Moderator effects of job complexity on the validity of forced-choice personality inventories for predicting job performance

Jesús F. Salgado*
University of Santiago de Compostela, Spain

A B S T R A C T

Research has demonstrated that job complexity moderates the validity of general mental ability (GMA), the relationship between personality and job satisfaction, and the relationship between GMA and job satisfaction. However, no published research has investigated whether job complexity moderates the criterion validity of the Five-Factor Model (FFM) of personality for predicting job performance. This paper reports a meta-analytic examination of the moderator effects of job complexity on the criterion validity of the FFM of personality as assessed with forced-choice inventories. In accordance with the hypotheses, the results showed that job complexity moderates negatively the validity of conscientiousness and emotional stability and that it moderates positively the validity of openness. The implications for personnel selection research and practice are discussed.

© 2017 Published by Elsevier España, S.L.U. on behalf of Colegio Oficial de Psicólogos de Madrid. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Efectos moderadores de la complejidad del puesto sobre la validez de los cuestionarios de personalidad de elección forzosa para predecir el desempeño

RESUMEN

La investigación ha demostrado que la complejidad del puesto de trabajo modera la validez de la capacidad mental general (CMG), la relación entre la personalidad y la satisfacción en el trabajo y las relaciones entre la CMG y la satisfacción en el trabajo. Sin embargo, no se ha publicado ninguna investigación que haya examinado si la complejidad del puesto de trabajo modera la validez de criterio del modelo de los cinco grandes factores (MCGF) de personalidad para predecir el desempeño en el trabajo. Este artículo presenta un metaanálisis sobre los efectos moderadores de la complejidad del puesto en la validez del MCGF de personalidad cuando se emplean cuestionarios de elección forzosa (CEF). De acuerdo con las hipótesis planteadas, los resultados muestran que la complejidad del puesto modera negativamente la validez de criterio de los factores de responsabilidad y de estabilidad emocional y positivamente la validez del factor de apertura a la experiencia. Finalmente, se plantea algunas posibles implicaciones para la teoría y la práctica de la selección de personal.

© 2017 Publicado por Elsevier España, S.L.U. en nombre de Colegio Oficial de Psicólogos de Madrid. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Recent surveys have shown that personality inventories are popular instruments for making personnel decisions in the United States (US) and the European Union (EU) (Alonso, Moscoso, & Cuadrado, 2015; Tett, Christiansen, Robie, & Simonet, 2011; Zibarras & Woods, 2010) and research on personality at work has also shown they are very useful procedures for predicting important organizational criteria. For example, personality measures predict job performance, training proficiency, counter-productive behaviors, well-being, accidents, productivity data, salary, promotions, and occupational attainment, among other work criteria (Barrick & Mount, 1991; Barrick, Mount, & Judge, 2001; Clarke & Robert-

In the domain of personality at work, the Five-Factor Model (FFM) of personality (i.e., emotional stability, extraversion, openness to experience, agreeableness, and conscientiousness) has received more attention than any alternative model. The extant meta-analytic evidence has demonstrated that conscientiousness and emotional stability generalized validity across samples, criteria, occupations, and countries, and that the other three personality dimensions were valid predictors for specific criteria and specific occupations. For example, openness to experience predicted training proficiency, and extraversion and agreeableness predicted performance in occupations characterized by a large number of interpersonal relationships (e.g., Barrick et al., 2001; Judge, Rodell, Klinger, Simon, & Crawford, 2013).

Nevertheless, agreement is not unanimous about the relevance of personality measures for personnel selection. For example, Murphy and Dziewczynski (2005) posited that the theories linking personality constructs and job performance were often vague and unconvincing, that little was known about how to match personality dimensions and occupations, and that some of the most valid personality-related measures (e.g., integrity tests) included poorly defined constructs. On the other hand, researchers suggested that the validity of personality measures was small and that the measures based on self-reports can be faked, independently of the administration mode (Grieve & Hayes, 2016; Morgeson et al., 2007a, 2007b; Salgado, 2016).

In part, these criticisms have been contradicted by recent research that showed that (1) the format of the personality inventories is an important moderator of the criterion-related validity of the Big Five dimensions (Salgado, Anderson et al., 2015; Salgado & Tauriz, 2014), (2) the facets of the Big Five do not show evidence of criterion-related validity for predicting job performance when the variance of the facets is residualized (Salgado, Moscoso, & Berges, 2013; Salgado, Anderson et al., 2015), and (3) there is robust evidence of the construct validity of the FFM (e.g., Judge et al., 2013). With regard to the first issue, Salgado and Tauriz (2014) and Salgado, Anderson et al. (2015) found that criterion-related validity increased noticeably when quasi-ipsative forced-choice (QFC) formats are used. For example, the operational validity of conscientiousness was found to be .39 when a QFC format was used. In addition, some empirical evidence showed that the forced-choice (FC) format can be more resistant to faking than the most frequently used formats, such as Likert’s (Jackson, Wrobleski, & Ashton, 2006; Nguyen & McDaniel, 2000).

Therefore, there is currently empirical evidence that the FC personality inventories are valid predictors of job performance and that they are also widely used in organizations for making personnel decisions. However, no previous research has examined the potential moderator effects of job complexity on the FC inventories as a unique category, nor have the moderator effects for the particular types of FC inventories (i.e., normative, ipsative, and quasi-ipsative) been examined.

The objective of this study is to shed light on this issue that has been ignored in the meta-analytic research conducted to examine the validity of the FC personality inventories. Consequently, the main goal of this study is to meta-analytically examine whether job complexity is a moderator of the criterion validity of FC inventories. The second goal is to check whether job complexity has similar effects for the three types of FC personality scores which can be obtained from FC inventories. Thus, the main contribution of this paper lies in highlighting the role that job complexity plays in the validity of FC personality inventories for predicting job performance.

**Forced-Choice Personality Inventories**

The first FC personality inventories were developed during the 1940s and 1950s (Hicks, 1970) and the FC models used in those days have remained relatively unchanged until now. Usually, the FC method asks the individual to make a choice between several alternatives, most frequently three or four. In order to make the decision the individual must indicate what alternative he/she likes most and what alternative he/she likes least when those alternatives are applied to the individual. The alternatives are paired in terms of similar levels of social discrimination and preference. Therefore, the FC method distinguishes from the most typical personality assessment methods, such as Likert, True-False, Agree-Indecisive-Disagree (collectively called single-stimulus [SS] methods), in that the individual has to make a choice between two or more alternatives rather than to rate each single statement or phase as is typically done with SS personality inventories.

Even though the FC method always consists of a choice between alternatives, the FC inventories can produce three types of scores depending on how the choice is made (Cattell, 1944; Clemans, 1960; Hicks, 1970). The FC personality inventories can result in normative, ipsative, and quasi ipsative scores (see Salgado, Anderson et al., 2015, and Salgado & Tauriz, 2014, for a detailed account of these three scores). This contrasts with the SS personality inventories which always produce normative scores. Therefore, it is important to take into account the score type produced by the FC inventory because each of them has important psychometric characteristics.

The normative scores allow comparisons among individuals and groups on each personality variable. Therefore, they are inter-individual scores. The ipsative scores are dependent on the individual level in the other variables included in the choice. Consequently, ipsative scores permit the comparison of one individual across different personality factors. In other words, the ipsative scores are intra-individual ones. The quasi-ipsative scores allow comparisons between individuals and between groups, but produce simultaneously some degree of dependence among the variables assessed.

Several characteristics of the FC personality inventories which are relevant for personnel assessment have to be mentioned. First, they appear to correlate with general mental ability (GMA) when individuals respond as job applicants (Vasilopoulos, Cucina, Dyomina, Morewitz, & Reilly, 2006), so they can be more cognitively loaded than the typical SS personality formats (e.g., Likert, Yes, No). Therefore, the validity of the FC personality measures might be moderated by job complexity, as this variable also moderates the validity of GMA. Second, the FC-based measures may produce gender differences in some cases and, consequently, equal opportunities may also be negatively affected (Anderson & Slep, 2004). Third, FC personality inventories showed stronger resistance to faking than SS personality inventories, although they are not totally unaffected by faking (Jackson et al., 2000; Nguyen & McDaniel, 2000). Fourth, recent advances in IRT methodology have produced methods for recovering normative scores from ipsative scores (Brown & Maydeu-Olivares, 2011; Chernyshenko, Stark, Drasgow, & Roberts, 2007; Heggestad, Morrison, Reeve, & McCoy, 2006; Maydeu-Olivares & Brown, 2010; McCoy, Heggestad, & Reeve, 2005; Stark, Chernyshenko, & Drasgow, 2005; Stark, Chernyshenko, Drasgow, & Williams, 2006). Fifth, they are currently used in around 30% of organizations, according to a survey conducted by Tett et al. (2011).
Job Complexity

Hunter and Hunter (1984) defined job complexity as the cognitive difficulty of the requirements and demands of the occupation. This definition can be extended to include not only the number of complex tasks that should be solved, but also the degree to which these tasks are not repetitive, to what extent the goals are hard to define, the number of opportunities that exist for making personal decisions, and the degree to which the procedures of problem solving are not standardized. According to Hunter and Hunter, job complexity is to a great extent captured in the “data” dimension of Fine’s (1955) Functional Analysis, which has defined three extensive occupational families. Other two smaller occupational families were defined by Fine’s “things” dimension. Based on this scheme, Hunter and Hunter classified occupations in five big occupational families using the Dictionary of Occupational Titles (DOT; U.S. Department of Labor, 1977/1991). These occupational families were: design (I), feed (II), synthesize (III), coordinate, analyze/compile/calculate (IV), and compare/copy (V). According to their degree of complexity, the most complex would be family I; families III and IV would be of an intermediate level of complexity, and II and V would be of low complexity. The percentage of job positions that would exist in each complexity family would be 2.5, 2.4, 14.7, 62.7, and 17.7 percent for families I, II, III, IV, and V, respectively. This means that approximately 17.2% of the occupations would be of high complexity, 62.7% would be of intermediate complexity, and 20.1% would be of low complexity. Later, Hunter and Hunter’s classification was reduced to three levels of complexity by Hunter, Schmidt, and their colleagues (Hunter, Schmidt, & Judiesch, 1990; Schmidt, Shaffer, & Oh, 2008).

Job complexity is a construct that has been shown to be relevant as a moderator of the relationships between a series of individual differences variables (e.g., intelligence and personality) and also organizational variables (e.g., job performance and job satisfaction). For example, Hunter and Hunter (1984) demonstrated that job complexity moderated the validity of GMA, so higher complexity was associated with larger validity coefficients of GMA for predicting both job performance and training proficiency. This finding was later replicated by other researchers (e.g., Hartigan & Wigdor, 1989; Salgado et al., 2003; Schmidt et al., 2008). Ganzach (1998) found that job complexity moderates the relationship between intelligence and job satisfaction. Wilk and Sackett (1996) demonstrated that there is a positive relationship between job complexity and cognitive ability for people in the process of labor mobility. Judge, Bono, and Locke (2000) found that job complexity moderates the relationship between personality and job satisfaction. The moderator effects of job complexity have also been examined for the criterion validity of situational interviews (SI) and behavior-description interviews (BDI) by Huffcut, Conway, Roth, and Klehe (2004), who found moderator effects of job complexity for SIs but not effects for BDIs. However, no meta-analysis has been published to examine whether job complexity moderates the validity of personality measures in general and of those FC inventories in particular.

Job Complexity-Personality Relationships: Reasons to Expect Moderator Effects

There are theoretical and empirical reasons to expect a moderator effect of job complexity on the validity of FC personality inventories. First, if the wide definition of job complexity mentioned above is used, i.e., the one that includes not only cognitive factors but also situational and personal ones, it could be argued that some personality factors could be related to job complexity and, therefore, job complexity could moderate the validity of these personality factors for predicting job performance. For example, Judge et al. (2000) have demonstrated that there is a moderator effect of job complexity on the relationship between personality and job satisfaction, which suggests that such moderator effects could be present in other relationships in which personality could be a determinant, such as for example job performance. In addition, some researchers have suggested that conscientiousness might not predict performance well for all occupations and situations. For example, it could be contraindicated in those occupations in which innovation, creativity, or leadership are important (Costa, Pérez, Sánchez, Garaiordobil, & Gondim, 2015; Hülsheger, Anderson, & Salgado, 2009; Hough, 1997, 1998; Robertson, Baron, Gibbons, Maclver, & Nyfield, 2000; Robertson & Callinan, 1998). Given that creativity, innovation, and leadership are elements which would contribute to or reflect the complexity of a job, then job complexity would produce a negative effect on the validity of the conscientiousness dimension for predicting job performance. The implications of job complexity for emotional stability seems obvious if we take into account that the most cognitively complex jobs also tend to be characterized by their occupants having more control over and ability to decide on their behaviors and the consequences of their actions. This is the case, for example, for managers and professionals (e.g., physicians, attorneys). However, less complex jobs are more often subjected to unpredictable situations and external control. When control is associated with experiencing or not experiencing stress or having or not having emotional adjustments at work, the potential relationship between emotional stability and job complexity also seems clear. For example, the literature on “learned helplessness” is illustrative of this case (Abramson, Seligman, & Teasdale, 1978; Seligman, 1975). People who define situations as unpredictable and that believe that they do not control the outcomes show poorer performance, less motivation, and more negative emotional reactions. Finally, given that openness to experience shows some correlation with GMA (Judge, Jackson, Shaw, Scott, & Rich, 2007; McCrae, 1996) and that the cognitive difficulty of the tasks is an important element for defining job complexity, one would expect that job complexity moderates positively the relationship between openness and job performance, in contrast to what occurs with the dimensions of conscientiousness and emotional stability. With regard to extroversion and agreeableness, it does not seem that job complexity moderates its validity or it is at least conceptually difficult to speculate over possible relations.

Finally, as some authors (e.g., Vasilopoulos et al., 2006) have suggested that FC personality inventories can be more cognitively loaded than SS inventories, and due to the fact that job complexity moderates the relationships between GMA and job performance, some effects of job complexity on the validity of FC inventories can be expected because of their cognitive variance. Taking all these reasons into account, the following two hypotheses can be posited:

H1. Job complexity negatively moderates the validity of conscientiousness and emotional stability, so that their validity will be lower for occupations with higher complexity levels.

H2. Job complexity positively moderates the validity of the openness to experience dimension, so that the validity will be greater for the occupations with higher complexity level.

Method

Literature Review

Exhaustive manual and computer-assisted searches were carried out to identify validity studies in which FC personality inventories were used and in which job performance was the criterion. The examined period covers from January 1960 to December
2015. The search was done using three strategies. First, the name of the most popular FC personality inventories (e.g., MBTI, OPQ, GPIP, EPPS) was used as a keyword in the computer searches. Second, I examined a number of proprietary electronic databases of well-known journal publishers, including Academy of Management, Ammons, Sage, Wiley, Springer, ProQuest, and Elsevier. In these searches, the terms ipsative, normative, partially-ipsative, quasi-ipsative, forced-choice, job performance, as well as the acronyms of the most popular personality inventories were used. Third, exhaustive searches were also done in Google and Scholar Google in order to identify unpublished papers. Fourth, manual searches were done in a dozen of the most important journals in which validity studies are frequently published (e.g., Journal of Applied Psychology, Journal of Occupational Psychology, Personnel Psychology, etc.). Fifth, I examined the list of references of the most highly cited meta-analyses (e.g., Bartram, 2007; Judge et al., 2013; Salgado, 1997, 2003; Salgado, Anderson et al., 2015; Salgado & Tauriz, 2014; Tet, Rothstein, & Jackson, 1991) to identify articles and unpublished papers not identified in the previous searches. Sixth, I wrote to several researchers who have conducted research on FC personality inventories and asked for published and unpublished studies. Finally, I reviewed the technical manual of the most popular FC personality inventories (e.g., DSD, EPPS, GPIP, and OPQ, among others). Once a preliminary data base of the articles, unpublished papers, technical reports, manuals, doctoral dissertations, and papers presented at conferences was created, I separated the studies that reported the relationship between personality measures and job performance and excluded the studies that used other alternative criteria (e.g., academic performance, training success, salaries, and counterproductive behavior at work). In other words, I did not include studies that used grades, instructor ratings, salaries, and reports of counterproductive behaviors as criteria. Next, I classified the scales of the inventories into the Five-Factor Model using the classification scheme of Salgado and Tauriz (2014).

If available, I recorded the following information for each study: (1) sample size, (2) occupational title, (3) type of FC measure (i.e., ipsative, quasi-ipsative, and normative), (4) personality measures, (5) criterion type, (6) predictor reliability, (7) criterion reliability, (8) range restriction in the predictor, (9) correlation between predictor and criterion, and (10) correlations among the predictors if more than one was used. In the case of studies with conceptual replications (e.g., two or more scales or inventories were used to assess the same personality factor), I used Mosier’s formula to form a linear composite (Schmidt & Hunter, 2015).

### Job Complexity Coding

Once the characteristics of the studies were recorded, the next step was to classify the occupations according to their respective level of job complexity. After an inspection of the jobs included in the data base, I decided to create three levels of complexity only, in accordance with the analytic strategy used by Hunter et al. (1990), Salgado et al. (2003), and Schmidt et al. (2008). The highest level of complexity consists of occupations coded 0 and 1 in the “data” dimension of the DOT and the occupations coded 0 in the “things” dimension. The medium complexity level includes the occupations coded 2, 3 and 4 in the “data” dimensions of the DOT. The lowest level of job complexity included the occupations coded 5 and 6 in the “data” dimension and the occupations coded 6 in the “things” dimension. With this classification system, the jobs included in the occupational categories of engineers, counselors, managers, lawyers, pilots, military commanders, and professionals were included in the category of medium complexity. Finally, jobs included in the occupational categories of unskilled workers, maintenance personnel, clerical assistants, and soldiers were classified in the category of low complexity level. The classification resulted in 56 coefficients for conscientiousness, 47 for emotional stability, 47 for extraversion, 42 for agreeableness, and 40 for openness. The studies included in each complexity level-FC format combination appear in the Appendix.

### Meta-analysis Method and Artifact Distributions

After the studies were classified, the following step was to apply the psychometric meta-analysis method of Schmidt and Hunter (2015), implemented in a software program by Schmidt and Le (2014), which uses the correction for indirect range restriction (IRR). Sampling error, predictor reliability, criterion reliability, and range restriction (RR) were considered as artifacts that affect the validity size as well as the variance of the reliability coefficients. For the criterion reliability, I used a mean value of .52 (SD = .09) which was reported by Salgado, Moscoso, and Anderson (2016; see also Salgado & Moscoso, 1996), that is also the same value found by Viswesvaran, Ones, and Schmidt (1996; see also Salgado, 2015). In the case of the reliability of the measures of the Big Five factors of personality, I used the estimates reported by Salgado and Tauriz (2014) and Salgado, Anderson et al. (2015) for the FC personality inventories.

With regard to RR, I created a distribution for each complexity level. There were two reasons for proceeding in this way. First, previous research revealed that the degree of RR is not the same across the three levels of job complexity (Salgado et al., 2003; Schmidt et al., 2008). The second reason is that the three complexity levels contain occupational categories which practically do not overlap, thereby they may have different distributions of the RR values. Therefore, if only one distribution of RR was to be used for all the jobs, then they might result in an underestimation of validity for the highly complex jobs and an overestimation of the validity for the medium and low complexity jobs. The average and the standard deviation of the distributions appear in Table 1. As can be seen in this table, the RR values of low complexity jobs are a bit lower than the values found for medium and high complexity levels. It must be noted here that RR values were practically the same for the five personality factors in each complexity level. For all the above reasons, I grouped all the coefficients of the Big Five for the same complexity level.

![Table 1](http://www.elsevier.es)
Table 2

<table>
<thead>
<tr>
<th>Personality Dimension</th>
<th>K</th>
<th>N</th>
<th>r_{xy}</th>
<th>SD_{r}</th>
<th>p</th>
<th>SDp</th>
<th>%VE</th>
<th>90%CV</th>
<th>CIU</th>
<th>CIL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Complexity – Normative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraversion</td>
<td>4</td>
<td>1994</td>
<td>-0.53</td>
<td>0.09</td>
<td>-0.089</td>
<td>.000</td>
<td>100</td>
<td>-0.089</td>
<td>-1.53</td>
<td>-0.25</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>4</td>
<td>1994</td>
<td>0.055</td>
<td>0.012</td>
<td>0.092</td>
<td>.000</td>
<td>100</td>
<td>0.092</td>
<td>0.125</td>
<td>0.076</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>5</td>
<td>2297</td>
<td>0.072</td>
<td>0.067</td>
<td>0.121</td>
<td>.076</td>
<td>51</td>
<td>0.023</td>
<td>0.218</td>
<td>0.023</td>
</tr>
<tr>
<td><strong>High Complexity – Ipsative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>20</td>
<td>4505</td>
<td>0.008</td>
<td>0.069</td>
<td>0.013</td>
<td>.029</td>
<td>93</td>
<td>-0.025</td>
<td>0.063</td>
<td>-0.037</td>
</tr>
<tr>
<td>Extraversion</td>
<td>19</td>
<td>4437</td>
<td>0.061</td>
<td>0.078</td>
<td>0.103</td>
<td>.068</td>
<td>73</td>
<td>0.016</td>
<td>0.162</td>
<td>0.044</td>
</tr>
<tr>
<td>Openness</td>
<td>19</td>
<td>4437</td>
<td>0.049</td>
<td>0.072</td>
<td>0.082</td>
<td>.046</td>
<td>85</td>
<td>0.023</td>
<td>0.137</td>
<td>0.028</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>19</td>
<td>4437</td>
<td>-0.17</td>
<td>0.082</td>
<td>-0.228</td>
<td>0.082</td>
<td>64</td>
<td>-0.133</td>
<td>-0.090</td>
<td>0.034</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>19</td>
<td>4437</td>
<td>0.055</td>
<td>0.083</td>
<td>0.091</td>
<td>0.081</td>
<td>65</td>
<td>-0.013</td>
<td>-0.153</td>
<td>0.028</td>
</tr>
<tr>
<td><strong>Medium Complexity – Ipsative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>7</td>
<td>1459</td>
<td>0.053</td>
<td>0.070</td>
<td>0.086</td>
<td>.015</td>
<td>98</td>
<td>0.068</td>
<td>0.171</td>
<td>0.001</td>
</tr>
<tr>
<td>Extraversion</td>
<td>7</td>
<td>1459</td>
<td>0.078</td>
<td>0.058</td>
<td>0.132</td>
<td>.000</td>
<td>100</td>
<td>0.132</td>
<td>0.005</td>
<td>0.060</td>
</tr>
<tr>
<td>Openness</td>
<td>6</td>
<td>1027</td>
<td>0.130</td>
<td>0.095</td>
<td>0.222</td>
<td>.094</td>
<td>65</td>
<td>0.010</td>
<td>0.123</td>
<td>0.093</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>5</td>
<td>1058</td>
<td>0.060</td>
<td>0.088</td>
<td>0.101</td>
<td>.089</td>
<td>62</td>
<td>0.093</td>
<td>0.070</td>
<td>0.023</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>8</td>
<td>1666</td>
<td>0.115</td>
<td>0.090</td>
<td>0.190</td>
<td>.092</td>
<td>60</td>
<td>0.072</td>
<td>0.284</td>
<td>0.087</td>
</tr>
<tr>
<td><strong>High Complexity – Quasi-Ipsative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>9</td>
<td>1237</td>
<td>0.028</td>
<td>0.181</td>
<td>0.046</td>
<td>.255</td>
<td>23</td>
<td>-0.280</td>
<td>-0.237</td>
<td>-1.145</td>
</tr>
<tr>
<td>Extraversion</td>
<td>5</td>
<td>1066</td>
<td>0.143</td>
<td>0.067</td>
<td>0.240</td>
<td>.000</td>
<td>100</td>
<td>0.240</td>
<td>0.339</td>
<td>0.142</td>
</tr>
<tr>
<td>Openness</td>
<td>5</td>
<td>1066</td>
<td>0.205</td>
<td>0.054</td>
<td>0.341</td>
<td>.000</td>
<td>100</td>
<td>0.341</td>
<td>0.420</td>
<td>0.262</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>5</td>
<td>1066</td>
<td>0.068</td>
<td>0.102</td>
<td>0.114</td>
<td>0.124</td>
<td>46</td>
<td>0.013</td>
<td>0.230</td>
<td>0.027</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>6</td>
<td>1113</td>
<td>0.130</td>
<td>0.100</td>
<td>0.216</td>
<td>0.108</td>
<td>56</td>
<td>0.077</td>
<td>0.348</td>
<td>0.083</td>
</tr>
<tr>
<td><strong>Low Complexity – Quasi-Ipsative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>6</td>
<td>848</td>
<td>0.145</td>
<td>0.187</td>
<td>0.239</td>
<td>0.258</td>
<td>27</td>
<td>-0.091</td>
<td>0.451</td>
<td>0.026</td>
</tr>
<tr>
<td>Extraversion</td>
<td>6</td>
<td>848</td>
<td>0.145</td>
<td>0.187</td>
<td>0.239</td>
<td>0.258</td>
<td>27</td>
<td>-0.091</td>
<td>0.451</td>
<td>0.026</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>7</td>
<td>674</td>
<td>0.169</td>
<td>0.141</td>
<td>0.280</td>
<td>0.159</td>
<td>52</td>
<td>0.077</td>
<td>0.453</td>
<td>0.107</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>11</td>
<td>1250</td>
<td>0.220</td>
<td>0.097</td>
<td>0.363</td>
<td>0.043</td>
<td>92</td>
<td>0.307</td>
<td>0.457</td>
<td>0.268</td>
</tr>
<tr>
<td>Note: K = number of independent samples; N = total sample size; r_{xy} = observed validity; SD_{r} = standard deviation of observed validity; p = validity corrected for measurement error in X and Y and indirect range restriction in predictor; SD_{p} = standard deviation of p; %VE = percentage of variance accounted for by artifactual errors; 90%CV = 90% credibility value based on p; CIU = upper limit of the 95% confidence interval of p; CIL = lower limit of the 95% confidence interval of p.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results

The results of the meta-analysis of the relationships between the five personality dimensions and job performance for the FC personality inventories appear in Table 2, classified according to the degree of complexity of the occupations. From left to right, the first two columns show the number of independent validations and the total sample size. The next four columns show the observed validity, the observed standard deviation, the validity corrected for measurement error in predictor and criterion, and for IRR. The last four columns report the variance accounted for by the four statistical artifacts, the 90% credibility value, and the 95% confidence interval of p.

In the case of the normative FC personality inventories, I found validity coefficients for the high complexity level only. Therefore, it was not possible to estimate the potential moderator effects of job complexity for this FC format. With very small differences, the results for the normative FC were practically the same ones reported by Salgado, Moscoso et al. (2015). The corrected validity is very small for the four personality factors included in the Table 2, although validity generalizes in the four cases, as the 90% credibility value is greater than 0.

With regard to the validity of the ipsative FC personality inventories, I found studies to compare the effects of job complexity for the occupations of high and medium complexity levels. The results indicate that validity is smaller for all the personality factors in the high complexity occupations than in the medium complexity ones. In other words, job complexity moderates validity negatively, so the larger the job complexity the lower the validity. In regard to the validity size, it is very small for the five personality factors in the case of high complexity occupations, with validity values ranging from -0.22 to 0.216 for agreeableness to 0.91 for conscientiousness. There is evidence of validity generalization for extraversion and openness only. The validity size of the Big Five for the ipsative FC personality inventories in the case of medium complexity occupations is small, too, ranging from 0.086 for emotional stability 0.222 for openness to experience. There is evidence of validity generalization for extraversion, openness, and conscientiousness, as the 90% credibility is greater than 0 in the three cases. These results partially supported Hypotheses 1 and 2.

In respect to the quasi-ipsative FC personality inventories, the meta-analytic findings show that job complexity moderates the validity, although the pattern of the results vary across the five personality factors. For emotional stability, the results show that the largest validity corresponds to the occupations of the medium complexity level, followed by the lower level, and with the smallest validity for the higher complexity level. Therefore, these findings partially supported Hypothesis 1. For the three levels of job complexity, the validity size is small, ranging from 0.046 (high complexity) to 0.203 (medium complexity). In the case of extraversion, the validity is practically the same for the occupations of high and medium complexity levels (p = 24), and is larger for the low complexity level (p = 29). Therefore, the relationship between job complexity and the validity for extraversion is negative, as the validity increases when job complexity decreases. In the case of openness, I found validity studies for occupations of high and
medium levels of job complexity. For this last personality factor, the validity was larger for the occupation of the high complexity level than for the medium level. Furthermore, it must be noted that the validity is considerable for the high complexity level ($r = .34$) and small and negative for the medium complexity level ($r = -.107$). Consequently, these findings totally support Hypothesis 2. The results for agreeableness show, to a certain extent, a pattern similar to the one found for emotional stability. The validity was noticeably larger for the medium level of complexity than for the other two complexity levels, but for agreeableness the second largest value was for the higher level of complexity and the validity was 0 for the lower level.

Finally, the results for conscientiousness are the most interesting for three reasons. First, they confirm that conscientiousness is consistently a valid predictor of job performance. Second, the pattern of the results showed that job complexity is an important moderator of the criterion validity of conscientiousness. The validity size ranged from .22 for the high complexity level to .43 for the lower level of complexity. The validity for the medium level was in the middle of these two. Therefore, job complexity moderated the validity of conscientiousness negatively. In other words, the findings revealed that conscientiousness is a much better predictor of job performance for the low level of job complexity than for the medium and high levels. In fact, the validity size for the low complexity level is greater than the validity of many of the most commonly used procedures in personnel selection (Schmidt & Hunter, 1998). Moreover, these findings support Hypothesis 1.

As a whole, the findings of this meta-analytic effort showed that job complexity is a relevant moderator of the validity of the five personality factors assessed with FC inventories when they are used for predicting job performance. The results supported Hypotheses 1 and 2, particularly in the case of the quasi-ipsative FC inventories. In second place, the findings showed also that there is not a single pattern of moderator effects, as can be seen in Figures 1 and 2. Finally, as conscientiousness occupies a special place among the personality factors due to the fact that it has been found to be the only factor that consistently predicted job performance, a specific comment about conscientiousness is in order. For both ipsative and quasi-ipsative FC personality inventories, the effects of job complexity run negatively with respect to the validity size. As job complexity increases, the validity of conscientiousness decreases.

Discussion

This study was designed to test two hypotheses, according to which job complexity would have negative effects on the validity of conscientiousness and emotional stability (Hypothesis 1), and positive effects on the validity of openness to experience (Hypothesis 2). The results have supported Hypothesis 1, as the criterion validity of emotional stability and conscientiousness was larger for occupations of the lower levels of job complexity than for the occupations of high job complexity. The findings also supported Hypothesis 2 for both ipsative and quasi-ipsative FC personality inventories, as openness to experience was shown to be a better predictor of job performance for the occupations of high complexity level.

More specifically, in the case of conscientiousness, job complexity moderates largely and negatively the criterion validity, so that the size of the validity is smaller for the occupations of high and medium complexity in comparison with low complexity. Furthermore, the hypotheses were supported for both ipsative and quasi-ipsative FC inventories. These findings support those authors who have suggested that conscientiousness might not be such a good predictor of performance in those jobs where more flexibility, initiative, creativity and, clearly, more cognitive complexity are required (Da Costa et al., 2015; Hülshéger et al., 2009; Hough, 1997; Robertson & Callinan, 1998; Robertson et al., 2000). For both emotional stability and conscientiousness, the relationship between job complexity level and criterion validity is described by a negative linear shape.

With regard to openness to experience, the results indicated that this personality factor is a valid predictor for high complexity jobs, but that it is not a predictor for occupations of medium and low complexity level. Therefore, the relationship is described by a positive linear shape. This finding totally agreed with Hypothesis 2. Consequently, this supports the assumption that, due to the relationship between openness and GMA, the highest value for openness validity should be for high complexity level occupations.

In the case of extraversion, job complexity has no effect on its criterion validity, as for both the ipsative and the quasi-ipsative FC inventories the validity size is practically the same across the job complexity levels. The smaller differences are probably due to sampling error.

The results for agreeableness are particularly interesting and unexpected (as they were not hypothesized). Many previous meta-analyses of the FFM found that agreeableness showed very low validity or it was not a predictor at all (Barrick et al., 2001; Judge et al., 2013; Salgado & Tauriz, 2014). However, this meta-analysis shows that agreeableness can be a valid and relevant predictor of job performance for the occupations of medium complexity level, if this factor is assessed with a quasi-ipsative FC inventory. For agreeableness, the effects of job complexity on the validity take the shape of an inverted U.
In summary, the findings suggest two relevant conclusions. First, job complexity is a robust moderator of the validity of the FFM for predicting job performance of the ipsative and quasi-ipsative FC personality inventories, with the exception of extraversion. Second, the pattern (shape) of the job complexity effect is not the same for all the personality factors. Job complexity can have positive, negative, and inverted U consequences for the validity size of emotional stability conscientiousness, openness, and agreeableness, respectively. Future meta-analytic studies should be done to confirm or to refute these patterns of relationships. A third conclusion is that the average validity of the FFM reported in previous meta-analyses should be considered in the light of the present findings, as there are important differences in the validity size for the three levels of job complexity. The overall average validity used unreservedly can hide the relevant effect of job complexity for at least four of the five big personality factors.

The findings of this meta-analysis have implications for the research and practice of personnel selection and personnel decisions. The first implication for research is for the previous meta-analyses of the criterion validity of the FFM. The proportions of studies corresponding to the high, medium, and lower levels of job complexity should be examined and compared as variations in these proportions can explain the differences in the validities obtained. This fact, together with the effects of IRR (Hunter, Schmidt, & Le, 2006) and the imperfect measurement of the constructs (Salgado, 2003) can produce considerable differences in the size of the validity. From an applied point of view, when making personnel decisions practitioners should take into account the degree of job complexity to evaluate in what way and to what extent personality factors predict job performance. This implication is especially relevant with regard to agreeableness, as the results of this meta-analysis provide clues for a better understanding of how this factor works in the labor domain. Previous meta-analyses of the validity of the FFM assessed with both SS and FC personality inventories have concluded that agreeableness was a predictor that did not generalize validity across occupations in order to predict job performance (e.g., Barrick et al., 2001; Hurtz & Donovan, 2000; Salgado, 1997, 2003; Salgado & Tauriz, 2014). However, the results of the validity of agreeableness for the three levels of complexity indicate that this factor can be a valid predictor of performance in those jobs characterized by a medium level of job complexity.

There are several limitations that should be noted. The first limitation is that the number of studies has not permitted us to examine the moderator effect of job complexity for normative FC inventories. Therefore, primary studies are needed for this format. Studies are also needed for the occupations of low complexity level in the case of ipsative FC personalities. A second limitation is that the number of estimates of range restriction did not allow for the development of a specific distribution for each combination of FC format—job complexity level-personality factor. A third limitation is that a relevant number of the studies included in this meta-analysis were conducted over four decades, and it is impossible to know if the validity remained stable despite potential changes in the nature of the occupations and the measures of job performance in the last two decades. All the limitations mentioned suggest that new validity studies should be conducted, in particular taking into account that quasi-ipsative FC inventories were shown to be more valid predictors than SS personality inventories (see Barrick et al., 2001; Judge et al., 2013; Salgado, Moscoso et al., 2015), in spite of the fact that the latter are used more frequently for personnel decisions than the former.

Considering the results obtained as a whole, the final conclusion to be inferred is that job complexity is a strong moderator of the validity of personality measures. Job complexity moderates the size of validity coefficients and makes some of them (e.g., openness, agreeableness) valid predictors, when it was believed that they were not, and others were shown to have larger validity than the one obtained in previous meta-analyses, provided the analysis is based on the medium and low complexity levels (conscientiousness and emotional stability).

**Conflict of Interest**

The author of this article declares no conflict of interest.

**Financial Support**

The research reported in this manuscript was supported by Grant PSI2014-56615-P from the Spanish Ministry of Economy and Competitiveness and Grant 2016 GPC GI-1458 from the Consellería de Cultura, Educación e Orden, Xunta de Galicia.

---

**Appendix.**

Validity Coefficients and Sample Size of the Studies Included in the Meta-analysis.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample</th>
<th>ES</th>
<th>EX</th>
<th>O</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High Complexity – Ipsative Forced Choice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>.02</td>
<td>.07</td>
<td>.02</td>
<td>.09</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>86</td>
<td>-.04</td>
<td>.05</td>
<td>.09</td>
<td>.18</td>
<td>.05</td>
</tr>
<tr>
<td>McDaniel, Yost, Ludwick, Hense, and Hartman (2004)</td>
<td>384</td>
<td>-.06</td>
<td>.16</td>
<td>.09</td>
<td>-.04</td>
<td>.03</td>
</tr>
<tr>
<td>Nelson (2008)</td>
<td>114</td>
<td>-.11</td>
<td>-.02</td>
<td>.07</td>
<td>-.21</td>
<td>.00</td>
</tr>
<tr>
<td>Nyfield, Gibbons, Baron, and Robertson (1995)</td>
<td>503</td>
<td>.04</td>
<td>-.02</td>
<td>-.02</td>
<td>-.05</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>-.01</td>
<td>.15</td>
<td>-.15</td>
<td>-.06</td>
<td>-.03</td>
</tr>
<tr>
<td>Perkins and Corr (2005)</td>
<td>68</td>
<td>.00</td>
<td>.15</td>
<td>.05</td>
<td>.14</td>
<td>.03</td>
</tr>
<tr>
<td>Robertson, Gibbons, Baron, Machyer, and Nyfield (1999)</td>
<td>114</td>
<td>-.03</td>
<td>-.02</td>
<td>-.05</td>
<td>.14</td>
<td>.03</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>.24</td>
<td>.01</td>
<td>.10</td>
<td>-.20</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>-.09</td>
<td>.12</td>
<td>-.09</td>
<td>.15</td>
<td>-.12</td>
</tr>
<tr>
<td></td>
<td>131</td>
<td>.02</td>
<td>-.16</td>
<td>.10</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>34</td>
<td>.03</td>
<td>-.15</td>
<td>-.02</td>
<td>-.30</td>
<td>.03</td>
</tr>
<tr>
<td>Salgado (1991)</td>
<td>189</td>
<td>.03</td>
<td>-.08</td>
<td>-.04</td>
<td>-.08</td>
<td>.09</td>
</tr>
<tr>
<td>Saville, Sik, Nyfield, Hackston, and Macher (1996)</td>
<td>440</td>
<td>-.03</td>
<td>.03</td>
<td>.07</td>
<td>.00</td>
<td>-.02</td>
</tr>
<tr>
<td>Schippmann and Pren (1989)</td>
<td>270</td>
<td>-.02</td>
<td>.12</td>
<td>.12</td>
<td>-.02</td>
<td>.08</td>
</tr>
<tr>
<td>SHL (2006)</td>
<td>120</td>
<td>.03</td>
<td>.06</td>
<td>.14</td>
<td>-.05</td>
<td>.01</td>
</tr>
<tr>
<td>Slocum and Hand (1971)</td>
<td>57</td>
<td>.17</td>
<td>.07</td>
<td>.05</td>
<td>.02</td>
<td>-.07</td>
</tr>
<tr>
<td>Whetzel, McDaniel, Yost, and Kim (2010)</td>
<td>1152</td>
<td>-.02</td>
<td>.07</td>
<td>.08</td>
<td>-.04</td>
<td>.08</td>
</tr>
</tbody>
</table>
### References


---

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


