brislin, 2000; Black & Mendenhall, 1989; Deshpande & Viswesvaran, 1992; Dowling et al., 1999; Earley, 1987). It seems reasonable, therefore, to begin by selecting managers on the basis of their capacity to successfully master the cross-cultural training. Along these lines, Caligiuri (2000) asserted that practitioners should think of expatriate selection as the precursor to cross-cultural training. She further argued that organizations “should identify those expatriate candidates with the requisite personality characteristics, and then offer cross-cultural training to those identified. Cross-cultural training may only be effective when the expatriates are predisposed to success in the first place” (Caligiuri, 2000, p. 85).

However, recent reviews have indicated that the selection process of international managers is still intuitive and unsystematic (Deller, 1997; Ones & Viswesvaran, 1997; Sinangil & Ones, 2001). One of the problems is that the selection of people for foreign assignments is often based solely on job knowledge and technical competence (Aryee, 1997; Schmitt & Chan, 1998; Sinangil & Ones, 2001; Tung, 1981). Another problem is that in the past, researchers have mainly tried to determine a list of (inter) personal factors responsible for expatriate adjustment versus failure (e.g., Mendenhall & Oddou, 1985; Ones & Viswesvaran, 1997; Ronen, 1989) but that “there is apparently little research on designing selection systems which might be effective predictors of success in overseas assignments” (Arvey, Bhagat, & Salas, 1991, p. 378). More generally, it has been stated that there is a clear need to improve and broaden existing techniques for selecting people for international assignments (Arthur & Bennett, 1997; Church, 1982; Hough & Oswald, 2000). This need for more international selection research is especially striking in light of the large research base about predictors in traditional domestic selection (Schmidt & Hunter, 1998).
To respond to this gap in the literature, in this study we sought to examine the validity of four commonly used predictors, namely, cognitive center exercises, personality, and dimensions measured by assessment center exercises and a behavior description interview for selection into a cross-cultural training program. This study extends existing international selection research in three important ways. First, in response to the limited amount of research on expatriate selection, we broaden existing methods for selecting people for international assignments by including a wide range of predictors. Second, we examine the incremental validity of various predictors in an international context. Specifically, we focus on the incremental validity of dimensions measured in assessment center exercises and a behavior description interview over cognitive ability and personality in predicting cross-cultural training performance. Third, we examine a relatively unique criterion, namely, performance in cross-cultural training. Whereas various selection techniques have been used to predict training performance within domestic programs, there is almost no research with regard to their validity in predicting training performance in a cross-cultural context, despite the growing practical and scientific importance of cross-cultural training.

In the next sections, we review the relevant literature regarding the predictors used and formulate hypotheses on the basis of research in the domestic context. Beforehand, however, we discuss the practical and scientific importance of cross-cultural training and address why predictors of cross-cultural training success may differ from predictors of success in traditional training. We also briefly describe the cross-cultural training program under investigation.

Importance of Cross-Cultural Training

Cross-cultural training has long been proposed as an anticipatory mechanism to increase adjustment to foreign cultures (Mendenhall & Oddou, 1985; Tung, 1981). In cross-cultural training programs, a wide variety of training methods are typically used, including lectures, video films, experiential exercises, culture assimilators, and behavior modification (Bhawuk & Brislin, 2000; Black & Mendenhall, 1989; Mendenhall & Stahl, 2000). Bhagat and Prien (1996) discussed the main differences between traditional training and cross-cultural training. As they described, traditional training is characterized by a focus on the “acquisition of information, rather than on change in attitudes” (Bhagat & Prien, 1996, p. 223). By way of comparison, Bhagat and Prien observed that cross-cultural training addresses the acceptance of differences between cultures.

Several recent surveys (cited in Sinangil & Ones, 2001) illustrate that cross-cultural training is increasingly used as a staffing practice in international human resources management. A survey by Andersen Consulting, for example, revealed that among the best 32 of Fortune 500 organizations, 94% offered language training and 69% offered other cross-cultural training to international assignees. Sinangil and Ones (2001) cited another recent survey of 250 companies, showing that approximately 63% offered cross-cultural training to their expatriates. These usage percentages are much higher than those reported in earlier surveys, in which usage was estimated at around 25% (e.g., Black & Gregersen, 1991).

Furthermore, empirical research evidence attests to the effectiveness of cross-cultural training as a means to facilitate cross-cultural interactions in the foreign country. Black and Mendenhall (1990) reviewed 29 empirical studies that evaluated the effectiveness of various cross-cultural training programs. Their comprehensive literature review showed that cross-cultural training had a strong positive impact on participants’ self-confidence, on their interpersonal relationships with host nationals, and on their perceptions of the host culture. Equally important, Black and Mendenhall’s review (1990) confirmed the positive influence of cross-cultural training on expatriate adjustment and expatriate performance. Deshpande and Viswesvaran (1992) used meta-analysis to investigate the effectiveness of cross-cultural training. Across 21 empirical studies, a total of 1,611 participants were used to examine the effects of cross-cultural training on five criteria (expatriate perceptions, self-development, relations with host country nationals, adjustment, and performance). Results demonstrated the effectiveness of cross-cultural training as a key strategy for increasing expatriate success. For example, there was a corrected correlation between cross-cultural training and expatriate job performance of .39. Since this meta-analysis was published, other primary empirical studies have further confirmed the effectiveness of cross-cultural training programs (Bhawuk, 1998; Hammer & Martin, 1992; Harrison, 1992).

Given this growing practical and scientific value of cross-cultural training, it is important to select expatriates for participation in cross-cultural training (Caligiuri, 2000). Although there exists an extensive literature on the validity of different tools for selection into traditional training programs (Olea & Ree, 1994; Ree & Earles, 1991; Ree, Carretta, & Teachout, 1995; Salas & Cannon-Bowers, 2001; see also Table 2 in Schmidt & Hunter, 1998), there have been no studies about the predictors of success in a cross-cultural training program. Moreover, given the different requirements for success in traditional versus cross-cultural training (see Bhagat & Prien, 1996), it seems likely that the predictors may differ. As asserted by Hough and Oswald (2000), “validities of domestic selection instruments may not generalize to international sites, because different predictor and criterion constructs may be relevant, or, if the constructs are the same, the behavioral indicators may differ” (p. 649).

The Executive Training Program in Japan

The Executive Training Program (ETP) in Japan is a cross-cultural training program aimed at providing European managers with an in-depth understanding of Japanese business-related practices (including Japanese culture and society) and language. The European Commission financially supports the ETP and accordingly aims to facilitate European companies’ access to the Japanese market. The program consists of intensive language courses, university seminars, company visits, and in-house training. Thus, the ETP appears to be a blend of the experiential (e.g., field trips, role-playing) and analytical training methods (e.g., language training, case studies) in Black and Mendenhall’s (1989) cross-cultural training model. The learning in the ETP typically occurs in small groups (e.g., language learning occurs in groups of three). During 12 months, participants are familiarized with the Japanese language and with Japanese business and management. After these 12 months, participants work in Japanese host companies for 6 months. During these 6 months, executives work with Japanese coworkers and supervisors on various projects. Accordingly, it provides them with an opportunity to demonstrate that they have
acquired the language and business-related skills necessary to perform successfully in Japan.

The purpose of the selection procedure was to identify managers who would successfully complete the ETP in Japan. To determine the specific performance dimensions related to managers’ performance in the ETP in Japan, we began by holding a series of workshops with the ETP staff. The choice of the criteria was based on (a) the objectives of the ETP (see above), (b) a review of the current role requirements and skills present in successful participants of the ETP, and (c) a review of the extant literature on expatriate adjustment. Besides mastery of the Japanese language, seven performance dimensions were specified to represent the business-related component of cross-cultural training performance: self-discipline, tenacity–resilience, teamwork, communication, adaptability, cross-cultural awareness, and organizational and commercial awareness (see Table 1, first column). We also conducted workshops with the ETP staff to develop specific behavioral definitions of these seven dimensions (see Table 1).

Inspection of these performance dimensions confirms the aforementioned assertions about the distinctions between traditional training and cross-cultural training (Bhagat & Prien, 1996). In fact, some of the dimensions might be relevant for any training pro-

<table>
<thead>
<tr>
<th>Criteria (performance dimensions)</th>
<th>Big Five factors and selected OPQ scales</th>
<th>Assessment center exercises and behavior description interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tenacity–resilience (keeps difficulties in perspective, stays positive despite disappointments and setbacks)</td>
<td>Emotional Stability</td>
<td>Analysis–presentation (tenacity–resilience rating)</td>
</tr>
<tr>
<td></td>
<td>● Relaxed (calm, relaxed, cool under pressure, free from anxiety, can switch off)</td>
<td>Behavior description interview (tenacity–resilience rating)</td>
</tr>
<tr>
<td></td>
<td>● Tough-minded (difficult to hurt or upset, can brush off insults, unaffected by unfair remarks)</td>
<td></td>
</tr>
<tr>
<td>Communication (is able to communicate clearly, fluently, and to the point; talks at a pace and level that holds people’s attention, both in group and individual situations)</td>
<td>Extraversion</td>
<td>Analysis–presentation (communication rating)</td>
</tr>
<tr>
<td></td>
<td>● Affiliative (has many friends, enjoys being in groups, likes companionship, shares things with friends)</td>
<td>Group discussion (communication rating)</td>
</tr>
<tr>
<td></td>
<td>● Outgoing (fun-loving, humorous, sociable, vibrant, talkative, jovial)</td>
<td></td>
</tr>
<tr>
<td>Adaptability (adapts readily to new situations and ways of working, receptive to new ideas, willing and able to adjust to changing demands and objectives)</td>
<td>Openness</td>
<td>Analysis–presentation (adaptability rating)</td>
</tr>
<tr>
<td></td>
<td>● Change oriented (enjoys doing new things, seeks variety, prefers novelty to routine, accepts changes)</td>
<td>Group discussion (adaptability rating)</td>
</tr>
<tr>
<td></td>
<td>● Independent (has strong views on things, is difficult to manage, speaks up, argues, dislikes ties)</td>
<td></td>
</tr>
<tr>
<td>Teamwork (cooperates and works well with others in the pursuit of team goals, shares information, develops supportive relationships with colleagues and creates a sense of team spirit)</td>
<td>Agreeableness</td>
<td>Group discussion (teamwork rating)</td>
</tr>
<tr>
<td></td>
<td>● Democratic (encourages others to contribute, consults, listens, and refers to others)</td>
<td>Behavior description interview (teamwork rating)</td>
</tr>
<tr>
<td>Self-discipline (is committed, consistent, and dependable; can be relied on to deliver what has been agreed; is punctual and conscientious)</td>
<td>Conscientiousness</td>
<td>Behavior description interview (self-discipline rating)</td>
</tr>
<tr>
<td></td>
<td>● Conscientious (sticks to deadlines, completes jobs, perseveres with routine, likes fixed schedules)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>● Achieving (ambitious, sets sights high, career centered, results oriented)</td>
<td></td>
</tr>
<tr>
<td>Cross-cultural awareness (is able to see issues from the perspective of people of other cultures)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizational and commercial awareness (is alert to changing organizational dynamics, is knowledgeable about financial and commercial issues, focuses on markets, and business opportunities that bring the largest return)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. If two subscales measured a Big Five factor, subscale scores were averaged to obtain a score on the Big Five factor (see Caligiuri, 2000). OPQ = Occupational Personality Questionnaire.
gram. Examples include self-discipline, tenacity–resilience, teamwork, and communication. However, other performance dimensions (e.g., language proficiency, adaptability, and cross-cultural awareness) would not typically be critical in a domestic training program. Among these dimensions, language proficiency deals with the acquisition of information, whereas adaptability and cross-cultural awareness pertain to the acceptance of differences.

Validity of Predictors

Cognitive Ability

General cognitive ability, or g, has been found to be a consistent predictor of job performance across a variety of occupations (Schmidt & Hunter, 1998). This is especially the case for more complex job levels (Hunter, 1986). In addition, general cognitive ability has been identified as a key determinant of learning and skill acquisition (Kanfer & Ackerman, 1989; Salas & Cannon-Bowers, 2001). This is evidenced by several large-scale studies in military settings, which found a strong relationship between cognitive ability and training performance (Oleà & Ree, 1994; Ree et al., 1995; Ree & Earles, 1991). Finally, LePine, Colquitt, and Erez (2000) demonstrated that cognitive ability was an important determinant of effective coping with unforeseen changes in the task context. Thus, these studies in the domestic context show that general cognitive ability is consistently and strongly related to job and training performance and that this relationship increases in more complex, novel, and changing tasks.

Because international managers are required to learn and to deal with a variety of novel and complex situations in a cross-cultural training program (Spreitzer et al., 1997), cognitive ability measures were included in this study. In particular, managers’ general cognitive ability was estimated with tests of verbal and numerical reasoning. We chose these subtests because they are indicators of general cognitive ability (Ree et al., 1995). We expected that general cognitive ability would be significantly related to cross-cultural training performance (Hypothesis 1).

Personality

Over the past decade, the use of the five-factor model has served as a unifying theoretical framework to substantially advance our understanding of personality-based predictors in the context of domestic selection (Barrick & Mount, 1991; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Hurtz & Donovan, 2000; Salgado, 1997; Tett, Jackson, & Rothstein, 1991). Although Barrick and Mount’s (1991) meta-analysis of the Big Five personality dimensions focused on Conscientiousness as a predictor of job performance, their meta-analytic results also supported the validity of personality for predicting training performance. Specifically, in addition to Conscientiousness, both Extraversion (corrected \( r = .26 \)) and Openness (corrected \( r = .25 \)) exhibited sufficiently high validity coefficients with training proficiency to warrant further examination (Salas & Cannon-Bowers, 2001).

In an international context, personality-based predictors have also known a renaissance. Although prior research produced discouraging findings (see Ones & Viswesvaran, 1997, for an excellent review), recent studies have indeed found more promising validity results for the use of the five-factor model in a global context (Caligiuri, 2000; Dalton & Wilson, 2000; Schmit, Kihm, & Robie, 2000). In a concurrent validity study, Caligiuri (2000) concluded that all Big Five factors (with the exception of Openness) had statistically significant negative relationships with expatriates’ desire to prematurely terminate the assignment. In addition, Emotional Stability was significantly related to expatriates’ performance as rated by supervisors. Dalton and Wilson (2000) examined the relationship of the five-factor model to job performance ratings of 21 Middle Eastern expatriate managers. Both Conscientiousness and Agreeableness were related to home-country boss ratings but not to host-country boss ratings. With input from psychologists around the world, Schmit et al. (2000) constructed a “global” personality inventory based on the five-factor model. Initial predictive validity evidence was presented, showing that Conscientiousness, Extraversion, and Openness were significant predictors of job performance.

Because there seems to be an emerging consensus about the usefulness of the five-factor model in both domestic and international selection, three hypotheses were proposed. First, we hypothesized that Openness would be a significant predictor for managers’ cross-cultural training performance (Hypothesis 2a). This hypothesis was based on the fact that meta-analyses have shown that Openness is related to training performance in the domestic context (Barrick & Mount, 1991; Salgado, 1997). The significant relationship between Openness and training performance may be explained by the fact that open individuals are intellectually curious and independent thinkers who seek out new and unconventional experiences (Barrick & Mount, 1991). Accordingly, Openness may help people to better deal with the acceptance of differences between cultures, which is a prime component of cross-cultural training (Bhagat & Prien, 1996). Research has also shown that people high on Openness are more ready to engage in new learning experiences (Blickle, 1996), adapt more adequately to novel and unforeseen changes (LePine et al., 2000), and adjust more rapidly to a radically different work environment (Bing & Lounsbury, 2000). Similarly, Barrick and Mount (1991) found that Extraversion had a relatively large correlation with training performance. Individuals who are friendly and outgoing should be more likely to engage in interactive activities and therefore should be more successful in training. Also, recall that the training program used here relied heavily on working in small groups, so that Extraversion should enable trainees to be more effective in this setting. Hence, we hypothesized that Extraversion would be a significant predictor for managers’ cross-cultural training performance (Hypothesis 2b). Finally, Conscientiousness has been studied in relation to training programs and has been found to be an important predictor (e.g., Colquitt & Simmering, 1998). Trainees high on Conscientiousness are likely to be high on self-efficacy, motivation to learn, and other important training performance determinants (Colquitt, LePine, & Noe, 2000). On that basis, we hypothesized that Conscientiousness would be a significant predictor for managers’ cross-cultural training performance (Hypothesis 2c).

Dimensions Measured by Assessment Center Exercises

In the domestic context, assessment centers have been found to be good predictors of a variety of managerial criteria such as job and training performance (Gaugler, Rosenthal, Thornton, & Bent-
son, 1987; Pynes & Bernardin, 1989). The meta-analysis of Gau-  
gler et al. (1987) found a corrected validity coefficient of .35 for  
predicting training performance. In addition, a meta-analysis of  
work-sample tests of trainability (trainability tests) showed these  
tests to be good predictors of training success across a variety of  
jobs (Robertson & Downs, 1989). Yet an unanswered question is  
whether assessment centers are also predictive if used for inter-
national applications (Arthur & Bennett, 1997; Briscoe, 1997; Love,  
Bishop, Heinisch, & Montei, 1994; Ones & Viswesvaran, 1997).

Because assessment center exercises are methods that can be  
used to measure a wide array of constructs (Arthur, Day, McNelly,  
& Edens, in press), it should also be possible to design specific  
assessment center exercises to measure dimensions related to per-
formance in an international context. Hence, we constructed two  
assessment center exercises for measuring various dimensions  
related to cross-cultural training performance (see third column of  
Table 1). Specifically, an analysis–presentation exercise was de-
signed to measure the dimensions of tenacity–resilience, commu-
nication, adaptability, and organizational and commercial aware-
ness. Next, a group discussion exercise was constructed to capture  
the dimensions of teamwork, communication, adaptability, and  
analytical and commercial awareness. We hypothesized that the  
dimensions measured by these assessment center exercises would  
predict managers’ cross-cultural training performance (Hypothesis 3).

Dimensions Measured by High-Structure Interviews

Over the past decade, high-structure interviews have emerged as  
valid predictors of job performance in the domestic context (Huff-
cutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer,  
1994; Wiesner & Cronshaw, 1988). Two types of these interviews  
have dominated, namely, behavior description interviews (Janz,  
1982) and situational interviews (Latham, Saari, Pursell, & Cam-
pion, 1980). Besides being predictive of job performance, there is  
research suggesting that high-structure interviews are predictive of  
training performance. McDaniel et al. (1994) reported a correlation  
of .21 (corrected $r = .34$) between high-structure interviews and  
training performance, which was practically identical to the aver-
age validity of the interview for job performance. Contrary to this  
research in the domestic context, research on high-structure inter-
views in a cross-cultural context is scarce. An exception is Stahel  
(1995), who developed a high-structure interview for selecting  
managers for an assignment in Japan. Yet the sample was small  
($N = 8$), and no criterion-related validity evidence was gathered.

In this study, we included a behavior description interview in the  
selection procedure. Recent research comparing the behavior de-
scription interview and situational interview has shown that the  
former is more valid for higher level positions (Huffcutt, Weekley,  
Because a behavior description interview is essentially a method  
that can be used to measure a variety of constructs (Huffcutt,  
Conway, Roth, & Stone, 2001), it should also be possible to design  
a behavior description interview to measure dimensions related to  
performance in an international context. Therefore, we constructed  
a behavior description interview to measure several dimensions  
relevant to cross-cultural training performance (see third column  
of Table 1). In this study, a behavior description interview was  
developed around the dimensions of self-discipline, tenacity–  
resilience, teamwork, and cross-cultural awareness. We hypothe-
sized that the dimensions measured by this behavior description  
interview would predict managers’ cross-cultural training perfor-
mance (Hypothesis 4).

Incremental Validity

Apart from each predictor’s validity, it is both theoretically and  
practically pivotal to examine the predictive validity of different  
predictors over and above each other. From a theoretical view-
point, there is a need to understand whether these different mea-
ures are actually measuring similar or different constructs. From  
a practical standpoint, knowing whether to add additional predic-
tors that will explain additional variance in the criteria can help  
fine-tune a selection process. Typically referred to as incremental  
validity, the use of additional predictors is of value from a utility  
viewpoint only when they add additional variance explained in the  
criterion, beyond that which is accounted for by other, less expen-
sive predictors (Schmidt & Hunter, 1998). Because the validity of  
cognitive ability and personality (i.e., Conscientiousness) has been  
well established and because these predictors are less expensive,  
we examined whether dimensions measured by assessment center  
exercises and a behavior description interview provide incremental  
validity beyond these two predictors in explaining cross-cultural  
training performance.

In recent years, this specific incremental validity question has  
been addressed by some studies in the domestic context. In terms  
of the incremental validity of assessment centers over personality  
measures, research is scarce. Goffin, Rothstein, and Johnston  
(1996) reported almost no correlation between personality and  
assessment center ratings. Hence, both personality and assessment  
center ratings had significant incremental validity over one an-
other. There is somewhat more research on the relationship be-
tween assessment centers and cognitive ability. In their compre-
ensive review, Schmidt and Hunter (1998) included an earlier  
meta-analysis that found a high correlation ($r = .50$) between  
assessment center ratings and cognitive ability. They noted, how-
ever, that in this meta-analysis some assessment centers incorpo-
rated a cognitive ability measure, and so in certain instances, there  
might be contamination between assessment center ratings and  
cognitive ability. H. W. Goldstein, Yusko, Braverman, Smith, and  
Chung (1998) argued that the relationship varied as a function of  
the cognitive “loading” of assessment center exercises. That is,  
when exercises (e.g., an in-basket exercise) tapped more cognitive-
oriented dimensions (e.g., problem analysis), there was a stronger  
relationship between the assessment center exercise and cognitive  
ability. In support of this hypothesis, Goldstein et al. found  
exercise–cognitive ability test score correlations ranging between  
.09 and .29. Yet these are relatively modest correlations, and only  
one of the five exercises became statistically unrelated to perfor-
mance when cognitive ability was partialed out. Thus, Goldstein et  
al.’s results suggested that dimensions measured in assessment  
center exercises would provide incremental validity over cognitive  
ability, even with moderate correlations between these two predic-
tors. There is a relatively large literature examining the incremental  
validity of high-structure interviews over cognitive ability. Al-  
though earlier studies suggested that the high-structure interview  
measured little more than cognitive ability (e.g., Campion, Pursell,
& Brown, 1988), Huffcutt, Roth, and McDaniel’s (1996) meta-
analysis reported a correlation of .12 (corrected \( r = .18 \)) between
behavior description interviews and cognitive ability. In terms of
the relationship between interview ratings and personality, most of
the research has focused on Conscientiousness. Cortina, Goldstein,
Payne, Davison, and Gilliland (2000) concluded that the correla-
tion between interview ratings and other constructs varied, depend-
ing on the amount of structure imposed on the interview. At higher
levels of structure, they found an average correlation of .21 (cor-
corrected \( r = .26 \)) between interview ratings and Conscientiousness.
The relationship between interview ratings and other Big Five
factors (e.g., Agreeableness) has been the focus of far less re-
search. In one recent study, however, Huffcutt, Weekley, Wiesner,
DeGroot, and Jones (2001) reported small correlations between the
Big Five and a behavior description interview, with the exception
of a moderate correlation for Extraversion (\( r = .30 \)).

In sum, prior research in the domestic context about the incre-
mental validity of dimensions measured by assessment center
exercises and a behavior description interview over cognitive
ability and personality shows that there are at best moderate
correlations among these predictors. Therefore, we hypothesized
that dimensions measured by assessment center exercises and a
behavior description interview would add incremental validity
over and beyond personality and cognitive ability (Hypothesis 5).

Method

Sample

One hundred sixty-six European managers participated in the selection
procedure. The sample included 125 men and 41 women. Participants
ranged in age from 23 to 52 years, with an average age of 29.3 years
(\( SD = 4.4 \) years). In total, 15 European nationalities were involved. With
regard to job titles, the sample was relatively heterogeneous. Most man-
gers were active in export, marketing, sales, or general management. In
terms of managerial level, many were at the executive level, although
others were high-potential employees who were judged likely to become
executives. They were working in a variety of companies (services, man-
ufacturing, electronics, etc.). Most of the companies (71%) were doing
business with Japan. Forty percent were small companies (fewer than 50
employees), 20% were medium-sized companies (between 50 and 250
employees), and 38% were large companies (more than 250 employees).

The preselection of the managers had taken place in their home countries
by local divisions of the consultancy firm responsible for the actual
selection. This preselection was done on the basis of relevant prior expe-
rience and other biographical information (e.g., 2 years of managerial
experience, 6 months of international experience). The actual selection
procedure lasted 1 day. In total, 30 sessions were held. Each session was
attended by four to six managers (of differing nationalities; see Briscoe,
1997), who were evaluated by experienced psychologists (i.e., external
consultants) and staff members of the ETP.

Predictor Measures

Cognitive ability measures. The first test (VMG 1) was a verbal critical-
reasoning test that consisted of 52 items describing short passages of
complex and business-related information (Saville & Holdsworth, 1989).
The questions asked candidates to evaluate whether specific statements
were congruent with the information provided. The time limit was 25 min.
The second test (NMG 1) was a numerical critical-reasoning test that consisted
of 40 items depicting business-related tables and graphs (time
limit: 35 min; Saville & Holdsworth, 1989). Candidates were required to
interpret and use this complex numerical information to answer the ques-
tions. Candidates completed these tests in their own language.

Because these cognitive ability measures were the publisher’s property,
we received only candidates’ final scores and were not able to compute
internal consistencies. Prior research mentioned in the test manual, how-
ever, indicated internal consistencies varying (across different studies)
from .70 to .80 for VMG 1 and from .82 to .90 for NMG 1. Across different
validation studies and criteria, validity coefficients ranged from .17 (\( p < .05 \), \( N = 109 \)) to .37 (\( p < .05 \), \( N = 38 \)).
The correlation between VMG 1 and NMG 1 scores was .38 (\( p < .01 \)).
We averaged scores on the two tests to obtain a measure of general
cognitive ability (see Ree et al., 1995).

Personality inventory. To measure the Big Five personality factors, we
selected subscales from the Occupational Personality Questionnaire (OPQ;
Saville & Holdsworth, 1990). Although the OPQ was originally not de-
signed to measure the five-factor model, various studies (e.g., Beausjouan,
2000; Ferguson, Payne, & Anderson, 1994; Matthews & Stanton, 1994;
Matthews, Stanton, Graham, & Brimelow, 1990) have shown that the
subscales of the OPQ can be retranslated to the five-factor model. For
example, Matthews et al. (1990) demonstrated that the majority of covari-
ance among the OPQ scales could be explained by five factors, which
corresponded well to the five-factor model. Other studies have found that
the OPQ consistently predicted job performance across organizations,
nations, and time (Robertson & Kinder, 1993; Saville, Sik, Nyfield, Hack-
ston, & MacIver, 1996). On the basis of these prior studies, we chose nine
subscales and linked them to the Big Five factors and performance dimen-
sions (see also Barrick, Stewart, & Piotrowski, 2002; Caligiuri, 2000;
Cleveenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey, 2001). The
third column of Table 1 indicates which scales of the OPQ were selected.
For instance, Extraversion was measured with the Affiliative and Outgoing
subscales of the OPQ.

We used the Concept 4.2 version of the OPQ. In total, this version of the
OPQ consisted of 360 items measuring 31 scales. Candidates were required
to select one least and one most preferred item from a set of four alterna-
tives. They completed the OPQ in their own language. Because the OPQ is
the property of the publisher, we received only candidates’ scores on the
subscales. Hence, we were not able to compute the reliability of the scales.
However, prior studies that conducted analyses at the item level have found
satisfactory reliabilities (e.g., Matthews et al., 1990). For the scales in-
cluded in our study, test–retest reliabilities varied between .73 and .94, and
internal consistencies varied between .68 and .89 (Matthews et al., 1990).

Assessment center exercises. In the first exercise, an analysis–pre-
presentation exercise, each candidate assumed the role of a consultant and
analyzed a complex set of facts and figures relating to various departments
of a medium-sized organization. Each candidate had to determine and
present the strategy for the next 5 years. This exercise provided opportu-
nities to observe and evaluate the dimensions of tenacity–resilience, com-
unication, adaptability, and organizational and commercial awareness
(see Table 1). The second exercise was a group discussion in which
candidates had to reach consensus on cost reductions. This exercise mea-
sured the dimensions of communication, adaptability, teamwork, and or-
ganizational and commercial awareness (see Table 1). Assessors were both experienced psychologists and ETP staff members.
The psychologists came from a pool of 10 external consultants (7 women
and 3 men; average assessor experience 10 years, \( SD = 7 \) years). The
rationale for including ETP staff was that they were knowledgeable about
Japanese culture. Hence, they could give advice as to which cultural model
to apply when deciding on the effectiveness of behaviors (Briscoe, 1997).
Both psychologists and ETP staff had attended a 3-day training seminar in
accordance with the Guidelines and Ethical Considerations for Assessment
Center Operations (Task Force on Assessment Center Guidelines, 1989).
Assessors were given behavioral checklists to aid their observation,
recording, and classification tasks. After each exercise, they rated candidates
on a 9-point rating scale ranging from poor (1) to outstanding (9) per
dimension. There was one assessor per candidate. Candidates were rated by different assessors across exercises.

Behavior description interview. Candidates were asked to give specific examples of past situations wherein they had demonstrated particular skills that were expected to be related to success in cross-situational training. The dimensions measured were self-discipline, tenacity–resilience, teamwork, and cross-cultural awareness (see Table 1). Only experienced psychologists served as interviewers. There was one interviewer per candidate. Interviewers were given a detailed interview guide, which listed example questions and answers per dimension. A sample question asked candidates to discuss a stressful situation they had to deal with in the past (related to the tenacity–resilience dimension). Probes included such inquiries as “What made the situation so stressful?”, “How did you handle the situation?”, “What else could you have done?”, and “What was the final result?” Candidates were rated on a 9-point rating scale ranging from poor (1) to outstanding (9) per dimension. On average, the interview lasted 45 min.

Final decision. The ultimate decision as to whether to accept an applicant into the ETP was based on several ratings. First, candidates’ results on the aforementioned selection instruments were mechanically integrated into a final score. Second, the ETP staff conducted an interview whereby they asked applicants about the readiness and plans of their company for doing trade with Japan, the financial commitment of their company to the ETP, and related questions. Third, the external consultants and the ETP staff members clinically integrated candidates’ final selection procedure rating and ETP staff members’ judgments to determine the ETP fellows.

Criterion Measures

Training performance. We obtained instructors’ ratings on the performance dimensions (see our description of the ETP and Table 1) that captured the business-related component of the ETP. Thirty instructors from the school at which the language courses and business-related seminars and workshops were held rated the managers after the managers had left the school (i.e., after 12 months). These instructors were involved with the managers on a daily basis. None of these instructors was familiar with the managers’ evaluation at the time of the selection in Europe. It was made clear that the data gathered would be kept confidential and for research purposes only. Instructors were required to evaluate a manager on the consistency of this composite was .79.

Language acquisition. As a second criterion, we obtained objective data about the language component of the cross-cultural training. These data captured the information acquisition component that is typical for cross-cultural training (Bhagat & Prien, 1996). Managers’ proficiency of the Japanese language was measured at the halfway point (6 months) and at the end of the courses (12 months). Each exam consisted of various subsections (e.g., reading, understanding, writing, and speaking Japanese). The total score for each exam was obtained by summing the scores on each subsection. The maximum total score on each exam was 200. Mean exam scores at 12 months ($M = 127.81$) were significantly higher than mean exam scores at 6 months ($M = 76.32$). $t(75) = 32.01, p < .001 (d = 1.91)$. These two exam scores correlated .91.

These two training performance criteria are independent because the language exams were conducted by an external organization and the instructors who gave the ratings on the dimensions were not familiar with the exam scores. In addition, the correlation between the final exam score and the instructors’ ratings was relatively low ($r = .26, p < .05$), indicating that these two criteria were empirically distinct.

Because only 86 of the 166 managers (52%) were selected for the ETP in Japan, 6 managers did not attend the ETP (although they were selected), and 2 managers left the ETP, training performance data were available for 78 managers.

Results

Validity of Predictors

Table 2 presents the means, standard deviations, and correlations among the study’s variables. The last two rows display the bivariate correlations between the various predictors and the criteria. We corrected correlations for direct range restriction (Thurstone, 1949) and for unreliability in the criterion. To correct for unreliability in instructor ratings of training performance, we used the same value (.80) as McDaniel et al. (1994). To correct for unreliability in the language proficiency criterion, we used the correlation between the two exams (.91). Statistical significance was determined prior to correcting the correlations (Sackett & Yang, 2000).

Hypothesis 1 proposed that cognitive ability would be significantly related to cross-cultural training performance. In support of this hypothesis, cognitive ability was significantly correlated with the test measuring language acquisition ($r = .23, p < .05$; corrected $r = .27$). However, cognitive ability was not significantly correlated with instructors’ ratings of training performance ($r = .09, ns$; corrected $r = .11$). Thus, there was partial support for Hypothesis 1.

The next set of hypotheses proposed significant relationships between cross-cultural training performance and Openness (Hypothesis 2a), Extraversion (Hypothesis 2b), and Conscientiousness (Hypothesis 2c). There was mixed support for this set of hypotheses. On the positive side, Openness was significantly correlated with instructors’ ratings of cross-cultural training performance ($r = .31, p < .01$; corrected $r = .33$). Yet Extraversion was not related to instructors’ ratings of training performance ($r = -.03, ns$; corrected $r = -.04$), nor was Conscientiousness ($r = .18, ns$; corrected $r = .20$).

Although we did not have specific hypotheses about Emotional Stability or Agreeableness, we examined their correlations with instructors’ ratings of training performance. Emotional Stability was not significantly related to instructor ratings ($r = -.03, ns$; corrected $r = -.04$). Agreeableness was significantly, albeit negatively, related to instructor ratings ($r = -.24, p < .05$; corrected $r = -.26$). Emotional Stability was the only personality factor that was significantly, albeit negatively, correlated with the language proficiency test ($r = -.28, p < .05$; corrected $r = -.29$).

Hypothesis 3 assumed that the dimensions as measured by assessment center exercises would be significantly related to cross-cultural training performance. Partial support was obtained for this hypothesis. All dimensions measured in the group discussion exercise were significantly correlated with instructor ratings and with language proficiency. Teamwork, communication, adaptability, and organizational and commercial awareness as measured in the group discussion exercise had uncorrected correlations varying from .25 ($p < .05$) to .31 ($p < .01$) with instructor ratings (range of corrected $r$s = .31 to .40) and from .28 ($p < .05$) to .38 ($p < .01$) with the language proficiency test (range of corrected $r$s = .33 to .44). However, none of the dimensions measured in the analysis–presentation exercise was significantly correlated with...
### Table 2
**Means, Standard Deviations, and Correlations of Study Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7a</th>
<th>8a</th>
<th>8b</th>
<th>9a</th>
<th>9b</th>
<th>10a</th>
<th>10b</th>
<th>11a</th>
<th>11b</th>
<th>12a</th>
<th>13a</th>
<th>13b</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cognitive ability</td>
<td>5.28</td>
<td>1.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Emotional stability</td>
<td>5.61</td>
<td>1.56</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Extraversion</td>
<td>6.02</td>
<td>1.57</td>
<td>-.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Openness</td>
<td>5.82</td>
<td>1.45</td>
<td>.08</td>
<td>-.03</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Agreeableness</td>
<td>6.10</td>
<td>2.07</td>
<td>-.12</td>
<td>-.07</td>
<td>.05</td>
<td>-.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Conscientiousness</td>
<td>5.71</td>
<td>1.40</td>
<td>-.03</td>
<td>-.03</td>
<td>-.22</td>
<td>-.13</td>
<td>-.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Self-discipline</td>
<td>6.31</td>
<td>1.29</td>
<td>.01</td>
<td>-.02</td>
<td>-.23</td>
<td>-.02</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Tenacity–resilience</td>
<td>5.60</td>
<td>1.55</td>
<td>.09</td>
<td>.05</td>
<td>.14</td>
<td>-.01</td>
<td>-.09</td>
<td>.02</td>
<td>.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Interview</td>
<td>6.09</td>
<td>1.21</td>
<td>.11</td>
<td>-.02</td>
<td>-.06</td>
<td>.00</td>
<td>-.03</td>
<td>-.01</td>
<td>.41</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Interview</td>
<td>5.35</td>
<td>1.76</td>
<td>-.24</td>
<td>-.01</td>
<td>.10</td>
<td>.00</td>
<td>-.04</td>
<td>.13</td>
<td>.31</td>
<td>.26</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Teamwork</td>
<td>5.80</td>
<td>1.25</td>
<td>.20</td>
<td>.00</td>
<td>.09</td>
<td>.00</td>
<td>.13</td>
<td>.01</td>
<td>.36</td>
<td>.20</td>
<td>.36</td>
<td>.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Group discussion</td>
<td>5.33</td>
<td>1.66</td>
<td>.24</td>
<td>-.02</td>
<td>.08</td>
<td>.05</td>
<td>-.12</td>
<td>.08</td>
<td>.21</td>
<td>.31</td>
<td>.08</td>
<td>.88</td>
<td>.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Communication</td>
<td>4.69</td>
<td>1.77</td>
<td>.29</td>
<td>.01</td>
<td>.09</td>
<td>.01</td>
<td>-.02</td>
<td>.13</td>
<td>.27</td>
<td>.38</td>
<td>.15</td>
<td>.88</td>
<td>.21</td>
<td>.83</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Group discussion</td>
<td>5.01</td>
<td>1.74</td>
<td>.25</td>
<td>.00</td>
<td>.16</td>
<td>.06</td>
<td>-.14</td>
<td>.01</td>
<td>.20</td>
<td>.77</td>
<td>.22</td>
<td>.37</td>
<td>.26</td>
<td>.42</td>
<td>.76</td>
<td>.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Analysis–presentation</td>
<td>6.05</td>
<td>1.37</td>
<td>.03</td>
<td>-.11</td>
<td>.13</td>
<td>-.08</td>
<td>-.03</td>
<td>-.07</td>
<td>.36</td>
<td>.23</td>
<td>.27</td>
<td>.28</td>
<td>.42</td>
<td>.22</td>
<td>.24</td>
<td>.27</td>
<td>.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Adaptability</td>
<td>3.95</td>
<td>1.79</td>
<td>.25</td>
<td>.04</td>
<td>.05</td>
<td>-.01</td>
<td>.03</td>
<td>.08</td>
<td>.23</td>
<td>.36</td>
<td>.13</td>
<td>.77</td>
<td>.17</td>
<td>.76</td>
<td>.37</td>
<td>.88</td>
<td>.44</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Group discussion</td>
<td>4.61</td>
<td>1.78</td>
<td>.22</td>
<td>-.05</td>
<td>.12</td>
<td>.04</td>
<td>-.05</td>
<td>.01</td>
<td>.20</td>
<td>.67</td>
<td>.21</td>
<td>.45</td>
<td>.26</td>
<td>.47</td>
<td>.69</td>
<td>.54</td>
<td>.84</td>
<td>.22</td>
<td>.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Analysis–presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Cross-cultural awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Organizational and commercial awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Group discussion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Analysis–presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Training performance</td>
<td>3.35</td>
<td>0.53</td>
<td>.09</td>
<td>-.03</td>
<td>-.03</td>
<td>.31</td>
<td>-.24</td>
<td>.18</td>
<td>.12</td>
<td>.10</td>
<td>-.04</td>
<td>.30</td>
<td>.15</td>
<td>.31</td>
<td>.06</td>
<td>.28</td>
<td>.11</td>
<td>.05</td>
<td>.25</td>
<td>.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Language proficiency</td>
<td>127.81</td>
<td>30.53</td>
<td>.23</td>
<td>-.28</td>
<td>.07</td>
<td>.10</td>
<td>.03</td>
<td>.12</td>
<td>.01</td>
<td>-.09</td>
<td>-.12</td>
<td>.38</td>
<td>-.01</td>
<td>.35</td>
<td>-.10</td>
<td>.33</td>
<td>-.02</td>
<td>.14</td>
<td>.28</td>
<td>.10</td>
<td>.26</td>
<td></td>
</tr>
<tr>
<td>(N = 78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Correlations between predictors and training performance that were corrected for range restriction and criterion unreliability appear in parentheses. Statistical significance was determined prior to the range restriction correction. Uncorrected correlations with the criteria of training performance and language acquisition above .22 are significant at p < .05 and above .29 at p < .01. Other uncorrected correlations above .16 are significant at p < .05 and above .20 at p < .01.
the criteria (range of uncorrected $r$ = $-.04$ to $.13$ for training performance and $-.12$ to $.10$ for language acquisition).

Hypothesis 4 proposed that there would be a significant relationship between dimensions measured by the behavior description interview and cross-cultural training performance. No support was found for this hypothesis because none of the dimensions measured by the behavior description interview yielded significant correlations with our criteria.

**Incremental Validity**

Hypothesis 5 posited that dimensions measured by assessment center exercises and a behavior description interview would have incremental validity over personality and cognitive ability. Our general strategy for testing this hypothesis consisted of holding the construct under investigation constant. Accordingly, our analyses avoided the pitfall of confounding the content (constructs measured) and the methods (techniques used to measure the specific content) (Arthur et al., in press; Chan & Schmitt, 1997; Schmitt & Chan, 1998; Schmitt & Mills, 2001).

To test Hypothesis 5, for instance, we conducted a hierarchical regression analysis in which we entered one Big Five personality factor (e.g., Openness) and cognitive ability in the first step. In the second step, we entered a conceptually related dimension (e.g., Adaptability) as measured by an analysis–presentation exercise and a group discussion exercise. Hence, we could examine whether Adaptability as measured by an analysis–presentation exercise and a group discussion exercise added additional variance over cognitive ability and the conceptually related personality trait of Openness. We linked Adaptability to Openness because prior research showed that these were related constructs (LePine et al., 2000). In another hierarchical regression analysis, we investigated whether teamwork as measured by a group discussion and a behavior description interview explained additional variance, even when cognitive ability and Agreeableness were already accounted for. We linked teamwork to Agreeableness because prior research demonstrated that these were related constructs (Mount, Barrick, & Stewart, 1998). Table 1 shows which three other conceptually relevant linkages were examined to test Hypothesis 5.

Because multivariate range restriction and criterion unreliability might also affect the regression results, we applied the appropriate corrections (see Ree et al., 1995) to the matrix of correlations (Table 2) and used this corrected matrix as input for the hierarchical regression analyses. Statistical significance was determined prior to applying the corrections (by conducting hierarchical regressions on the uncorrected matrix of correlations).

Table 3 presents the results of these hierarchical regression analyses for both of our training performance criteria. Hypothesis 5 was partially confirmed because three dimensions accounted for a significant amount of additional variance in cross-cultural training performance. Teamwork significantly explained additional variance in our two criteria of training performance over and beyond cognitive ability and Agreeableness. Communication accounted for a significant additional portion of the variance in training performance over and beyond cognitive ability and Extraversion. Adaptability added a significant amount of variance over cognitive ability and Openness. The dimensions of self-confidence and tenacity–resilience did not provide incremental validity over cognitive ability and conceptually related personality traits.

**Discussion**

In this study we examined the validity of a broad set of predictors for selecting European managers for a cross-cultural training program in Japan. The selection procedure assessed personality traits and cognitive ability and also dimensions measured by assessment center exercises and a behavior description interview. This study has several important contributions with regard to the validity of these predictors in an international context, their incremental validity, and the criterion (cross-cultural training performance) used to evaluate them.

**Validity of Individual Predictors in an International Context**

Although there exists a large research base about predictors in a traditional domestic selection context (Schmidt & Hunter, 1998), relatively few studies have examined these predictors in an international context. This is especially the case for predictors of cross-cultural training, despite the growing practical and scientific importance of cross-cultural training. Therefore, one of the contributions of this study consists in demonstrating which validity results obtained in a domestic context hold in an international selection context.

Many, but certainly not all, of our hypotheses for the validity results were supported. A traditionally valid predictor as cognitive ability was significantly related to the language acquisition dimension of cross-cultural training performance, but it was not significantly related to other aspects of cross-cultural training performance.

With regard to the personality traits, Openness was significantly related to instructors’ ratings of cross-cultural training performance. In line with our hypothesis, this result extends the training findings for Openness to an international context. It also fits well with the general contention that open individuals are more adaptable and therefore more ready to accept differences between various cultures (Bhagat & Prien, 1996; Bing & Lounsbury, 2000). Conversely, no support was found for our hypotheses that Extraversion and Conscientiousness would be significantly related to cross-cultural training success. A closer look at our correlation between Conscientiousness and cross-cultural training ($r = .18$; corrected $r = .20$), nevertheless, indicates that it is quite close to the meta-analytic value ($r = .13$; corrected $r = .23$) reported by Barrick and Mount (1991). Thus, although our correlation between Conscientiousness and instructor ratings did not reach statistical significance here, it is approximately the same as that reported by Barrick and Mount for overall training proficiency.

Contrary to the extant literature, some validity coefficients for the Big Five factors (e.g., Agreeableness) were negative. On the one hand, these divergent results may stem from the use of an ipsative personality measure (Baron, 1996; Hicks, 1970). On the other hand, as argued by Tett, Jackson, Rothstein, and Reddon (1999), these negative correlations may reflect true relationships between the predictors and the criteria (e.g., for an explanation for the negative correlation of Agreeableness in an international context, see Caligiuri, 2000), which raises the possibility that person-
Cross-validity was computed with the formula of Cattin (1980). Because of rounding,

| Note. N correlated with the instructors work as measured in the group discussion were significantly
| correlated with performance in the meta-analysis of Gaugler et al. (1987). When
| these dimensions were measured in the analysis—presentation, they were not significantly correlated with training performance.

The most straightforward explanation for these results is that these dimensions closely mirrored the skills needed in the small-group learning setting of the ETP. In other words, executives who scored well on these dimensions (communication, adaptability, and teamwork) in the group discussion also received high ratings from the instructors on the dimensions and were better able to study the Japanese language in small groups. In addition, assessors told us that the group exercise provided them with ample opportunities for evaluating each candidate’s quality of contribution.

In partial support of our hypothesis with regard to the assessment center dimensions, communication, adaptability, and teamwork as measured in the group discussion were significantly correlated with the instructors’ training performance ratings and with the language acquisition scores. The corrected validities obtained for these dimensions are even higher than the corrected mean validity (.35) of assessment centers for predicting training performance in the meta-analysis of Gaugler et al. (1987). When these dimensions were measured in the analysis—presentation, they were not significantly correlated with training performance.

The most straightforward explanation for these results is that these dimensions closely mirrored the skills needed in the small-group learning setting of the ETP. In other words, executives who scored well on these dimensions (communication, adaptability, and teamwork) in the group discussion also received high ratings from the instructors on the dimensions and were better able to study the Japanese language in small groups. In addition, assessors told us that the group exercise provided them with ample opportunities for evaluating each candidate’s quality of contribution. Perhaps the other exercise (analysis—presentation) did not provide enough opportunities for observing and evaluating these dimensions. Apparently, to obtain good predictive validity, practitioners should pay attention to simulate the specific (cultural) environment that international managers will work in. Future studies are needed to single out which other assessment center design factors are of key importance to guarantee valid predictions in an international context (see Briscoe, 1997). For instance, should assessors come from both the sending and the receiving country? And should assesses with different cultural backgrounds be assessed together?

Finally, we found no support for our hypothesis regarding the validity of high-structure interviews. Given the specified dimensions that it was designed to measure, the behavior description

### Table 3

**Summary of Hierarchical Regression Analyses of Cross-Cultural Training Performance and Language Proficiency on Predictors**

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictors</th>
<th>Training performance</th>
<th>Language acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\beta$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1</td>
<td>Cognitive ability</td>
<td>.18</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>Emotional Stability</td>
<td>.08</td>
<td>.47**</td>
</tr>
<tr>
<td>2</td>
<td>Tenacity (analysis—presentation)</td>
<td>.10</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>Tenacity (interview)</td>
<td>-.05</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Cross-validity = .02</td>
<td>Adj. $R^2$ = .20</td>
<td>Cross-validity = .18</td>
</tr>
<tr>
<td>1</td>
<td>Cognitive ability</td>
<td>.08</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Openness</td>
<td>.11</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>Adaptability (analysis—presentation)</td>
<td>.11</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Adaptability (group discussion)</td>
<td>.11</td>
<td>.29</td>
</tr>
<tr>
<td></td>
<td>Cross-validity = .22</td>
<td>Adj. $R^2$ = .25</td>
<td>Cross-validity = .19</td>
</tr>
<tr>
<td>1</td>
<td>Cognitive ability</td>
<td>.12</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Conscientiousness</td>
<td>.16</td>
<td>.08</td>
</tr>
<tr>
<td>2</td>
<td>Self-discipline</td>
<td>.16</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>Cross-validity = .03</td>
<td>Adj. $R^2$ = .04</td>
<td>Cross-validity = .03</td>
</tr>
</tbody>
</table>

*Note. N = 77. Estimates are for second step, not entry. Input correlation matrix was corrected for multivariate range restriction and criterion unreliability. Cross-validity was computed with the formula of Cattin (1980). Because of rounding, $\Delta R^2$ differs .01 from cumulative $R^2$. Adj. = adjusted.

*p < .05. **p < .01.
Incremental Validity in an International Context

In recent years, a growing amount of studies have examined the incremental validity of multidimensional predictors such as assessment centers (Goffin et al., 1996), high-structure interviews (Cortina et al., 2000), situational judgment tests (Clevenger et al., 2001), and biodata (Mount, Witt, & Barrick, 2000) over cognitive ability and the Big Five personality traits. Our study contributed to this growing literature by examining incremental validity issues in an international context. An important finding was that only the dimensions of communication, teamwork, and adaptability as measured by a group discussion exercise provided incremental variance in cross-cultural training performance beyond cognitive ability and conceptually related personality traits.

Another contribution to the developing incremental validity literature in personnel selection is that we used a construct-driven analytical strategy for examining incremental validity. When prior research found that assessment centers provided incremental variance beyond personality (Goffin et al., 1996), no meaningful conclusions could be drawn, as this result might have been found because assessment centers have a different selection procedure than a personality inventory or because assessment centers capture constructs other than the Big Five personality factors. In our analyses, however, the construct under investigation was held constant so that we were able to conclude that dimensions as measured by the assessment center method provided incremental variance over and above conceptually related constructs measured by a different selection method (i.e., a personality inventory). More generally, we hope that this study encourages researchers to adopt a construct-driven focus in incremental validity investigations.

Cross-Cultural Training Performance as Criterion

This study focused on predictors of cross-cultural training performance. Hereby we assumed that if people did not succeed in the cross-cultural training in the first place, it would be unlikely that they would perform successfully in the cross-cultural assignment (see Caligiuri, 2000). To examine whether the ETP also predicted job performance, we gathered data regarding executives’ performance in Japanese companies. As mentioned above, the European executives worked in Japanese companies after the ETP. We asked the Japanese supervisors of the executives to rate them after 6 months on similar rating forms, as described in the Method section. These supervisors were familiar neither with the executives’ ETP evaluation nor with their selection evaluation in Europe. Rating forms were returned for 33 of the 78 managers. Both criteria of cross-cultural training performance were significantly correlated with a composite measure \( (\alpha = .89) \) of executives’ performance in Japanese companies \( (r = .38, p < .05, \text{for instructors’ ratings and } r = .45, p < .05, \text{for language proficiency}) \). These correlations are in the same range as the meta-analytic correlations found by Deshpande and Viswesvaran (1992) and demonstrate that ETP performance is related to subsequent job performance in Japan.\footnote{Because language acquisition scores significantly improved during the ETP (see Method section), we have evidence that the significant relationship between language acquisition scores and job performance in Japanese companies was due to the language training component of the ETP. However, as noted by an anonymous reviewer, because no data were available showing that the executives improved on the performance dimensions during the ETP, the significant correlation between instructors’ ratings of training performance and supervisors’ ratings of job performance might be due to factors (e.g., the fact that both instructors and supervisors were Japanese) other than the skill training received during the ETP.}

Similar to Caligiuri (2000), this study used a process of selecting people into cross-cultural training, providing those selected with cross-cultural training, and then sending abroad those who passed the training. An important advantage of this process is that it may reduce the costs associated with international assignments because only people who have passed the selection and therefore are considered to be predisposed for expatriate success are sent to the training and abroad. Similar cost arguments have been made in the context of selecting trainees for traditional training (L. Goldstein & Ford, 2002). In addition, because the selection and training procedures might serve as mini–work samples for future expatriate performance, they might shape realistic perceptions about the foreign culture. Besides this model of selection, training, and assignment, another process simply consists of selecting people for the international assignment and sending those selected abroad. Research is needed to compare the viability of these and other international human resources staffing models.

Limitations of Study

This study has several limitations. First, this study involved European managers participating in cross-cultural training. Whether our results are due to the use of non–North American managers or whether they are due to success in a cross-cultural training program as the criterion needs to be determined through future research. So far, meta-analytic validity results for personality (Salgado, 1997) and cognitive ability (Salgado & Anderson, 2002) that were obtained in Europe have shown good correspondence with North American results. Second, as with most studies with international managers (Arthur & Bennett, 1997), this study had a relatively small sample size. As a result, the statistical power to detect significant relationships was low (although we found many significant results nevertheless). Because of these limitations, our findings should be interpreted with caution, and studies with other and larger samples and different criterion measures are needed to confirm our findings. Third, even though a variety of predictors were used for selecting people into a cross-cultural training program, some potentially
interesting predictors were not included (see Colquitt et al., 2000; I. L. Goldstein & Ford, 2002). An example is self-efficacy. In the domestic context, self-efficacy predicts training performance, and training programs built to modify self-efficacy also appear to be effective in changing job performance (Ford, Quinones, Sego, & Sorra, 1992).

Summary
The selection of international managers is a field full of opportunities and challenges for both practitioners and researchers. The selection procedure in this study broadened existing expatriate selection procedures by incorporating cognitive ability, personality, and dimensions measured by assessment center exercises and a behavior description interview. The overall results are encouraging because the validity coefficients are relatively consistent with the existing literature in the domestic context, as is the incremental validity for assessment center dimensions, although some of the validity coefficients operated differently than expected.

References


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.

Received February 27, 2002
Revision received December 13, 2002
Accepted December 19, 2002