Psicothema

Psicothema 2016, Vol. 28, No. 4, 471-478
doi: 10.7334/psicothema2016.68

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

A computerized adaptive test for enterprising personality assessment in youth

Ignacio Pedrosa, Javier Suárez-Álvarez, Eduardo García-Cueto and José Muñiz
Universidad de Oviedo

Abstract

Backgound: Assessing specific personality traits has shown better predictive power of enterprising personality than have broad personality traits. Hitherto, there have been no instruments that evaluate the combination of specific personality traits of enterprising personality in an adaptive format. So, the aim was to develop a Computerized Adaptive Test (CAT) to assess enterprising personality in young people. **Methods:** A pool of 161 items was developed and applied to two sets of participants (n_1 = 357 students, M_{age} = 17.89; SD_{age} = 3.26; n_2 = 2,693 students; M_{age} = 16.52, SD_{age} = 1.38) using a stratified sampling method. **Results:** 107 items that assess achievement motivation, risk-taking, innovativeness, autonomy, self-efficacy, stress tolerance, internal locus of control, and optimism were selected. The assumption of unidimensionality was tested. The CAT demonstrated high precision for a wide range of θ, using a mean of 10 items and demonstrating a relatively low Standard Error (0.378). **Conclusions:** A brief, valid, and precise instrument was obtained with relevant implications for educational and entrepreneurial contexts.

Keywords: Entrepreneurship, Assessment, Computerized Adaptive Test, Youth.

Resumen

Un Test Adaptativo Informatizado para la evaluación de la personalidad emprendedora en jóvenes. **Antecedentes:** la evaluación de rasgos específicos de personalidad, respecto a rasgos generales, ha demostrado un mayor poder predictivo de la personalidad emprendedora. Actualmente, no existe ningún instrumento que evalúe el conjunto de rasgos específicos de la personalidad emprendedora que se consideran relevantes en un formato adaptativo. El objetivo fue desarrollar un Test Adaptativo Informatizado (TAI) que permita evaluar la personalidad emprendedora en jóvenes. **Método:** se desarrolló un banco inicial de 161 ítems, el cual se aplicó a dos conjuntos de participantes (n_1 = 357 estudiantes, M_{edad} = 17,89; DT_{edad} = 3,26; n_2 = 2.693 estudiantes; M_{edad} = 16,52, DT_{edad} = 1,38) mediante un muestreo estratificado. **Resultados:** se seleccionaron 107 ítems que evalúan motivación de logro, toma de riesgos, innovación, autonomía, autoeficacia, tolerancia al estrés, locus de control interno y optimismo, sobre los que se comprobó el supuesto de unidimensionalidad. El TAI desarrollado muestra una elevada precisión para un amplio rango de θ, empleando una media de 10 ítems y presentando un error típico de las estimaciones relativamente bajo (0,378). **Conclusiones:** se cuenta con un instrumento breve, válido y preciso con múltiples implicaciones en el contexto educativo y emprendedor.

Palabras clave: personalidad emprendedora, evaluación, test adaptativo informatizado, jóvenes.

Entrepreneurs play a crucial role in modern economy owing to their ability to offset negative effects of socioeconomic development, to take advantage of new opportunities, and to generate new ideas for business growth (OECD/The European Commission, 2013). Consequently, research in the area of entrepreneurship has increased significantly in recent years (Liñán & Fayolle, 2015).

While economic and social factors play an important role when starting a business (Obschonka et al., 2015), various explanatory models suggest that individual variables, especially personality variables, are particularly relevant (Baum, Frese, Baron, & Katz, 2007; Brandstatter, 2011). The study of the enterprising personality may be tackled from the perspective of broad personality traits such as the Big-Five (Brandstätter, 2011; Zhao, Seibert, & Lumpkin, 2010) or centered on more specific traits (Rauch & Frese, 2007a, 2007b; Suárez-Álvarez, Pedrosa, García-Cueto, & Muñiz, 2014). In this second group, previous studies have identified the following as the main specific personality traits: achievement motivation, risk-taking, innovativeness, autonomy, self-efficacy, stress tolerance, internal and external locus of control, and optimism (Baum et al., 2007; Miller, 2015; Muñiz, Suárez-Álvarez, Pedrosa, Fonseca-Pedrero, & García-Cueto, 2014; Rauch & Frese, 2007a, 2007b; Zhao et al., 2010).

The attempt to predict entrepreneurial success from the simultaneous evaluation of both approaches has confirmed that the specific traits have more predictive power. In fact, when combining both measures, the effect of the Big-Five is considerably reduced, with extraversion being the only trait to have a statistically significant influence on the prediction (Leutner, Ahmetoglu,

Received: February 22, 2016 • Accepted: June 10, 2016
Corresponding author: Ignacio Pedrosa
Facultad Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
e-mail: npedrosa@cop.es

Akhtar, & Chamorro-Premuzic, 2014). An evaluation based on specific personality traits, therefore, will allow more accurate predictions of the success of entrepreneurs than predictions obtained from broad personality traits (Rauch & Frese, 2007a, b).

There are many tools for the evaluation of the enterprising personality (Suárez-Álvarez & Pedrosa, 2016), such as the Battery for the Assessment of the Enterprising Personality (BEPE; Muñiz et al., 2014), which offers a combined evaluation of the aforementioned specific personality traits and has been validated in young population; or the Measure of Entrepreneurial Tendencies and Abilities (META; Almeida, Ahmetoglu, & Chamorro-Premuzic, 2014), which is "*the strongest and most consistent predictor of entrepreneurial activity*" (Almeida et al., 2014, p. 1). However, despite the diversity of evaluation instruments in this context, there are none which evaluate the enterprising personality adaptively, via a Computerized Adaptive Test (CAT).

CATs have numerous advantages over conventional tests (Muñiz, 1997; Van der Linden & Glas, 2010) due to their being based on IRT and to their intrinsic characteristics. The development of a CAT will allow the quick and precise evaluation of specific personality traits both in terms of assessment and of making decisions based on test scores. This is particularly important in contexts where there are numerous evaluations, such as education. The application of a CAT would make it easier to perform quick assessments of all of the students in a school, and speed up the decision-making process regarding academic and professional orientation for whom interested on run their own idea. An additional advantage comes from item invariance, which allows identifying which groups can be more precisely evaluated. Finally, it would facilitate making better decisions based on the scores of each participant, increasing their validity (Lane, 2014) and possibly leading to fewer business failures and associated negative consequences.

Consequently, the development of a CAT would be a significant contribution to the field. Also, its validation in young population would make it easier to make decisions that would promote the growth of entrepreneurship, reduce the risk of failure, and guide potential entrepreneurs, improving their training and opening up possibilities for their future professional success. The aim of this paper is, therefore, to develop a CAT that allows the assessment of young potential entrepreneurs though specific personality traits.

Method

Participants

The items developed in a first study were applied to a sample of 357 participants (M = 17.89, SD = 3.26; 54% male) from the Principality of Asturias (Spain).

The second study used a stratified sampling method based on geographical area (21.1% rural, 9.1% coastal, and 69.8% urban), the ownership of the school (60.8% public, 35.7% state-subsidized private, and 3.5% private), and the educational level (34.2% compulsory/10th grade, 57.6% continuing education/12th grade, and 8.2% vocational training). Using this method, 2,693 students were selected from various regions in the North of Spain (92.8% Asturias, 3.2% Cantabria, and 4% Leon). The age range was between 16 and 23 years (55% aged 14 to 16.38% aged 17 to 18, and 7% aged 19 to 23), with a mean age of 16.52 (SD = 1.38), of whom 51% were male.

These participants were previously assessed by Suárez-Álvarez et al. (2014), and Muñiz et al. (2014) to achieve different objectives.

Instruments

An initial pool of 161 Likert-type items with five response categories was developed to evaluate the nine previously identified personality traits or facets (Pedrosa, Súarez-Álvarez, & García-Cueto, 2013). Items were distributed as follows: achievement motivation (20 items), risk-taking (19 items), innovativeness (19 items), autonomy (21 items), self-efficacy (21 items), stress tolerance (19 items), internal locus of control (12 items), external locus of control (17 items), and optimism (13 items). The recommendations for test construction provided in the current psychometric literature were followed (AERA, APA, NCME, 2014; Moreno, Martínez, & Muñiz, 2015). The items were developed to be easily understood by young people, using vocabulary in accordance with their comprehension skill and content suitable for their age group.

Procedure

The items were applied by psychologists in classrooms provided by the schools, in a single, group session. Participation in the study was voluntary, and participants did not receive any kind of compensation or reward.

Data analyses

In the first study, qualitative and quantitative studies were performed to analyze both the relevance of the specific personality traits of the model and the representativeness of the content (Sireci & Faulkner-Bond, 2014) of the developed items. Ten experts were asked to indicate both the relevance of each of the proposed traits and to classify the items according to the trait they thought it belonged to. Based on their answers, Aiken's V index was calculated. Item discrimination indices for each subscale were estimated, and items with indices lower than .20 were eliminated (Muñiz, Fidalgo, García-Cueto, Martínez, & Moreno, 2005). An exploratory factor analysis of each subscale was performed, using the polychoric correlation matrix and the method of generalized least squares. Items with a factorial loading of less than .20 were removed (Muñiz et al., 2005). The dimensionality of each trait was determined through optimal implementation of parallel analysis (Timmerman & Lorenzo-Seva, 2011) with 100 random correlation matrices. Psychometric properties of the complete battery were estimated by applying a second-order exploratory factor analysis on the matrix of correlations between the traits in the model. The robust maximum likelihood estimation method was used both to demonstrate a better fit of the data to the model and for the violation of the assumption of multivariate normality.

In the second study, the psychometric properties of the eight selected facets were analyzed through the analysis of the items for each facet separately. The discrimination indices were calculated. Differential item functioning in terms of gender was examined using the logistic regression method (Gómez-Benito, Hidalgo, & Zumbo, 2013). Exploratory factor analysis was performed using the polychoric correlation matrix and the method of weighted least squares. Items with discrimination indices lower than .20 and/or factor loadings less than .25 (Muñiz et al., 2005) were eliminated. To ascertain the dimensionality of each subscale, the optimal implementation of parallel analysis was performed, using 10,000 random correlation matrices. To test the fit to a unidimensional structure, the percentage of variance explained by each factor was considered, as well as the goodness of fit index (GFI) and the root mean square residual (RMSR). The reliability of the subscales was estimated using Cronbach's alpha coefficient for ordinal data (Elosua & Zumbo, 2008).

Within the framework of IRT, the unidimensionality assumption was checked first. Given the clear correlation shown between the different personality traits (Muñiz et al., 2014; Suárez-Álvarez et al., 2014), a unidimensional hypothesis for the battery was established. This supposition was tested both by exploratory and confirmatory approaches. The participants were randomly divided into three data subsets. An exploratory factor analysis was applied to the first subsample (n_1 = 852). A confirmatory factor analysis was applied to the second subsample (n_2 = 955), correlating the errors of measurement, with the assumption of unidimensionality. The solution from that confirmatory factor analysis was replicated on a third subsample (n_3 = 886) to check the consistency of the fit indices previously obtained (Byrne, 2001). Robust maximum likelihood estimation method was used in all factor analyses, to show a better fit of the data to the model and for the violation of the assumption of multivariate normality.

The calibration of items was performed under the Samejima graded response model (Samejima, 1969) and their fit was determined via a graphical analysis of standardized residuals using ResidPlots-2 software (Liang, Han, & Hambleton, 2008).

The precision of the pool of items was studied using three approaches: (a) estimating the information function from the participants' responses; (b) through the correlation between raw scores from the original instrument on a paper and pencil format and θ estimated from a pool of items using IRT; and (c) estimating the precision for a larger and more heterogeneous sample by simulation; a total of 13 samples were simulated with a range of θ = ±3 with intervals of θ in steps of 0.5. Each sample was made up of 10,000 participants, resulting in 130,000 participants. For each of these samples, θ was estimated and the corresponding standard error (SE) according to the information function test was calculated.

Finally, the performance of the CAT was tested following these algorithms: (a) a first item is randomly selected from the item pool based on its *a-parameter;* (b) A response is given to the previously selected item in order to estimate the participant's provisional θ. For this purpose, using maximum likelihood procedures, an item from the total data matrix with a maximum information function at that estimated ability level (θ*) is presented; (c) Using maximum likelihood procedures, a new estimation of θ and SE for that participant is made; (d) Steps 2 and 3 are repeated until two criteria are met; and (e) Once the CAT is finished, both the participant's ability level and the corresponding SE are estimated again using maximum likelihood procedures.

Testing the performance of the CAT, a simulation method is used following the presented algorithms and based on the responses given by each of the 2,693 participants who took part in the item development phase: a random item with an *a-parameter* greater than 1 is shown (20 of the 107 items belonging to various traits met this condition). The response given by the participant in the previous phase is given to this item and the participant's provisional θ is estimated. The procedure is followed until the SE is reduced, a new item is presented only if it is able to reduce the SE by at least 5%; and at least five items are presented. The first criterion makes efficient use of CAT item pools, administering fewer items when predictive gains in information are small and increasing measurement precision when information is abundant (Choi, Graddy, & Dodd, 2011). The latter criterion was selected in order to increase the face validity of the test and because it struck the best balance of highest item savings, and generally fewer costs to validity and accuracy (Rudick, Yam, & Simms, 2013). When reaching both criteria, the participant's ability level and the corresponding SE are estimated. This practice was applied separately to every participant.

In addition, the correlation between the estimations from the CAT and those found from the participants' pooled responses were calculated.

Results

Study 1. Development and study of the psychometric properties of the initial pool of items

Following the experts' assessment, two items were eliminated because they did not reach a minimum level of representativeness (33% inter-judge agreement). In terms of the applicability of the personality traits included in the model, the mean score was 3.85 on a scale of 1 to 5, with external locus of control showing the lowest score (M = 1.9). Aiken's V coefficient of content validity was .71 [.56-.85; CI = 95%], which indicates an acceptable level of agreement (Penfield & Giacobbi, 2004).

By estimating the item discrimination indices of each of the subscales, 15 items were removed using the above-mentioned criteria. Additionally, 17 items were eliminated due to reduced factor loadings.

To test the possibility that these nine specific traits could be defined as a single general factor, a second-order exploratory factor analysis was performed. It seems reasonable to propose the existence of a factor which may be called entrepreneurship (Table 1), which explains 49% of the total variance and has appropriate fit indices (Byrne, 2001).

As previously noted, the external locus of control component was described as not very relevant when it comes to defining entrepreneurship, and it showed a reduced factor loading in comparison with the other traits. Consequently, we decided to

Table 1
Second order exploratory factor analysis of entrepreneurship subscales

	Entrepreneurship
Self-efficacy	.92
Achievement motivation	.88
Innovation	.74
Internal locus of control	.63
Optimism	.60
Autonomy	.55
Risk-taking	.55
Stress tolerance	.49
External locus of control	-.35
Explained variance (%)	49.07
Goodness of FitIndex (GFI)	.97
Root Mean Square Residual [Standard Error]	.07 [.06]

eliminate it from the proposed instrument. As a result, a pool of items comprising 115 items evaluating eight specific personality traits was obtained.

Study 2. Psychometric properties of the eight selected facets

Firstly, an analysis of the items was carried out. Two items were eliminated following the examination of item discrimination indices. The DIF study led to the removal of four further items. Finally, the internal structure analysis of each of the facets led to the deletion of two more items, leading to a final pool of items comprised of 107 items. Following this refinement, the GFI was greater than .95; the RMSR did not exceed .08; and the percentage of variance explained by the factor was at least 30% in all cases, demonstrating an adequate fit (Kline, 2011).

Testing the assumptions of Item Response Theory Model

The assumption of unidimensionality was tested both through exploratory and confirmatory approaches with the aim of verifying whether the pool constituted a single dimension. Firstly, an exploratory factor analysis was carried out on the first subsample which produced the following fit indices. The RMSEA was .045, 90% CI [.044, -.046], the coefficient χ^2/df = 2.75, and the incremental comparative fit index (CFI) was .76. In addition, the defined factor explains 22.16% of the total variance, with the first eigenvalue being 22.89, and the second one 5.38. While the incremental fit index is lower than desired (Kline, 2011), both the value of the standardized residuals and the percentage of variation explained, as well as the ratio between the first and second eigenvalues (greater than 4:1) suggest that unidimensionality may be assumed (Lord, 1980; Reckase, 1979). This result was checked by applying two confirmatory factor analyses with the aim of carrying out a cross-validation. The fit indices in the two subsamples were as follows: χ^2/df_{n_2} = 2.36; χ^2/df_{n_3} = 2.22; $RMSEA_{n_2}$ = .038, 90% CI [.037, -.039]; $RMSEA_{n_3}$ = .037, 90% CI [.036, -.038CFI$_{n_2}$ = .74; CFI_{n_3} = .75. The fit indices met acceptable levels in both cases.

In terms of fit to the Samejima graded response model, standardized residual analysis produced a mean of 0.05 and a standard deviation of 1.16 for all items. These results are close to the ideal values for the distribution of standardized residuals (i.e. $M = 0$, $SD = 1$). When analyzing the results graphically, it can be seen that the distribution of the standardized residuals (Figure 1) forms a function similar to the desired ideal, so it seems reasonable to conclude an acceptable fit of the data to the model.

Estimation of pool item parameters

The calibration of the 107 pool items was performed from the data collected from the application of paper and pencil tests. The *a-parameter* of the items demonstrated appropriate values, distributed between 0.36 and 1.96, with the majority above 0.65 (Baker, 2001). One of the aspects to highlight in this point is that, in most cases, the information contributed by each item is very similar for a range of θ scores. Figure 2 is an example of this. It shows, on the left hand side, the item characteristic curve, and on the right, in the dotted line, the item information function of one of the pool items.

Precision of the pool of items

The precision of the pool of items was assessed, as well as the SE according to the estimated score of the participants (θ). Firstly the test information function was estimated for a range of scores ±4 (Figure 3). In Figure 3, the solid line represents the information contributed by the test while the dotted line shows the SE. The precision of the pool is especially good for θ levels between -3 and +2, and reduced at higher levels. Despite this reduction, the SE is lower than 0.2 for most of the θ range (M = 0.179).

Complementarily, the correlation between the estimated θ from the complete pool of items and the participants' raw score in the paper and pencil test was estimated, providing a value of $r_{\theta x}$ = .974.

Once the precision of the pool of items had been verified by empirical data, its precision was simulated when applied to a larger, more heterogeneous sample. To that end, the answers of 130,000 participants were simulated with the previously stated characteristics. In the same way as the test information function, the precision of the estimation is better for θ levels between -3 and +2. Although the SE increases slightly for very high levels of θ, it has a mean of 0.182 for the total range of estimated scores.

Simulation of performance of CAT

Testing the performance of the CAT (BEPE-A), between 6 and 18 items must be presented, with a mode of 10 items, and highlighting that 13 items or fewer were presented to 99.44% of the participants. A mean SE of 0.378 was obtained (SE_{mode} = 0.354; SE_{median} = 0.355). A larger value in comparison with the SE obtained when applying the complete pool of items (SE_{mean} = 0.174; SE_{mode} = 0.165; SE_{median} = 0.170). However, it is important to note that only about 10% of the items that comprises the total pool are presented.

The correlation between the estimations of θ obtained from the application of the complete pool of items and the adaptive format via the simulation using real data provided r = .91.

Discussion

Reviews in the area of research into entrepreneurs' personalities (Brandstatter, 2011; Zhao et al., 2010) indicate the need to combine the contributions of broad personality traits with specific traits that produce more precise predictions owing to their better predictive capabilities (Leutner et al., 2014). Therefore, it is especially important to have appropriate instruments to measure those traits.

This study is a novel contribution to the assessment of the enterprising personality, with numerous advantages related to the format of the evaluation in terms of speed, precision, and invariance of the properties of the instrument.

Various qualitative and quantitative pilot trials were done with the pool of initial items which allowed the selection of 107 items with appropriate psychometric properties. Unidimensionality and adequate fit to the Samejima graded response model were both confirmed.

Regarding the estimation of item parameters, it is worth commenting that practically any pool item assesses the participants in a similarly precise way, regardless of their level of entrepreneurship. This result has clear advantages in the application of the CAT. Firstly, there is a high probability that all of the items will be presented at some time, reducing a possible item parameter drift (Han, Wells, & Sireci, 2012). In this case, it is worth considering that variation in the application format (paper and pencil vs. computer) and the presentation order of the items could change those parameters (Olea, Abad, Ponsoda, & Ximénez, 2004). In addition, the precision over a wide range of scores permits multiple, short, fixed-length tests to be prepared, and allows applying the instrument to the same participant over short time periods, avoiding the memory effect. This would be especially useful to implement business training in an educational center, allowing a learner to be continuously assessed. Finally, having an informative instrument for a broad range of θ suggests that the estimations of scores are essentially unbiased (Abad, Olea, Real, & Ponsoda, 2002).

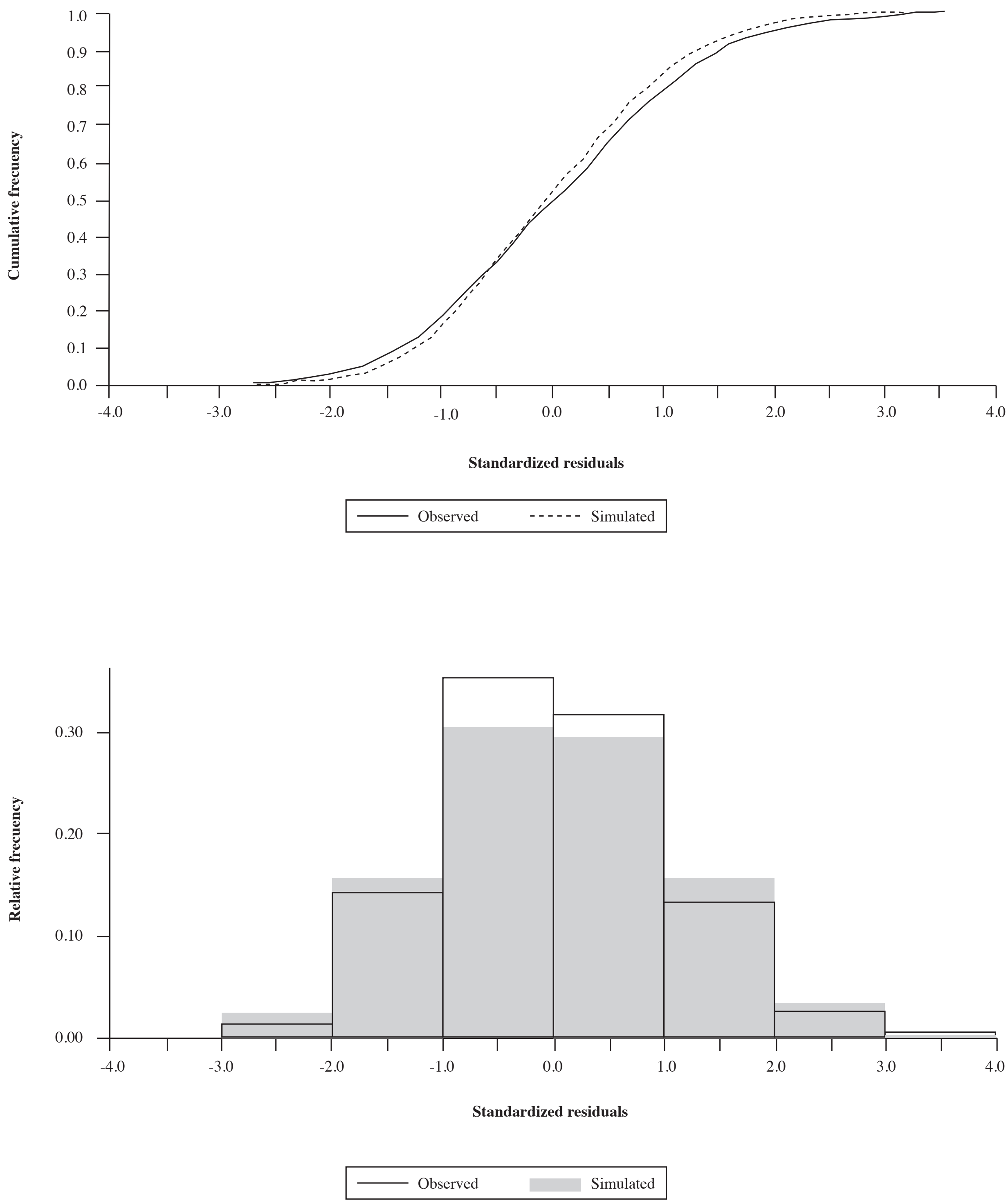

Figure 1. *Fit of the standardized residuals of the pool items to the Samejima Graded Response Model*

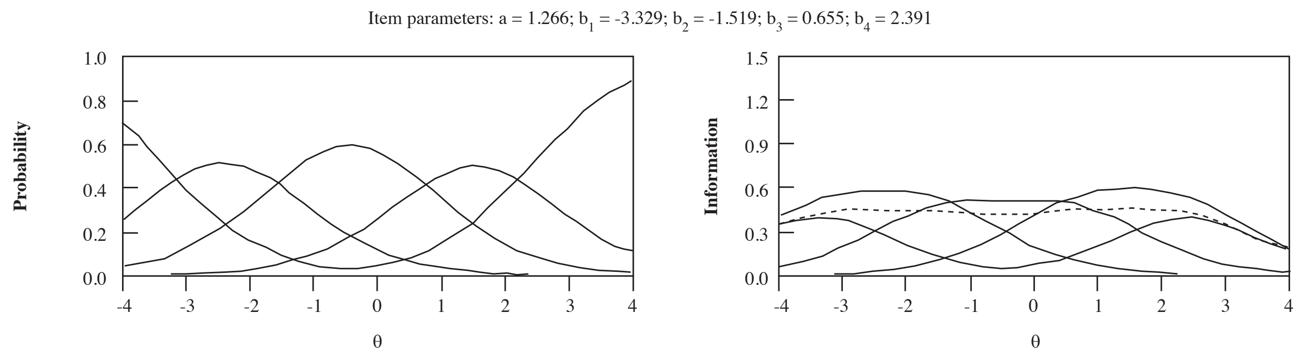

Figure 2. *Example of the Item Characteristic Curve and Item Information Function of one of the items which makes up the pool*

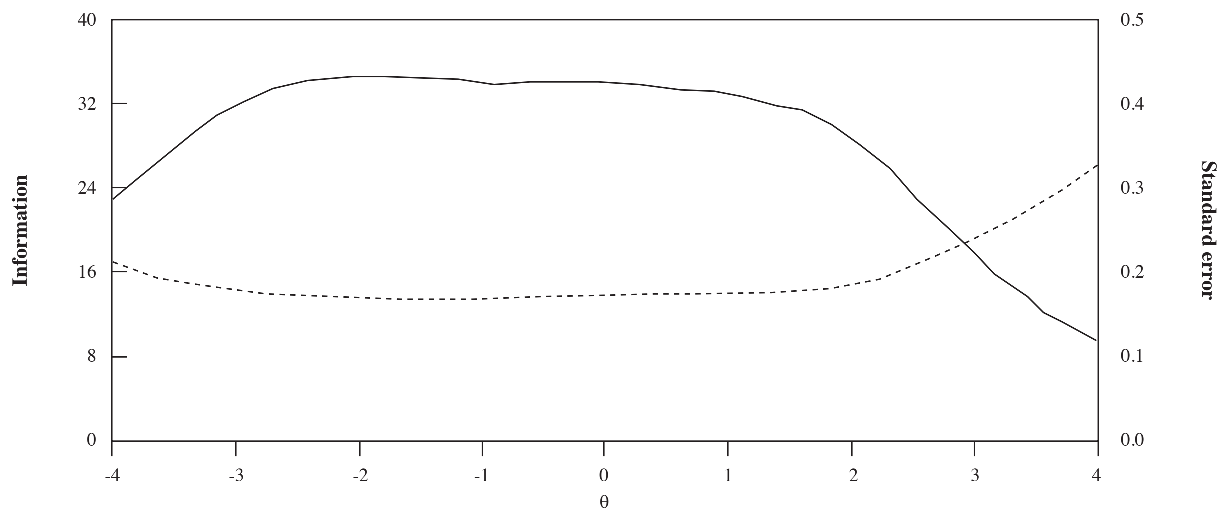

Figure 3. *Test Information Function of the pool of items*

In terms of precision, the pool of items demonstrated a reduced SE over a wide range of θ and was especially precise between -3 and +2, both based on the raw score from the paper and pencil and when the pool was applied to a larger sample size (130,000 participants).

Finally, the CAT (BEPE-A) has demonstrated its precision and brevity. In almost all cases it was possible to achieve a precise assessment (M_{SE} = 0.378) by presenting between 6 and 13 items, with an average of 10 items. Additionally, a clear correlation was obtained between the estimations of θ found by application of the whole pool and the CAT (*r* = .91).

Considering the above results, this current project may prove beneficial in various ways. The results confirm the establishment of an item pool which is appropriately calibrated for the assessment of specific enterprising personality traits. From that pool, the development of the CAT provides professionals with a much faster, more rigorous and more effective assessment of enterprising personality than is currently performed. Furthermore, having developed a computer-based platform opens the possibility of applying online evaluations with all the benefits in terms of location of testing, ease of implementation, and saving on personnel, materials and cost.

Lastly, the application of the CAT allows the inclusion of evaluation of the enterprising personality within the sphere of work and education, improving the precision of forecasts and the early identification of people who have a propensity for business, giving them better academic and professional direction.

One main limitation which must be noted is the lack of criterion validity evidence. Although the instrument has demonstrated appropriate psychometric properties in a population of young people, there is still no clear understanding of the relationship between entrepreneurs who fail and those who are successful after starting a business.

Future lines of research should address various aspects. Firstly, the instrument must be applied to entrepreneurs who have already started businesses, providing evidence of external validity; an optimal cutoff point to distinguish potential entrepreneurs from the general population must be set; and we must understand the instrument's performance and measuring properties in a sample of adults. Furthermore, it is necessary to obtain measures of other individual and socio-economic variables involved in the theoretical model. This would allow an explanation of the effect that each of them has on entrepreneurship and job performance (e.g. Berry & Zhao, 2015). The combined evaluation of broad and specific personality traits will be essential to define the precision and predictive capacity of both approaches when predicting entrepreneurship. Additionally, a multidimensional model should be developed which frames the eight facets defining the enterprising personality (Muñiz et al., 2014). This would provide more precise estimations of scores by considering the relationships between the dimensions (Reckase, 2009). Related to this proposal, an alternative item presentation should be applied as a control method to ensure items related to the eight facets are presented, guaranteeing the content validity of the CAT. Likewise, simulation studies comparing different starting and stopping methods could be useful to increase the efficiency of the CAT. Finally, it would be advisable to complement the data with information from sources other than the individual him/herself or other response formats such forced-choice tests (e.g. Hontangas, Leenen, de la Torre, Ponsoda, Morillo, & Abad, 2016).

Acknowledgments

This study was funded by the Spanish Ministry of Economy and Competitiveness, the University of Oviedo, the FPU program of the Spanish Ministry of Education, Culture and Sport, and the Asociación Española de Metodología de las Ciencias del Comportamiento. References: PSI2011-28638, IB-05-02, UNOV-11-BECDOC, AP2010-1999, PSI2014-56114-P, and AEMCCO-2015.

References

Abad, F. J., Olea, J., Real, E., & Ponsoda, P. (2002). Estimación de habilidad y precisión en tests adaptativos informatizados y tests óptimos: un caso práctico *[Estimation of ability and precision on computerized adaptative test and optimal test: A case study]. Revista Electrónica de Metodología Aplicada, 7*(1), 1-20.

Almeida, P. I. L., Ahmetoglu, G., & Chamorro-Premuzic, T. (2014). Who wants to be and entrepreneur? The relationship between vocational interests and individual differences in entrepreneurship. *Journal of Career Assessment*, *22*, 102-112.

AERA, APA, & NCME (2014). *Standards for educational and psychological testing*. Washington, DC.

Baker, F. (2001). *The basics of item response theory*. University of Maryland: College Park: ERIC Clearinghouse on Assessment and Evaluation

Baum, J. R., Frese, M., Baron, R. A., & Katz, J. A. (2007). Entrepreneurship as an area of psychology study: An introduction. In J.R. Baum, M. Frese & R.A. Baron (Eds.), *The Psychology of Entrepreneurship* (pp. 1-18). London: Lawrence Erlbaum.

Berry, C. M., & Zhao, P. (2015). Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance. *Journal of Applied Psychology, 100*(1), 162-179.

Brandstatter, H. (2011). Personality aspects of entrepreneurship: A look at five meta-analyses. *Personality and Individual Differences, 51*(3), 222-230.

Byrne, B. M. (2001). Structural equation modeling with AMOS. Basic concepts, applications, and programming. In K.A. Bollen (Ed.), *Testing structural equation models* (pp. 136-162). Londres: SAGE.

Choi, S. W., Graddy, M. W., & Dodd, B. G. (2011). A new stopping rule for computerized adaptive testing. *Educational and Psychological Measurement, 71*(1) 37-53.

Elosua, P., & Zumbo, B. D. (2008). Reliability coefficients for ordinal response scales. *Psicothema, 20*, 896-901.

Gómez-Benito, J., Hidalgo, M. D., & Zumbo, B. D. (2013). Effectiveness of combining statistical tests and effect sizes when using logistic discriminant function regression to detect differential item functioning for polytomous items. *Educational and Psychological Measurement, 73*(5), 875-897.

Han, K. T., Wells, C. S., & Sireci, S. G. (2012). The impact of multidirectional item parameter drift on IRT scaling coefficients and proficiency estimates. *Applied Measurement in Education, 25*(2), 97-117.

Hontangas, P. M., Leenen, I., de la Torre, J., Ponsoda, V., Morillo, D., & Abad, F. J. (2016). Traditional scores versus IRT estimates on forced-choice tests based on a dominance model. *Psicothema, 28*, 76-82.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Lane, S. (2014). Validity evidences based on testing consequences. *Psicothema, 26*, 127-135.

Leutner, F., Ahmetoglu, G., Akhtar, R., & Chamorro-Premuzic, T. (2014). The relationship between the entrepreneurial personality and the Big Five personality traits. *Personality and Individual Differences, 63*, 58-63.

Liang, T., Han, K. T., & Hambleton, R. K. (2008). *User's guide for ResidPlots-2: Computer software for IRT graphical residual analyses, Version 2.0*. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Liñán, F., & Fayolle, A. (2015). A systematic literature review on entrepreneurial intentions: Citation, thematic analyses, and research agenda. *International Entrepreneurship and Management Journal, 11*(4), 907-933.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum.

Miller, D. (2015). A downside to the entrepreneurial personality? *Entrepreneurship Theory and Practice, 39*(1), 1-8.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems [Introduction to Item Response Theory]*. Madrid: Pirámide.

Muñiz, J., Fidalgo, A. M., García-Cueto, E., Martínez, R., & Moreno, R. (2005). *Análisis de los ítems [Item analysis]*. Madrid: La Muralla.

Muñiz, J., Suárez-Álvarez, J., Pedrosa, I., Fonseca-Pedrero, E., & García-Cueto, E. (2014). Enterprising personality profile in youth: Components and assessment. *Psicothema, 26*, 545-553.

Moreno, R., Martínez, R. J., & Muñiz, J. (2015). Guidelines based on validity criteria for the development of multiple choice items. *Psicothema, 27*, 388-394.

Obschonka, M., Stuetzer, M., Gosling, S. D., Rentfrow, P. J., Lamb, M. E., Potter. J., & Audretsch, D. B. (2015). Entrepreneurial regions: Do macro-psychological cultural characteristics of regions help solve the "knowledge paradox" of economics? *PLoS ONE 10*(6): e0129332.

OECD/The European Commission (2013). *The missing entrepreneurs: Policies for inclusive entrepreneurship in Europe*. Brussels: OECD Publishing.

Olea, J., Abad, F. J., Ponsoda, V., & Ximénez, M. C. (2004). Un test adaptativo informatizado para evaluar el conocimiento del inglés escrito: diseño y comprobaciones psicométricas [A computerized adaptive test to assess knowledge of written English: Design and psychometric testing]. *Psicothema, 16*, 519-525.

Pedrosa, I., Suárez-Álvarez, J., & García-Cueto, E. (2013). Content validity evidences: Theoretical advances and estimation methods. *Acción Psicológica, 10*(2), 3-18.

Penfield, R., & Giacobbi, P. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, *8*(4), 213-225.

Rauch, A., & Frese, M. (2007a). Born to be an entrepreneur? Revisiting the personality approach to entrepreneurship. In J. R. Baum, M. Frese & R.J . Baron (Eds.), *The psychology of entrepreneurship* (pp. 41-65). Mahwah, NJ: Erlbaum.

Rauch, A., & Frese, M. (2007b). Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success. *European Journal of Work and Organizational Psychology, 16*(4), 353-385.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207-230.

Reckase, M. D. (2009). *Multidimensional item response theory. Statistics for Social and Behavioral Sciences*. London: Springer

Rudick, M. M., Yam, W. H., & Simms, L. J. (2013). Comparing countdown- and IRT-based approaches to computerized adaptive personality testing. *Psychological Assessment, 25*(3), 769-779.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Vol. No. 17). Richmond, VA: Psychometric Society.

Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidences based on test content. *Psicothema, 26*, 100-107.

Suárez-Álvarez, J., & Pedrosa, I. (2016). Enterprising personality assessment: Current status and future directions. *Papeles del Psicólogo, 37*(1), 62-68.

Suárez-Álvarez, J., Pedrosa, I., García-Cueto, E., & Muñiz, J. (2014). Screening enterprising personality in youth: An empirical model. *The Spanish Journal of Psychology, 17*, 1-9.

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods, 16*(2), 209-220.

Van der Linden, W. J., & Glas, C. A. W. (2010). *Elements of adaptive testing*. London: Springer.

Zhao, H., Seibert, S. E., & Lumpkin, G. T. (2010). The relationship of personality to entrepreneurial intentions and performance: A meta-analytic review. *Journal of Management, 36*, 381-404.