Personality and Intelligence Correlates of Assessment Center Exercises

Adrian Furnham* & John Taylor
University College London
Tomas Chamorro-Premuzic
Goldsmiths College, University of London

*Adrian Furnham; Department of Psychology; University College London; 26 Bedford Way; London WC1H OAP; a.furnham@ucl.ac.uk (email).

ABSTRACT - One hundred and eighty-eight prospective managers completed a Five Factor personality test (NEO-FFI) and an intelligence test (Ravens Progressive Matrices) and then a short quarter-day assessment exercise with five tasks, two more time-pressured and three less time-pressured. The IQ test was correlated significantly with Openness and three of the five assessment tasks. Only Openness was correlated with the assessment center task scores. The five tasks were factored and two emerged, one related to greater time-pressured, short tasks (proof reading and problem solving) and three to less time-pressured tasks. A regression showed bright, open, and low conscientious people did best at the timed task. Overall, intelligence was the best predictor of the various assessment center tasks.

This study is concerned with how well personality and intelligence test scores predict assessment center task scores. Assessment centers are widely used both for development and selection processes (Woodruffe, 1995). When reviewers have compared the predictive validity of various assessment methods from references, work samples, personality tests, cognitive ability tests, assessment center scores (etc) most have found assessment centers have among the best validity statistics (Anderson, & Cunningham-Snell, 2000). Most assessment centers are designed to measure highly specified, and often unique, organizational “competencies” by a variety of methods: e.g., subjective self-reports, objective performance tests, and observational data. Traditionally they are expensive in terms of time, effort and
money but given their predictive validity usually considered worth the effort (Krause, Kersting, Heggstedt & Thornton, 2006).

A good deal has been written about the history and philosophy of assessment centers (Woodruffe, 1995). Whilst they have changed somewhat over the years, they have tended to maintain their original purpose and processes (Howard, 1997). Inevitably there has been a great deal of interest in the construct, predictive and incremental validity of assessment centers (Donahue, Truxillo, Cornwell & Gerrity, 1997; Kleinmann & Koller, 1997; Kudisch, Ladd & Dobbins, 1997; Thornton, Tziner, Dahan, Clevenger & Meir, 1997).

Whilst some of these centers used established, validated tests most celebrate the fact that the tests/tasks devised for the center are particularly relevant or salient to the group of individuals tested and the job being advertised. That is, they are tailor-made to assess a range of competencies by a range of methods including self-report, observation and test-taking. This also means that until established by empirical research they only have face validity and not predictive validity which is why indeed they are constructed. To do this research a robust, sensitive and salient criterion, or better criteria, need to be found, measured and related to assessment center task scores to demonstrate their validity. However, frequently assessment centers are commissioned by organizations in the belief that their face validity and relationship to the job analysis-based requirements of the job must mean they are valid. That is validity is assumed, not investigated or empirically established.

Over the past 20 years various reviewers have tried to compare the validity of different assessment methods such as cognitive ability tests, interviews, work samples, etc. (Anderson & Cunningham-Snell, 2000; Cook, 2004; Schmitt, 1989). Many reviewers place assessment centers above other methods (like personality tests) in terms of their validity. However, it is not always possible to separate out different methods because more complex assessment centers tend to use cognitive ability and personality tests as part of the process to derive data.

There have been many studies that have examined personality, cognitive ability and assessment center ratings. Some have looked very specifically at the relationship between personality trait scores and assessment center ratings (Craik et al., 2002; Furnham & Crump, 2005; Haaland & Christianse, 2002; Kolk, Born & Van der Flier, 2004; Lievens, De Fruyt & Van Dam, 2001; Riggio, Mayes, & Schleicher, 2003; Waldman & Korbar, 2004). They have tended to show significant, but low and predictable, correlations between self-report derived trait scores and observer ratings on assessment center tasks. Other studies have looked at the combined relationships of cognitive ability and personality trait scores on assessment center ratings (Carless & Allwood, 1997; Dayan, Kasten & Fox, 2002; de Meijer, Born, Terlouw & Molen,
2006; Hardison, 2006; Krajewski et al., 2006; Lievens, Harris & Van Keer, 2003; Moser, Schuler & Funke, 1999; Spector et al., 2000). Most have been concerned with the relationship between the individual difference measures and the assessment ratings, but also the incremental validity of the latter over the former in predicting work outcomes. Other studies have looked at the incremental validity of cognitive ability over assessment center exercises in predicting specific outcomes (Goldstein et al., 1998).

This study seeks to explore the relationship between three types of tests: a well known measure of cognitive ability, one of personality, and assessment center task performance. Recently there have been a number of attempts to explore the relationship between personality and intelligence and how these two marginally related factors predict educational and work related behavior (Chamorro-Premuzic & Furnham, 2005). It is evident now from review and meta-analytic studies that both cognitive ability (Bertua, Anderson & Salgado, 2005) and personality tests (Bono & Judge, 2003) logically and systematically predict job performance criteria. This study explores the relationship between two very well established psychometric tests and five tests designed specifically for an assessment center to measure middle management competencies.

The Assessment Center tasks were designed specifically for this selection exercise based on a job analysis of positions that the candidates were applying for as well as the “competency list” of qualities the selectors were looking for. The brief was to design exercises specifically aimed at assessing managerial decision making under pressure; drafting clear and succinct reports without error and prioritization. This is a standard approach to the development of assessment center exercises even though it means they only have face validity. In this instance selection relied on an extensive structured interview, personal references as well the data from the tests mentioned in this study to make their decision. The test data from the assessment center was thought to be particularly relevant to assess ability to cope with the intellectual demands of management specifically strategic and tactical problem solving. Other relevant competencies such as interpersonal skills were thought to be adequately assessed in the interview process.

In their study of the validity of personality, cognitive ability and dimensions measured by an assessment center, Lievens et al (2003) tested various hypotheses based on the Big Five personality factors. Based on that study, as well as the work of Chamorro-Premuzic and Furnham (2005), it was predicted that (H1) Neuroticism would be negatively correlated, but (H2) Openness and (H3) Intelligence positively correlated with all Assessment Center Test scores, which were all essentially ability-type measures. That is bright, stable, open individuals tend to do well on decision
making tasks. Neuroticism has been demonstrated to be related to poor decision making because unstable individuals tend to be flooded with anxiety particularly when making decisions under pressure. Openness on the other hand is often associated with curiosity and imagination which usually helps decision making. Further it tends to be correlated with intelligence which is nearly always associated with assessment center tasks particularly those associated with problem solving and decision making.

**Method**

**Participants**

There were 188 participants, of which 128 were male. They ranged in age from 23 to 59 years, the mean being 36.52 (SD = 8.83) years. They were all candidates for a (British) government, senior management, job concerned with criminal intelligence. All had previous work experience, many in the police-force and were required to complete tests and the assessment center exercises as part of their selection process.

**Instruments**

**Personality.** The NEO-FFI (Costa & McCrae, 1992). This 60 item, non-timed questionnaire, measures the Big Five personality factors, i.e. Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness. Items involve questions about typical behaviors or reactions, which are answered on a five-point Likert scale, ranging from “strongly disagree” to “strongly agree”. The manual shows adequate indices of reliability and validity (see Costa & McCrae, 1992). Together with the NEO PI-R, this inventory is probably currently the most frequently used personality measures in research circles (Howard & Howard, 2001; Matthews, Deary & Whiteman, 2003).

**Intelligence.** Standard Progressive Matrices – Sets A – E (Raven, 1965). This is a test of nonverbal or abstract reasoning widely regarded as one of the best measures of general intelligence (Jensen, 1980; Kline, 1994). Participants were given 30 minutes to complete the 60 items which come in five blocks. It is one of the most widely used of all intelligence tests (Lynn, Allik & Irwing, 2004).

**Assessment Center Tasks**

These were devised and piloted by a senior Human Resources consultant with experience in the area. His brief was to devise ability type tasks that would assess decision making and document drafting under pressure. Each was approved (and slightly modified) by the organization who decided to use an assessment center. It
was decided both to have tasks of different length and complexity but also to ensure that some had a real sense of time pressure. All however had clear time limits.

**Intray.** Candidates were given 15 fairly typical managerial problems that may be encountered by this group. The situations were targeted at testing managerial skills such as delegation, performance management, team building, etc. Candidates were given 45 minutes to write down solutions for each. A single efficiency score was the result.

**Prioritize.** Candidates were also asked to go through the 15 situations and prioritize them for their attention. That is, they were to rank order them according to the effort, energy and resources that they should dedicate to each. They were given 5 minutes to complete this task after doing the solution above. This was measured against a “correct” order list.

**Problem solving.** Candidates were each given a file with 30 – 35 papers relating to a particular tricky and complex business problem. They were asked to provide a written report summarizing the nature of the problem as well as recommendations for a solution. The task attempted to measure analytical skills. Candidates were given 45 minutes for this task.

**Document Drafting.** The problem solving exercise was concerned with the quality of solutions. This exercise was concerned with drafting/writing skills. The task was to write a clear, succinct account of the candidates understanding of the problem and the solution. They had 25 minutes to complete this task. An overall judgment was made on the draft provided.

**Proof-Reading.** Candidates were given a four hundred word paper that contained 25 grammatical and spelling mistakes. They had five minutes to identify and correct as many errors as they spotted.

All evaluations were made by the same assessor though a reliability check was undertaken to ensure scoring was consistent. The assessor was the person who designed the exercises. Around ten percent of the completed exercises were independently rated by a second expert in this area to ensure reliability. There was little or no disagreement with extreme score (i.e. 1 or 5) and overall levels of reliability exceeded .90. Each task was evaluated on a five point scale where 1 was excellent and 5 poor to indicate overall performance on that task.

**Procedure**

Candidates arrived at the Assessment Center and were typically tested in groups of 6 to 10. It took 2 years to collect the data for this study. The examiner explained the nature of the assessment exercises including the personality and intelligence tests. There were a number of examiners present each time. Overall it tended to take
Table 1

Correlations Between All Variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Sex</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 Age</td>
<td>-.22**</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 Ravens</td>
<td>.05</td>
<td></td>
<td>-.41***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Neuroticism</td>
<td>.10</td>
<td>.03</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Extraversion</td>
<td>-.02</td>
<td>.03</td>
<td>-.06</td>
<td>-.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Openness</td>
<td>.02</td>
<td>-.20**</td>
<td>.17*</td>
<td>-.17*</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Agreeableness</td>
<td>.19**</td>
<td>-.10</td>
<td>.06</td>
<td>-.17*</td>
<td>.18*</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Conscientiousness</td>
<td>.12</td>
<td>-.05</td>
<td>.04</td>
<td>-.28**</td>
<td>.43***</td>
<td>.09</td>
<td>.25**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 Intray</td>
<td>.00</td>
<td>.10</td>
<td>-.20***</td>
<td>.00</td>
<td>.05</td>
<td>-.10</td>
<td>.05</td>
<td>-.16*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Prioritize</td>
<td>.07</td>
<td>.17*</td>
<td>.09</td>
<td>.05</td>
<td>.06</td>
<td>-.09</td>
<td>-.04</td>
<td>.08</td>
<td>.18*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 Problem Solve</td>
<td>.07</td>
<td>.32***</td>
<td>-.44***</td>
<td>.06</td>
<td>.12</td>
<td>-.21*</td>
<td>.08</td>
<td>.04</td>
<td>.17*</td>
<td>.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 Drafting</td>
<td>-.10</td>
<td>.09</td>
<td>-.07</td>
<td>.08</td>
<td>.10</td>
<td>-.13</td>
<td>.03</td>
<td>.02</td>
<td>.29***</td>
<td>.17*</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>13 Proofing</td>
<td>.00</td>
<td>.17*</td>
<td>-.30***</td>
<td>-.09</td>
<td>.00</td>
<td>-.12</td>
<td>.07</td>
<td>.17*</td>
<td>.01</td>
<td>.02</td>
<td>.19*</td>
<td>.05</td>
</tr>
</tbody>
</table>

Because the N varies between 108 and 172 causes some identical correlations (r values) to be significant and others not to be significant; + The Assessment Center exercises are marked on a 5 point scale where 1 = “Very Good” and 5 = “Very Poor”

***p < .001 **p < .01 *p < .05
half a day to complete all measures. Also participants’ assessment center scores were marked blind to the intelligence and personality test scores. Candidates were later full debriefed as to their performance on all of the tests.

**Results**

Table 1 shows the correlations between the participants’ demography, personality, intelligence and assessment center scores.

Females scored significantly higher on Agreeableness. Older participants, who did less well on the Ravens Intelligence Test, were less Open-to-Experience and scored lower on the Problem solving task. This confirmed H2 and H 3. The Ravens was significantly correlated positively with only one personality trait, namely Openness, though it also correlated significantly with three of the five assessment center tasks: in-tray, problem solving and proofing. As may be expected Open individuals tended to do better on these tasks. None of the correlations between the five assessment center task scores and Neuroticism, Extraversion and Agreeableness were significant. Openness correlated significantly with only one task and Conscientiousness two, but correlations were in opposite directions.

Table 2 shows the results from five hierarchical regressions where demographics (age and sex; Step 1), cognitive ability (Ravens; Step 2), and the big five (Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness; Step 3) were used as predictor variables of each of the criterion variable tasks: in-tray, prioritizing, problem-solving, drafting and proof-reading exercises. Individual difference factors and demographics were significant predictors of in-tray, problem solving and proof reading, but not drafting and prioritizing. Specifically, it can be seen that cognitive ability predicted performance on these three assessment center exercises even when demographic variables were taken into account, explaining 6%, 9%, and 8% of unique variance in in-tray, problem solving, and proof reading, respectively. Although Conscientiousness explained unique variance in in-tray and proof reading exercises even when demographics and cognitive ability were taken into account, personality as a whole made no significant contribution in the prediction of assessment center exercises. This is evidenced in the Adj.$R^2$ value for Step 3, which are either the same or merely marginally higher than for Step 2.

Table 1 shows that the five assessment center task scores were modestly inter-correlated, with five being significant. A VARIMAX rotated factor analysis showed two clear factors that accounted for 55% of the variance. The three tasks that loaded on the first factor (Intray, drafting, prioritizing) were all strictly timed tasks while the two that loaded on the second factor (Proofing, problem-solving) placed candidates
under less time pressure. The first factor tasks seemed specifically related to problem solving capacity, while the second documentation.

**Table 2**

<table>
<thead>
<tr>
<th>Regressions Showing IQ Personality and Gender as Predictor Variables and Assessment Center Exercises as Criterion Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-Tray</strong></td>
</tr>
<tr>
<td>Beta</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>F(2,171)</td>
</tr>
<tr>
<td>Adj R²</td>
</tr>
<tr>
<td>Ravens</td>
</tr>
<tr>
<td>F(3,170)</td>
</tr>
<tr>
<td>Adj R²</td>
</tr>
<tr>
<td>Neur</td>
</tr>
<tr>
<td>Ext</td>
</tr>
<tr>
<td>Open</td>
</tr>
<tr>
<td>Agree</td>
</tr>
<tr>
<td>Con</td>
</tr>
<tr>
<td>F(8,165)</td>
</tr>
<tr>
<td>Adj R²</td>
</tr>
</tbody>
</table>

Note: *** p < .001; ** p < .01; * p < .05; all Beta values are standardized and from step 3 (thus they take into account the other 7 predictors)

Scores on the two factors were then combined into two factor scores and the regression re-run. The first regression was not significant, but the second was \( F(8,97) = 5.52 (p < .001) \) and accounted for a quarter of the variance. It showed bright (highly intelligent), \( (\text{Beta} = -.48, \ t = 4.84, \ p < .001) \) Open, \( (\text{Beta} = -.18, \ t = 2.03, \ p < .05) \) and low Conscientious \( (\text{Beta} = .22, \ t = 2.13, \ p < .05) \) people all tended to score higher on the less time pressurized tasks.

A final regression was then performed with the arithmetically combined assessment center score for all five tests as the criterion variable. The range of scores was from 8 to 22 with a mean of 15.2 \( (SD = 3.12) \). It measures overall success at the assessment center. The regression was significant \( F(8,104) = 3.67, \ p < .01, \ Adj \ R^2 = 17 \) and showed that intelligence \( (\text{Beta} = -.35, \ t = 3.33, \ p < .01) \) was the only significant predictor of the totaled Assessment Center Score.
Discussion

This study has demonstrated three things about the assessment of adults through psychometric tests and assessment center tasks. The first is that age is negatively related to ability test performance. This is well known and applies particularly to short timed tests of fluid ability (Chamorro-Premuzic & Furnham, 2005). Clearly this needs to be taken into consideration in evaluating candidates fairly, particularly those in late middle age.

Second, assessment center exercises that are designed to measure specific competencies should be differentiated in terms of whether exercises put people under considerable time pressure, as well a how long the tests last. It is possible the design of the task may be highly relevant in the outcomes. Third, as has been shown many times before, personality and intelligent tests, which themselves are modestly related, predict assessment center exercise ratings. The results of this study do confirm that of others like Lievens et al. (2003) who also found cognitive ability (most) but also Openness as significant positive predictors of various assessment center tasks. Thus, on an individual assessment center exercise demography, ability and personality scores can account for between 10% and 20% of the variance. However, once exercise scores are combined to presumably provide a more robust and reliable score these factors can account for a quarter of the variance. The results confirm all the data from meta-analyses in this area (Chamorro-Premuzic & Furnham, 2005) that the three factors most likely to be implicated are intelligence, Openness and Conscientiousness. Interestingly, in this study low Conscientiousness was predictive of assessment test success, not high Conscientiousness. The explanation for this perhaps counter-intuitive result lies in the work of Moutafi et al. (2006) who found Conscientiousness consistently negatively related to intelligence.

Ideally this study would have had some criteria variables like training performances, skill acquisition or job productivity so that the predictive validity of the assessment exercises could be established. In this sense this study was a simple correlational study between three types of measures. The assessment center exercises while fairly typical in this area, were at this stage of the research not related to any robust set of behavioral work criteria to demonstrate validity. Further it would then be particularly interesting to determine the incremental validity of these scores over and above the trait and cognitive ability tests. Again, this requires a set of criteria to demonstrate assessment center task validity.

However, this modest study did show two things of interest. First, that because many, but certainly not all, assessment center tasks are power tests (that is they are similar to ability tests) they are more closely correlated with other standard intelligence tests. It may well be that preference tests (personality type tests) are
more closely related to stylistic measures such as how participants chair meetings or give presentations. Alternatively this may indicate method rather than content based correlations. This is an important issue that warrants further research. Second, and related to the above, the factor analysis of the five assessment center exercises suggested that tasks may be related based on how much time pressure people are put under. Whilst there may be different interpretations to the factor themes in this paper, what does seem to be the case is that to avoid Type II error it is advisable to examine the structure of all assessment center task scores to understand how they are related.

References


