Journal of Work and Organizational Psychology

http://journals.copmadrid.org/jwop
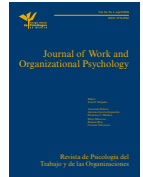
The Prediction of Training Proficiency in Firefighters:
A Study of Predictive Validity in Spain

Alfredo Berges, Elena Fernández-del-Río, and Pedro J. Ramos-Villagrasa

Universidad de Zaragoza, España

ABSTRACT

The present study provides results of criterion validity in the selection of firefighters in Spain. The predictors were cognitive skills, job knowledge, and physical aptitudes, and the criterion was training proficiency. The process involves 639 candidates, but only 44 complete successfully the selection process. Our results support previous evidence showing that general cognitive ability is the best predictor of training proficiency, with an operational validity of .57. With respect to the other predictors, job knowledge presented an operational validity of .55 and physical tests of .49. In addition, multiple regression analysis showed that cognitive aptitude explains 33% of the variance, but when physical aptitudes are included the explained variance increases to 50%. If we also add job knowledge, explained variance increases to 55%. Our study offers recent results of criterion validity in a barely investigated job, gathered in a country other than the one where prior research had been carried out.

La predicción del aprovechamiento de la formación en bomberos:
un estudio sobre su validez predictiva en España

RESUMEN

Este estudio ofrece resultados de la validez de criterio en la selección de bomberos en España. Los predictores fueron las aptitudes cognitivas, el conocimiento del puesto y las aptitudes físicas, siendo el criterio el aprovechamiento de la formación. El proceso comenzó con 639 candidatos, de los cuales solamente 44 superaron la selección. Nuestros resultados apoyan la evidencia previa, mostrando que la aptitud cognitiva general es el mejor predictor, con una validez operativa de .57, seguido del conocimiento del puesto con .55 y las pruebas físicas con .49. Además, el análisis de regresión múltiple mostró que la aptitud cognitiva explica un 33% de la varianza, que se incrementa hasta el 50% al incluir pruebas físicas y hasta el 55% si además se añade conocimiento del puesto. Estos resultados resultan especialmente interesantes al haber sido obtenidos en un país diferente al de las principales investigaciones de referencia (i.e., Estados Unidos de América).

Cognitive ability and personality are among the most outstanding predictors in personnel selection (Schmidt, Oh, & Shaffer, 2016). However, in the Spanish Public Administration, assessment of these variables is mandated by law for only two jobs: police and firefighters. In this sense, most regions have their own laws describing the functioning of these public services, including their selection process (e.g., Law 4/1992; July 8, art. 35, Coordination of the Local Police Force; Decree 222/1991; art. 31.4, Framework regulation of the Organization of the Local Police force of Aragón). Through

these selection processes, a large amount of information is gathered regarding individual differences that are used to predict applicants' job performance. Nevertheless, usually this valuable information is only used for the current process, which hinders the advance of science. Without the accumulation of updated empirical evidence, researchers and practitioners must trust classical works that may not reflect the current reality of the post. Studies in other countries have reported the same problem. For example, in his farewell article as editor of the *Journal of Applied Psychology*, Campbell regretted that

the vast majority of police and firefighter selection investigations were not published as journal articles, although technical reports were often produced that included criteria-oriented validity data (Campbell, 1982). Strangely, in the first issue of the same journal, an article about the evaluation of police and firefighters in a small American city was published at the request of the City Manager. In his own words, the author states that this effort is "an unusual experiment, perhaps the first of its kind to be made in this or any country" (Terman, 1917, p. 17).

Following the call of Campbell (1982), the present paper provides results of criteria-oriented validity corresponding to a selection process for firefighters. To the best of our knowledge, this is the first effort to share this kind of information for scientific purposes in a Spanish sample. Furthermore, this is also interesting for practitioners because the inclusion of psychometric tests in the selection of firefighters in the Autonomous Region of Aragon is a novelty introduced through the decree that regulates the organization and operation of fire extinguishing services (i.e., Decree 158/2014; Department of Territorial Policy and Interior, Government of Aragon). Previously, selection was carried out only through tests of knowledge and tests of physical abilities.

Cognitive Ability Test and Training Proficiency

Cognitive ability (whether general or specific) has demonstrated its relationship with organizational criteria such as task performance, overall performance, and training proficiency (e.g., Ones, Dilchert, & Viswesvaran, 2012; Salgado, 2016; Schmitt, 2014). As Scherbaum, Goldstein, Yusko, Ryan, and Hanges (2012) noted, cognitive ability is more important than ever in the workplace. The current work environment not only demands being able to deal with information (i.e., obtainment, analysis, decision making), using critical thinking, problem solving, and integrating technological tools, but also non-cognitive skills.

Focusing on training proficiency, several meta-analyses have analyzed the role of cognitive ability. Thus, Hunter (1983) reported a validity of .55 with data from 90 different samples and 6,496 workers in different positions. In Europe, Salgado, Anderson, Moscoso, Bertua, and de Fruyt (2003) also found a validity of .55, with data from 97 samples and 16,065 participants. Concerning specific countries, Salgado and Anderson (2002) analyzed the United Kingdom and found a validity of .47 in 25 samples (2,405 participants), whereas in Spain they reported a validity of .56 in 61 samples (20,305 participants). In another meta-analysis, Hülsheger, Maier, and Stumpp (2007) found a validity of .47 in 6 different samples from Germany with 1,089 participants.

As a whole, these meta-analyses demonstrate the validity of cognitive ability in very different jobs and positions. Another remarkable result is the moderating effect of job complexity, increasing the validity in high-complexity jobs. Thus, Salgado et al. (2003) establish a validity of .74 in posts for high-complexity, .53 for medium-complexity, and .36 for low-complexity jobs, respectively.

The prominent role of cognitive ability in the prediction of training proficiency pointed out by all this empirical evidence is essential for work settings (Ones et al., 2012). Cognitive ability facilitates learning job content and this, in turn, determines job performance (Hunter, 1986; Schmidt, Hunter, & Outerbridge, 1986). In other words, people with high cognitive ability can acquire more knowledge (and in less time) about how to apply their skills to deal with job duties (Hunter, 1986; McCloy, Campbell, & Cudeck, 1994). Thus, training proficiency is a proxy of the job knowledge acquired by workers. This is highly relevant in firefighters' job, where their actions tend to be carried out in emergency contexts. To deal with their job demands (including saving lives), firefighters must rely on their training and experience to apply procedures, techniques, and equipment in order to solve crises and make quick decisions.

Criterion Validity in Firefighter Selection Procedures

Criterion validity studies have been progressively replaced by content validity studies in firefighter research. This is due to litigation issues, because content validity is easily understandable for judges (Henderson, 2010). Although content validity studies have explanatory power about why a skill is relevant to performing a particular task on a job, they do not provide a quantitative estimation (i.e., statistical evidence), whereas studies of predictive validity do (Schmidt, 2012). The application of procedures that have shown scientific evidence is especially relevant in contexts where litigation is high, such as in public employment.

There are very few validity studies regarding the selection of firefighters. A notable exception to this trend is the work by the New York City Fire Department. As a consequence of a court judgment for discrimination in the selection process against Black and Hispanic minorities, the Fire Department conducted a validation process with three different approaches (PSI Services LLC, 2012): (a) content validity, (b) criterion validity, and (c) construct validity. Some other interesting efforts can be found in this case from the scientific literature. We refer to the works by Barrett, Polonsky, and McDaniel (1999), Henderson, Berry, and Matic (2007), and Henderson (2010). All this research is focused on the criterion validity of cognitive ability (Barrett et al., 1999; Henderson, 2010) and physical tests (Henderson, 2010; Henderson et al., 2007). We shall review each of these antecedents of our present research.

The first one was developed by Barrett et al. (1999), who carried out a meta-analysis to examine the validity of the predictors used in the selection of firefighters. They collected a total of 73 independent samples: 24 correlations of 2,791 participants for cognitive tests, 26 correlations of 3,087 participants for mechanical aptitude, and 23 correlations of 3,637 participants for a composite of cognitive-mechanical aptitude. They also collected samples that included results of training proficiency: 14 correlations of 2,007 participants for cognitive tests, 5 correlations of 869 participants for mechanical aptitude, and 9 correlations of 1,027 participants of a composite of both. Practically all the primary studies were really the technical results of selective processes not published in specialized journals, elaborated by the fire services, companies that distribute psycho-technical tests, and some doctoral dissertations. The authors made a considerable effort comprising the last two decades of research. The predictors they analyzed were cognitive ability tests, spatial/mechanical tests, and the composite criterion of cognitive-mechanical aptitude. Thus, they used job performance and training proficiency (i.e., supervisor-rating) as criteria. The results of the meta-analysis were as follows: for job performance, they found a mean correlation of .20 with cognitive ability, .26 with mechanical aptitude, and .28 with the composite criterion. After controlling for criterion reliability and range restriction, these correlations yielded operational validities of .42, .54, and .56, respectively. In the case of cognitive tests, the value of credibility at 90% was -.03, which means that this result cannot be generalized. On the other hand, the values of credibility at 90% for the operational validity of mechanical aptitude and the composite were .17 and .40 respectively, supporting that these results are generalizable. For training proficiency, the observed correlations were .50, .37, and .50, respectively, which yielded operational validities of .77, .62, and .77, respectively. The creditability values were .62, .73, and .40, supporting the generalization of results. In view of these results, Barrett et al. (1999) concluded that the validity of cognitive ability to predict job performance can be established by the number of coefficients and the size of the samples included in the study. In the case of the validity of training proficiency, they indicated that the results were more tentative because they were based on a significantly lower number of coefficients that were also based on smaller samples.

The second article regarding selection of firefighters was the one by Henderson et al. (2007). They analyzed 306 firefighters participating in a 16-week training course in the academy, measuring eight different

predictors, including anthropometric measures (e.g., estimated body fat percent) and physical ability tests (e.g., bench-press or step test). Using factor analysis performed at week 1 and week 14, they found that physical abilities were grouped into two factors: strength and endurance. The criteria were the outcomes in three situational tests with material used in actual interventions (e.g., hoses and stairs) and an instructor rating in terms of high, medium, and low performance. The correlations reached .54 on average, with a range of .45 to .59. Through regression analysis, they found that correlations increased as training progressed. They confirmed these results with structural equation modeling, showing that strength and endurance have direct effects on applicants' performance in situational tests.

The last work is a primary study carried out by Henderson (2010), in which he examined criterion validity of cognitive and physical ability measures. Seventy-four participants in a training course after a successful selection procedure were studied for 23 years. As this constitutes a considerable effort with a lot of valuable information, we are going to summarize his results in more detail.

Participants completed two different tests, the first one of cognitive abilities and job-knowledge, and the second one of physical abilities. The cognitive ability component of the test consisted of 120 questions grouped into six sections: (1) reading comprehension based on technical documents; (2) knowledge of fire prevention and extinction techniques and rescue based on the study material; (3) following instructions in logical reasoning tasks; (4) calculation and application of formulas related to firefighting; (5) drawing conclusions from written statements; and (6) identifying series of numbers, letters, and symbols. That is, the applicants had to demonstrate some prior knowledge related to fire prevention and extinction activity and also their ability in cognitive ability tests. Following Carroll's (1993) taxonomy, the six test sections reflect the following first-order cognitive factors: (1) Associative Memory, Meaningful Memory, and Visual Memory; (2) Reading Comprehension, Visualization and Mechanical Knowledge; (3) Integrative Processes and Sequential Reasoning; (4) Numerical Ability and Quantitative Reasoning; (5) Sequential Reasoning and Reading Comprehension, and (6) Induction. The author argues that because the six sections load between .72 and .81 on a factor in a principal components factorial analysis, it can be concluded that the test loads very high on g factor and, therefore, is a measure of general cognitive ability.

The second component of the test comprised physical abilities that covered strength and endurance. Additionally, the applicants completed two situational tests that were evaluated by the time invested in their performance. The first event simulated a fire scene arrival where the applicants had to drag a hose and handle a ladder. The second event simulated rescue evolution, where the candidates had to drag a sack through a low-headroom and narrow space.

Regarding criteria, Henderson (2010) proposed two different approaches: training proficiency and job performance. Training proficiency has several different measures that were significantly predicted by cognitive ability: (a) average examination grade, with an observed correlation of .67 and a corrected validity of .86; (b) emergency medical technician (EMT) state examination, with an observed correlation of .74 and a corrected validity of .92; (c) critical skill deficiencies in EMT and/or self-contained breathing apparatus (SCBA), with an observed correlation of .68 and a corrected validity of .90; (d) mean instructor rating of physical and practical skills, with an observed correlation of .22 and a corrected validity of .40; and (e) mean instructor rating of handling ground ladders, with an observed correlation of .19 and a corrected validity of .22. It is remarkable that cognitive tests predicted a criterion theoretically more related to physical ability. These data can be interpreted as the result of a halo effect of the qualifications given by instructors. Another criterion that is strictly physical (i.e., academy physical measures) was also predicted by cognitive ability (observed correlation of .23, corrected validity of .25) and, of course, by physical ability (observed correlation of .72 and a corrected validity of .90).

Regarding job performance, between 1985 and 2006, cognitive tests and physical tests were the predictors of performance rated by the officers. In 1992, the obtained observed correlations between the qualification of firefighting knowledge and judgment and physical and cognitive tests were .13 and .52 respectively. Observed correlations between the measure of physical strength and endurance, as rated by the supervisors, were .61 with physical tests, and .30 with cognitive test. All observed correlations were higher when corrected for criterion unreliability and indirect range restriction. In Henderson's (2010) opinion, these results showed a halo effect of the performance ratings that made it possible to obtain higher than expected correlations in very specific criteria (i.e., strength and endurance with cognitive tests, and judgment and knowledge with physical tests). In 2006, the "Elite Firefighter Squad" nominations (an acknowledgment system used by the service) had an observed correlation of .46 with cognitive tests (.73 corrected validity) and of .30 with physical tests (.50 corrected validity). The estimated total number of performances by each firefighter was predicted by cognitive abilities with a correlation of .25 and by physical abilities with a correlation of .36 (corrected validity of .43 and .53, respectively). Finally, a composite that collected the main criteria measures between 1992 and 2006 obtained a correlation of .47 with cognitive abilities and of .38 with physical abilities. Once corrected, these values reached .70 and .57, respectively.

Comparing these results with those of Barrett et al. (1999), the high correlations reported by Henderson (2010) are striking. The observed correlations are higher than the majority of the operational validations found in the general meta-analyses performed with heterogeneous samples and analyzing the moderating effect of job complexity. The explanation, according to Henderson's justification, is that the reliability of the criteria is very high and the range restriction is small.

As a summary of our review of the results on the criterion validity in firefighters and of the meta-analytical studies, we can conclude that: (a) training proficiency in firefighters is an antecedent of job performance; (b) general mental ability and specific cognitive abilities are good predictors of training proficiency; (c) general physical condition (strength and endurance) and some psychomotor skills are predictors of training proficiency when the training content is related to the same kind of physical activity; and (d) knowledge of the firefighter's job is a determinant of the criterion.

The Present Study

Although our review has shown some consensus regarding the relevance of firefighters' training proficiency and the role of cognitive and physical abilities as its predictors, more research is needed. In this sense, García-Izquierdo, Aguinis, and Ramos-Villagrasa (2010) called for more research in countries where legislation related to job issues is under development, as in Spain. As mentioned before, recent legislation makes it mandatory to consider cognitive ability tests in the selection of firefighters in Spain. Studies like this one may help to show the importance of this predictor, like in other countries.

With this idea in mind, the present study has two complementary objectives: (1) to provide criterion validity of general mental ability, specific cognitive abilities, physical abilities, and knowledge in the prediction of training proficiency in a sample of Spanish firefighters; (2) to analyze the joint role of all of these predictors, with training proficiency as criterion.

Method

Participants

Participants were aspiring firefighters who successively underwent the selective process. Of the 639 applicants who performed the first exercise, only 44 (6.89%) finally passed all the

tests successfully and began the period of training in the academy. Thus, there were 44 participants for the main calculations. To calculate the standard deviations of the unrestricted group and the reliability of the knowledge and physical tests, we used the total number of participants who completed each test, i.e., 639 (100%) of the applicants for the knowledge tests and 178 (27.85%) for the physical tests.

All 44 participants (43 men and 1 woman) had an average age of 31 years with a range of 22 to 44 years. All of them had at least a high school's degree or were vocational training technicians.

Measures

Knowledge. This was measured with an objective test of 110 questions with three response alternatives. The maximum possible score was 15 points. Reliability, calculated as internal consistency (Spearman-Brown), was .80.

Physical ability. The selective process consisted of six different tests: (1) rope climbing; (2) combined circuit of agility, speed, and strength; (3) long-distance racing (1500 meters); (4) swimming (100 meters); (5) control of vertigo and balance; and (6) claustrophobic control. The first four tests were rated between 0 and 10 and the last two were rated as pass/fail. The reliability (internal consistency with the Spearman-Brown formula) was .42. This reliability is intermediate but, considering the heterogeneity of the tests, this result is expected.

Cognitive abilities. The second part of the selection process consisted of two tests: a spatial reasoning test and an abstract reasoning test. Both tests belong to a battery of skills marketed by Tea Ediciones[1]. In addition, we obtained a composite criterion of the two tests constructed from the standardized scores. The reliabilities obtained from the normative group with which the applicants were rated were .92 for the spatial and the compound test and .91 for the abstract reasoning test. For the choice of specific skills, we used Barrett et al.'s (1999) meta-analysis and a more complete job analysis of firemen's occupation than those we have reviewed (Bureau of Fire Standards & Training, Division of State Fire Marshal, 2015). According to Schmidt (2002, 2012), the combination of two or more cognitive abilities constitutes a measure of general mental ability, which is the construct we consider is the spatial plus reasoning (S+R) compound used in the study.

Training proficiency. This was measured as the candidates' mean scores in the examinations proposed by the instructors of the different subjects studied in the 6-month course taken in the academy. The course combined academic activities in the classroom with diverse practices with equipment and simulation of real situations. The instructors jointly rated the applicants' performance in the exams and practices. Reliability (Spearman-Brown internal consistency) was .77.

Procedure

Data corresponding to the predictors were taken from the applicants' scores in the successive tests. Selection followed a non-compensatory model of multiple obstacles. The first test was a knowledge test; applicants who did not exceed the cut-off point did not pass on to the next test. The second test was a physical test composed of six mutually exclusive subtests, such that the candidates had to pass all of them to take the third test, which was the cognitive ability test. In the academy course, the candidates received theoretical and practical training oriented to their professional preparation as firemen. The different examinations and evaluations prepared and carried out by the applicants were the measure of the research criterion.

Analysis

The statistics and correlations were calculated with SPSS. The VALCOR program (Salgado, 1997) was used to correct the observed correlations. Structural equation models were performed with LISREL (Jöreskov & Sörbom, 1993).

Results

Table 1 presents the means and standard deviations of the restricted and unrestricted groups, and the inverse of the coefficient of heterogeneity (u). The correlations between the study variables and their reliabilities can also be observed. All the variables had a range restriction, with the highest one corresponding to the physical ability tests. All reliabilities had acceptable values with the exception of that obtained for physical tests, which was low (.42). As expected, the highest correlations were between cognitive abilities. All correlations with training proficiency were significant with the exception of physical tests (.29). The highest correlation with training proficiency corresponded to general mental ability (.45), followed by the knowledge test (.44). The high correlation between the knowledge test and the physical test (.51) is striking.

The observed correlation is an imperfect indicator of validity, as it is affected by several measurement errors. In this case, the sampling error cannot be controlled because, in practice, it is impossible to increase the number of participants. However, there are two other types of errors that can be neutralized with appropriate corrections (Van Iddekinge & Ployhart, 2008): measurement error in the criterion and range restriction. Table 2 shows the corrected validity of the three measures of knowledge, physical abilities, and cognitive abilities. First, we corrected the observed correlations due to measurement error or unreliability in the criterion. The obtained coefficient was then corrected for direct range restriction, and the confidence intervals were calculated with a 90% probability.

Our data support that general mental ability is the best predictor of training proficiency, with an operational validity of .57. The knowledge test reaches an operational validity of .55, corresponding to the second predictor of the criterion. Lastly, the physical ability tests have an operational validity of .49. The very important increase in the correlation observed in the corrected correlation of physical tests is due to the strong range restriction of this variable in the selection process. The amplitude in confidence intervals supports the validity of the measures used.

Table 1. Descriptive Statistics and Correlations

Variable	*M*	*SD*	*sd*	u	1	2	3	4	5	6
1. Spatial	24.41	4.70	3.84	.82	(.92)					
2. Reasoning	27.02	4.30	3.44	.80	.57**	(.91)				
3. GMA	127.10	33.50	28.50	.85	.89**	.88**	(.92)			
4. Knowledge	10.24	1.42	1.24	.88	.07	.26	.18	(.80)		
5. Physical	7.38	1.20	0.73	.60	-.07	.20	.07	.51**	(.42)	
6. Training	8.17	---	0.46	---	.39**	.40**	.45**	.44**	.29	(.77)

Note. Reliability (α) is on the diagonal in brackets. GMA = General Mental Ability; u: inverse coefficient of heterogeneity. *M* = mean; *SD* = standard deviation of unrestricted group; *sd* = standard deviation of the restricted group.

Table 2. Validity Corrected by Measurement Error and Range Restriction

Variable	(a)	(b)	(c)
1. Spatial	.44	.51	.30–.76
2. Reasoning	.46	.54	.34–.75
3. GMA	.51	.57	.39–.76
4. Knowledge	.50	.55	.36–.74
5. Physical	.33	.49	.21–.77

Note. (a) Validity corrected for measurement error of the criterion; (b) Validity corrected for range restriction in the predictor; (c) Confidence intervals at 90% probability; GMA = General Mental Ability.

In order to examine changes in validity according to which variables are included in the predictive model, we calculated a multiple regression using training proficiency as criterion and general cognitive ability, physical ability, and knowledge as predictors. As can be seen in Table 3, when we only used general cognitive ability as predictor, 33% of the variance was explained. Upon including physical ability tests, the explained variance rose to 50%, whereas when including knowledge of the job, the explained variance reached 55%.

Table 3. Correlations and Increase in Validity

Predictors increase	r/R	r/R^2	Dif.	% Validity
GMA	.57	.33	--	--
GMA and physical	.71	.50	.14	24.60%
GMA, physical, and knowledge	.74	.55	.03	4.23%

Note. GMA = General Mental Ability.

Discussion

The present paper aimed to examine the validity of the predictors used in the selection of firefighters, using training proficiency as criterion. The results obtained in this study confirm, once again, the criterion validity of general cognitive ability. The operational validity we obtained does not differ from that found by the most important meta-analyses (e.g., Hülsheger et al., 2007; Hunter, 1983; Salgado & Anderson, 2002; Salgado et al., 2003).

Regarding the results of Barrett et al. (1999), the validity we obtained is markedly lower (.57 vs. .77). The differences may be due to the fact that we used two specific aptitudes as a compound (abstract and spatial reasoning), whereas Barrett et al. (1999) used results from studies with measures of general cognitive ability and measures of mechanical and spatial skills. The validities observed in the two cases differ less than the corrected ones (.50, .37, and .50, respectively, vs. .40, .39, and .45 in the present study), so that a possible explanation of the differences may be the reliability of the criterion and the range restriction used to make the corrections.

Concerning Henderson (2010), our observed and operational validity are much lower than those obtained by him. In the case of corrected validations, Henderson assumed the existence of an indirect range restriction but without justifying which variables would be responsible for this restriction or providing the data used for calculation. As for the observed validations, as mentioned, it is surprising that in some cases they are higher than the operational validations found in the published meta-analyses.

Focusing on our study, we found a very high correlation between the knowledge test and the physical ability tests. This information is difficult to interpret. With the data available, we cannot ascertain if the relationship is spurious or due to the influence of a third variable. Exploring this issue may be interesting if further studies also find this result.

The results of multiple regression analyses confirm that general cognitive ability is the best predictor of training proficiency.

Nevertheless, the combination of cognitive and physical ability tests provides a better model for the prediction of firefighter selection, as they are independent of each other, along the lines of Henderson (2010). Regarding the knowledge test, we know that the acquisition of declarative and procedural knowledge depends on the intellectual level, and that the main effect of general cognitive ability on performance occurs through knowledge. In the practice of the selection of firefighters, the measure of prior job-related knowledge is a predictor of the use of training that contributes to increasing the validity by more than 4%.

Taking all of this into account, the results obtained represent an interesting contribution for several reasons. Firstly, because of their novelty. As mentioned above, in spite of the number of selection processes of firemen that take place in our country, the publication of results of criterion validity is rare. Secondly, when data involve predictive and non-concurrent validity, the possibility of generalization of results is greater. Thirdly, from a theoretical point of view, it is necessary to continue to carry out primary studies on the validity of criteria for cognitive abilities (Schmitt, 2014) because the most important meta-analyses carried out to date include 25-year-old studies or older. Fourthly, because its utility: firefighters' core tasks are carried out in emergency situations that demand maximum performance. A selection that guarantees the aptitude of the selected individuals will allow firemen to carry out their tasks appropriately and will ensure that the average performance is very high and with little variability (i.e., practically all firefighters work very well). Estimates by Schmidt, Hunter, Outerbridge, and Trattner (1986) pointed out that the use of more valid assessment methods provided a 9.74% increase in activity volume and a 61.4% decrease of poor performers.

The present paper has some limitations that should be commented on. First, the number of participants that were used ($N = 44$) limits the generalization of results. The socioeconomic conditions of our public administrations in recent years, in particular, the limitations of the replacement rate, lead to very small selection cohorts with very low selection ratios. Nevertheless, sample sizes like this one are usual in validation studies, in which the sample is reduced directly by the selection process itself, but valuable because the amount of samples to perform meta-analytic studies is increased. In any event, we note that, in this study, the selection began with 639 applicants to fill 44 vacancies, which provides a selection ratio of .07.

Another important limitation of the study is that it provides only data of the validity of training proficiency and not of task performance. However, we stressed the relevance of training for the future performance of firefighters. Last but not least, it would be very interesting to examine the validity of other non-cognitive predictive variables that would better explain the mastery of firefighters' performance.

Conflict of Interest

The authors of this article declare no conflict of interest.

Notes

[1]The specific information regarding the battery is omitted because it is commercial material and to preserve its possible use in further selection processes.

References

Barrett, G. V., Polomsky, M. D., & McDaniel, M. A. (1999). Selection tests for firefighters: A comprehensive review and meta-analysis. *Journal of Business and Psychology, 13*, 507-514. https://doi.org/10.1023/A:1022966820186

Bureau of Fire Standards & Training, Division of State Fire Marshal (2015). *Job Analysis Study Technical Report for the position of Career and Volunteer Firefighter in the State of Florida.* Industrial/Organizational Solutions. Retrieved from https://www.myfloridacfo.com/division/sfm/BFST/Documents/FLJobAnalysisStudyTechnicalReport-2015.pdf

Campbell, J. P. (1982). Some remarks from an outgoing editor. *Journal of Applied Psychology, 67,* 691-700. https://doi.org/10.1037/h0077946

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Decree 222/1991. *Framework regulation of the Organization of the Local Police force of Aragón.*

Decree 158/2014. Department of Territorial Policy and Interior. Government of Aragon.

García-Izquierdo, A. L., Aguinis, H., & Ramos-Villagrasa, P. J. (2010). Science-practice gap in e-recruitment. *International Journal of Selection and Assessment, 18,* 432-438. https://doi.org/10.1111/j.1468-2389.2010.00525.x

Henderson, N. D. (2010). Predicting long-term firefighter performance from cognitive and physical ability measures. *Personnel Psychology, 63,* 999-1039. https://doi.org/10.1111/j.1744-6570.2010.01196.x

Henderson, N. D., Berry, M. W., & Matic, T. (2007). Field measures of strength and fitness predict firefighter performance on physically demanding tasks. *Personnel Psychology, 60,* 431-473. https://doi.org/10.1111/j.1744-6570.2007.00079.x

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment, 15,* 3-18. https://doi.org/10.1111/j.1468-2389.2007.00363.x

Hunter, J. E. (1983). *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the general aptitude test battery (GATB).* US Department of Labor. US Employment Service. Washington, DC.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362. https://doi.org/10.1016/0001-8791(86)90013-8

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Law 4/1992, July 8. *Coordination of the Local police force.*

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79,* 493-505. https://doi.org/10.1037/0021-9010.79.4.493

Ones, D. S., Dilchert, S., &Viswesvaran, C. (2012). Cognitive abilities. N. Schmitt (Ed.). *Handbook of personnel assessment and selection* (pp. 179-224). Oxford, UK: Oxford University Press.

PSI Services LLC. (2012). *Firefighter test development and validation report.* Retrieved from http://appliedpersonnelresearch.com/papers/nycff/Test%20Development%20and%20Validation%20Report%20-%20Doc.%20976-1_OCR.pdf

Salgado, J. F. (1997). VALCOR: A program for estimating standard error, confidence intervals, and probability of corrected validity. *Behavior Research Methods, Instruments, & Computers, 29,* 464-467. https://doi.org/10.3758/BF03200601

Salgado, J. F. (2016). Using ability tests in selection. In H. Goldstein, E. Pulakos, J. Passmore, & C. Semedo (Eds.) *Handbook of the Psychology of Recruitment, Selection, and Retention* (pp. 115-149). United Kingdom: Wiley/Blackwell.

Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and Evidence. *Human Performance*, 15, 75-96. https://doi.org/10.1080/08959285.2002.9668084

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56,* 573-605. https://doi.org/10.1111/j.1744-6570.2003.tb00751.x

Scherbaum, C. A., Goldstein, H.W., Yusko, K. P., Ryan, R., & Hanges, P. J. (2012). Intelligence 2.0: Reestablishing a research program on g in I-O psychology. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 5,* 128-148. https://doi.org/10.1111/j.1754-9434.2012.01419.x

Schmidt, F. L. (2002). The role of general mental ability and job performance: Why there cannot be a debate. *Human Performance, 15,* 187-210. https://doi.org/10.1207/S15327043HUP1501&02_12

Schmidt, F. L. (2012). Cognitive tests used in selection can have content validity as well as criterion validity: A broader research review and implications for practice. *International Journal of Selection and Assessment, 20,* 1-13. https://doi.org/10.1111/j.1468-2389.2012.00573.x

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432-439. https://doi.org/10.1037/0021-9010.71.3.432

Schmidt, F. L., Hunter, J. E., Outerbridge, A. M., & Trattner, M. H. (1986). The economic impact of job selection methods on the size, productivity, and payroll costs of the federal work-force: An empirical demonstration. *Personnel Psychology, 39,* 1-29. https://doi.org/10.1111/j.1744-6570.1986.tb00572.x

Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings. *Fox School of Business Research Paper.* Retrieved from https://ssrn.com/abstract=2853669

Schmitt, N. (2014). Personality and cognitive ability as predictors of effective performance at work. *Annual Review of Organizational Psychology and Organizational Behavior, 1,* 45-65. https://doi.org/10.1146/annurev-orgpsych-031413-091255

Terman, L. M. (1917). A trial of mental and pedagogical tests in a civil service examination for policemen and fireman. *Journal of Applied Psychology, 1,* 17-29. https://doi.org/10.1037/h0073841

Van Iddekinge, C. H., & Ployhart, R. E. (2008). Developments in the criterion-related validation of selection procedures: a critical review and recommendations for practice. *Personnel Psychology, 61,* 871-925. https://doi.org/10.1111/j.1744-6570.2008.00133.x