Special Collection: Challenges in Translating and Adapting Psychological Measures to Spanish/Portuguese

Original Article

Verbal and Nonverbal Culture-Fair Challenges

A Task-Switching Assessment Test With a Portuguese Sample

Beatriz Regina de Oliveira Trigo[1], Aristides I. Ferreira[2], and Paulo Almeida[3]

[1]Product Quality & Research, Enhanced Fertility, EFP Clinic, Lisbon, Portugal
[2]Business Research Unit, Iscte – Instituto Universitário de Lisboa, Lisbon, Portugal
[3]Neotalent Conclusion, Lisbon, Portugal

Abstract: Optimizing selection processes is a multilayer beneficial achievement. Moreover, task-switching is a commonly required ability in a variety of jobs (e.g., programmers). The present study aims to develop a rapid task-switching assessment for the Portuguese-speaking population that integrates verbal content with nonverbal components. One hundred and sixty-four participants answered the task-switching test, and 78 of those were evaluated in terms of their overall job performance in 2020, 2021, and 2022 by their managers. Fifteen items were kept in the final version of the tool, across three tasks. Statistical analyses showed there was positive evidence for the acceptable psychometric properties of the assessment tool and how the task-switching performance of the test correlated to job performance.

Keywords: task switching, assessment, candidate selection, human resources management, item response theory

Multitasking has become increasingly required of workers (Appelbaum et al., 2008; Morgan et al., 2013; Russ & Crews, 2014), with software developers being a prime example (Vasilescu et al., 2016). The need to multitask increased across a variety of jobs and occupations, particularly in knowledge-intensive business services (Suija-Markova et al., 2020) where ICT skills and resources are more prominent (Carrier et al., 2015; Crews & Russ, 2020). The aim of this present study was to develop a task-switching test for the Portuguese-speaking population. In this current study, following Ziegler's (2014) ABC of test construction and subsequent work developed by Ziegler (2020) about how a paper should be structured, we explain that the construct being measured is task-switching and the theories behind it. Then, we explain that HR selection processes are the intended uses of the measurement. Finally, in line with the aims of this paper, we developed a task-switching construct specific to the target Portuguese-speaking population. Evidence of structural and convergent validity will be provided in accordance with the guidance provided by Ziegler (2020).

This test integrates verbal content with nonverbal tasks. This approach, when developed or adapted to other native languages, may provide a good alternative to other non-cultural bias assessment tools if it includes simple, familiar words of low complexity in the language of the person being assessed (e.g., Portuguese). Beyond the need to avoid and minimize bias, it is important to consider a more comprehensive assessment tool for rapid task-switching skills. Assessment tools combining verbal and nonverbal tasks have been developed specifically for ICT staff; however, none of the verbal tasks have been widely adapted for use in languages other than English.

Cattell (1940), Groth-Marnat (2003), Rock and Price (2019), and Aghvinian et al. (2021) explain how an IQ score may not correlate to intelligence due to unaccounted cultural biases. For example, if a low score is obtained by a participant with a cultural background different from the one in which the IQ test was developed, conclusions drawn regarding their intelligence level may be unfounded and/or misinterpreted. These authors' findings illustrate the complexity of constructing assessment tools and that for results to be successfully captured, a variety of items should be adapted to account for cultural factors and linguistic diversity.

Given the culture-fair test model that considers the existence of two broad group factors corresponding to verbal-educational and practical-mechanical and spatial and visual aptitudes (e.g., Vernon, 1964), we believe that the addition of verbal content to such tests provides an interesting approach for psychological assessment.

Task-Switching

Multitasking is an ability required of certain occupations and professions (Vasilescu et al., 2016). Information and communication technology (ICT) professionals are commonly under pressure to perform multiple tasks, like conceptualizing, building, reviewing, testing, retesting, and launching products, and all of this under time constraints. These can be executed in a parallel (at the same time) or serial (sequential) way (Fischer & Plessow, 2015).

Current multitasking research focuses on dual-task performance and task-switching performance (Lee & Taatgen, 2002; Oswald et al., 2017; Pashler, 2000). The current article focuses on the latter. "The stimuli presented in task-switching situations afford not only the currently relevant task but also the other task(s)" (Strobach et al., 2018, p. 3), which involves two (or more) different tasks being performed sequentially (Kock et al., 2018) occurring in a predictable or unpredictable sequence, without temporal overlap (Kiesel et al., 2010).

From the nomological perspective, we conceptualize the construct of task switching as a complex multivariate dimension associated with the Cattell–Horn–Carroll (CHC) model of intelligence (W. J. Schneider & McGrew, 2012). This model of intelligence comprises 16 broad cognitive abilities measuring reasoning (G*f*), acquired knowledge (G*c*, G*kn*,, G*q*, G*rw*), memory and efficiency (G*sm*, G*lr*), sensory (G*c*, G*a*, G*o*, G*h*), motor (G*p*, G*k*), and speed and efficiency (G*s*, G*t*, G*ps*), which are subsumed by over 80 narrow abilities.

Drawing on this model of intelligence, we aim to develop a task-switching task that evaluates several cognitive domains from the CHC model of intelligence. Specifically, a complex task that integrates a knowledge/vocabulary (G*c*) test, a numerical reasoning (G*q*) test, a visual perception (G*a*) test, a velocimeter test to evaluate processing speed (G*s*), and finally the measurement of supervision and coordination ability associated with working memory processing (G*sm*), as suggested by Oberauer et al. (2003).

The ability to do so requires high-level cognitive processing (D. W. Schneider & Logan, 2009). Some researchers believe the switching cost associated with task switching to be due to the reconfiguring of the cognitive system to allocate resources to solve different tasks rapidly (Aagaard, 2019; Rogers & Monsell, 1995), but other researchers claim that after a task has been performed, the cognitive activation persists over time, which facilitates responding when the same task is repeated (Gilbert & Shallice, 2002). Liefooghe et al. (2010) found that there is less persisting activation in voluntary task-switching (i.e., random sequences of tasks are selected) than in task-switching procedures in which task sequences are indicated externally (i.e., with instructions). So, with instructions, task-switching costs decrease.

Job Performance

Employees of knowledge-intensive business services are commonly engaged in multitasking activities in their workplaces, reporting up to nine tasks per day and multiple interruptions caused by external and internal factors (Suija-Markova et al., 2020). The link between multitasking and productivity is context dependent because some jobs require more multitasking and/or are more interrupt-driven than others (Bannister & Remenyi, 2009).

Ophir et al. (2009) found that heavy media multitaskers were more susceptible to interference from irrelevant stimuli, i.e., they showed reduced ability to filter out, which made them perform worse on a test of task-switching ability. Moreover, Kohl et al. (2020) found that some developers associated multitasking behavior with losing focus, anxiety, procrastination, decreased code quality, and decreased productivity (in proportion to the quantity of context switching). Accordingly, Suija-Markova et al. (2020) affirmed that employees who multitask intensively report experiencing mental and physical tiredness, and those who work longer hours dislike multitasking and report lower well-being.

Therefore, besides the obvious need to design and implement protocols that prevent burnout, exhaustion, and diminished well-being, as well as control systems that verify code quality, it is also necessary to ensure that selection processes take into account a candidate's predisposition to manage stressful situations and yet be able to rapidly switch between and manage different tasks.

A Test for the Portuguese Population

The present study developed a task-switching assessment, included in Neotalent's Digital Talent Ecosystem (DTE; LISBOA-01-0247-FEDER-045216), which aims to set up an integrated web platform to digitize the talent market and the interaction between its various stakeholders, such as candidates, professionals (i.e., the talents), consumers, suppliers, and managers. The DTE is an information system comprising five modules: talent assessment, talent recommendation, team recommendation, career recommendation, and career logbook. The talent assessment module is equipped with a battery of tests, one of which is the rapid task-switching test.

This test, in the form of a computerized game, requires the simultaneous performance of four unrelated hard skills tasks. Action video games can increase a person's ability to take on additional tasks by increasing attentional capacity, i.e., enhancing performance on secondary tasks without interfering with the primary tasks (Chiappe et al., 2013).

The developed test follows a computerized adaptive test (CAT) format. The CAT is a computer-administered test that tailors test questions to each candidate's skill level (Van der Linden & Glas, 2000). The individual's proficiency, i.e., latent trait or the individual's skill in a knowledge area, is estimated during the administration of the test, and only items that efficiently measure the candidate's proficiency are selected. Because each item presented to the candidate is tailored to their specific aptitude, no item administered in the test is irrelevant. Fewer items are required than in a traditional test format for the same level of precision (Van der Linden & Glas, 2000).

Participants had to solve visual perception items (G*a*), word knowledge/vocabulary items (G*c*), numerical reasoning items (G*q*), and mental speed items (G*s*). Inspiration was drawn from SynWin (Hambrick et al., 2010), Control Tower (Redick et al., 2016), and Air Traffic Control Lab (Fothergill et al., 2009). The tasks appear on the screen in a nonpredictable sequence (Rogers & Monsell, 1995).

The mix of tasks has verbal and nonverbal cues specifically adapted for the Portuguese population. The verbal components of the tasks include not just the linguistic aptitude-related items but the instructions to complete the tasks themselves. It should be noted that inappropriate content, improper standardization samples, and inequitable social consequences, as pointed out by Reynolds (1998), are frequent problems that contribute to linguistically biased results.

According to Van de Vijver and Tanzer (2004), the variety of problems regarding this issue can be divided into three categories: construct, method, and item bias. "The use of a test in a new linguistic culture requires that it be redeveloped from the start" (p. 105). This is because it requires accommodations to structural differences or distinct meanings that, if not met, lead to unreliable assessments (Reynolds & Suzuki, 2012). Van de Vijver and Tanzer (2004) advise the use of bilingual or language-proficient experts to translate the items from one language to another. They further suggest eliminating examiner bias by making the test computerized, that there be detailed example-like instructions for items be included, and psychometric methods of item bias detection.

It is important to add that everyone can register, have access to the test, and answer the items (https://dte-talent.neotalent.pt). The test, however, belongs to Neotalent. Its usage for scientific or business purposes requires a formal request to Neotalent (https://www.neotalent.pt/).

Method

The study was registered in the Center for Open Science (https://doi.org/10.17605/OSF.IO/N8YSM, preregistration: https://osf.io/2xjf6, databases: https://osf.io/uyw64/). The DTE was made available in January 2021, with 164 answers collected until February 2022.

Sample

Lakens (2022) discusses six approaches to justify the sample size in a quantitative empirical study. One approach is based on collecting data from an entire population. Since the target population was Neotalent's ICT workers, we can affirm we collected data from almost their entire population. Another approach discussed by Lakens (2022) is choosing a sample size based on resource constraints, which is also applicable to the present project. Lastly, we also performed an a priori power analysis (Lakens, 2022), using https://www.danielsoper.com/statcalc/calculator.aspx?id=89, obtaining a recommended minimum sample size of 100.

All participants (N = 164) were IT workers or graduates, the large majority being developers, analysts, or consultants. The mean age of the sample is 38 years (SD = 9.7). Nationality was not important if participants were IT workers and able to communicate in Portuguese. Although participation was open to anyone who met the inclusion criteria, most were working for, or in the process of, being selected to work for Neotalent.

The current rapid task-switching is not exclusive to ICT workers, but due to (1) the increased need to recruit and retain these individuals (Bresnahan, 2002; Dzyubenko, 2021; Lockwood, 2006), (2) the difficulty task managers and recruiters face when trying to do so (Naqvi & Bashir, 2015), (3) the need to optimize selection processes and human resource allocation in ICT companies (J. Wright & Atkinson, 2019), and (4) the fact that task-switching is a required trait to perform ICT tasks (Vasilescu et al., 2016), the sample criteria included the specific requirement that participants were ICT workers.

Measures

Regarding the task-switching ability, to avoid undesirable effects, once participants had started the rapid task-switching

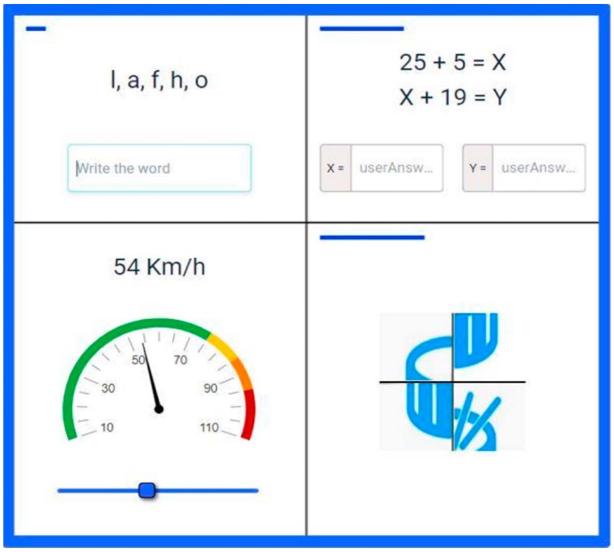

Figure 1. Example of a level of the rapid task-switching test.

test, they had to complete it in one go; in other words, they could not start, stop, and restart the task. However, the participants were advised not to answer the tests in the battery consecutively. The instructions for the test were in Portuguese, included examples, and provided information on the increasing difficulty of the tasks' sequence.

The test has five levels of increasing difficulty across all four subtasks, which appear on the screen, divided into four parts (Figure 1). In the top left-hand corner are the word knowledge/vocabulary items (G*c*), i.e., solve a word puzzle in which the participant is given a set of letters that they must arrange into a word that contains all the letters. The first level starts with four letters, and one is added per level, ending with eight in the fifth and final levels. In the top right-hand corner are the numerical reasoning items (G*q*), i.e., solve mathematical equations which at the lower levels consist of adding and subtracting and at the higher levels of multiplication and division. In the bottom right-hand corner are the solve visual perception items (G*a*), i.e., solve four-piece puzzles. The image becomes progressively more complex (the first level has an almost symmetrical image whereas the other levels do not). Lastly, in the bottom left-hand corner are the mental speed items (G*s*), i.e., a velocimeter that requires the participant to control its speed. The velocimeter becomes more *unstable*, changing velocities more often, as the levels get progressively more difficult.

The subtask levels advance as each participant completes the previous one. An inability to complete a level of any subtask within the time constraint or an inability to control the velocimeter would end the exercise. Participants receive one point for each correct answer. The task ends when the participant shows an inability to control and stabilize the velocimeter or when two consecutive wrong answers are made in the same task.

Furthermore, the participants who were working for Neotalent, and not just candidates, were evaluated regarding their job performance in 2020, 2021, and/or 2022 (depending on when they were hired). From the 164 participants who completed the task-switching assessment tool, 78 participants were evaluated regarding their job performance.

This evaluation was performed by Neotalent's HRM department. There were three types of evaluation: SKIN, HCM, and NOVA. SKIN is scored as 3 = *outstanding* (highest rating), 2 = *well-done* (medium rating), and 3 = *great* (lowest rating). HCM is scored as 1 = *A* (highest rating), 2 = *B* (medium rating), and 3 = *C* (lowest rating). NOVA is a numerical scale, in which the average score changes every year according to the evaluation scores assigned; to attribute the same method of ratings of 1 (*highest rating*), 2 (*medium rating*), and 3 (*lowest rating*), we converted the NOVA scored to *Z*-scores, and those who had below −1 *SD* were assigned a 3, those who had between −1 and 1 *SD* were assigned a 2, and those who had above +1 *SD* were assigned a 1.

Procedure

Previous authorization was sought from Neotalent to start the data collection. An ethics review board from ISCTE Instituto Universitário de Lisboa and Instituto Pedro Nunes approved the conduct and design of the study before it was implemented. All the participants signed an individual consent form that outlined a general description of the study, its title and aim, and their participation. Also outlined were any potential risks and benefits.

Van de Vijver and Tanzer's (2004) and Reynold's (1999) strategies to minimize biases and the indications for scale development such as the ones provided by Edelen et al (2007), Ferreira et al (2012), and Yang and Kao (2014) were followed. In all, the authors took into account potential construct bias (e.g., incomplete or partial coverage of relevant theoretical aspects of the construct), method bias (e.g., ambiguous instructions, communication, and familiarity with the items), and item bias (e.g., ambiguous items and cultural specificities). Furthermore, the psychometric properties of the rapid task-switching measure were evaluated, considering the classical test theory and the adoption of exploratory factorial analysis with factor (Ferrando & Lorenzo-Seva, 2017) and the Item Response Theory with WINSTEPS software (Linacre, 2005).

The hot-deck multiple imputation method was used to handle missing values in the exploratory factor analysis (Lorenzo-Seva & Van Ginkel, 2016). Unrestricted factor analysis was conducted with robust unweighted least squares and robust promin rotation (Lorenzo-Seva & Ferrando, 2019). Optimal implementation of parallel analysis (PA) was the procedure used to determine the number of dimensions (Timmerman & Lorenzo-Seva, 2011).

The item response theory with Rasch measurements was used to assess unidimensionality by using item and person fit values. According to Linacre (2002), outfit values between 0.5 and 1.5 suggest that the measure is productive for measurement. Values between 1.5 and 2 and below 0.5 mean that the item is less productive for measurement but cannot be degraded for construction. Values higher than 2.0 would show a severe misfit and would distort or degrade the measurement system. Linacre (2002) mentions that this misfit may be caused by only one or two observations.

The overall score obtained from solving the task-switching test was correlated with the job performance (i.e., supervisory rating) using Spearman's ρ because the sample did not follow a normal distribution.

Results

Before construct validity analyses, we followed the procedures of item development as suggested by Ziegler (2020). Accordingly, we developed a nomological network to represent the constructs of interest in the study (i.e., task switching considering verbal and nonverbal contents). The items were developed by two experts in the cognitive psychology and psychometric domains who took into account the recent advances around the construct (e.g., Lee & Taatgen, 2002; Kiesel et al., 2010; Kock et al., 2018; Oswald et al., 2017; Pashler, 2000; Strobach et al., 2018). After the initial item development phase, 15 items were considered for the next step. Then, items were submitted to five experts in the psychological assessment and cognitive psychology domains (M_{age} = 40.4 years) to assess the appropriateness of the items, clarity of instructions, and suitability of the items for the purpose of selecting people. A scale of 1 to 3 was used (1 = *highest rating*, 2 = *medium rating*, 3 = *lowest rating*), and consequently all the items were rated as being appropriate, suggesting content validity.

To measure construct validity, the Bartlett's sphericity test used to check whether there was an underlying structure for data revealed a value of 1,297.3 (df = 105; p < .01). Additionally, the Kaiser–Meyer–Olkin (KMO) was .900 and the bootstrap 95% confidence interval was KMO (.838; .906), which together suggest very good sample adequacy.

The retention criteria (PA based on the minimum rank factor analysis) suggested one factor with 45.95% explained variance, as seen in Table 1. However, for item 1, the loading was only .259 with a measure of sampling adequacy (MSA) below .500 due to a bootstrap 95%

Table 1. Factor loadings and person correlations

Variable	Task-switching factor (loadings)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
WORDS 1	.259	—													
WORDS 2	.669	.367	—												
WORDS 3	.804	.238	.717	—											
WORDS 4	.755	.153	.539	.693	—										
WORDS 5	.338	.038	.260	.325	.364	—									
MATH 1	.431	.049	.268	.213	.235	.143	—								
MATH 2	.816	.247	.598	.659	.569	.258	.503	—							
MATH 3	.899	.183	.557	.706	.697	.318	.354	.695	—						
MATH 4	.839	.257	.509	.622	.665	.238	.340	.627	.830	—					
MATH 5	.725	.151	.398	.551	.551	.307	.361	.554	.649	.757	—				
PUZZLE 1	.489	.040	.275	.326	.311	.069	.369	.413	.459	.388	.288	—			
PUZZLE 2	.543	.122	.312	.442	.320	.154	.326	.428	.460	.383	.399	.402	—		
PUZZLE 3	.646	.194	.417	.473	.474	.115	.306	.597	.624	.516	.386	.381	.394	—	
PUZZLE 4	.569	.115	.344	.422	.429	.227	.225	.462	.526	.489	.427	.333	.389	.304	—
PUZZLE 5	.646	.143	.377	.526	.487	.169	.223	.524	.566	.539	.514	.327	.389	.479	.367
Eigenvalue	6.892														
% of variance	45.95%														
EAP reliability	.944														

Note. EAP = expected a posteriori.

confidence interval (.408; .881). Hence, it is suggested that since item 1 (WORDS 1) apparently does not measure the same domain as the remaining items in the pool, it should be removed (Lorenzo-Seva & Ferrando, 2021).

Nonetheless, robust goodness of fit statistics after Losefer correction (Lorenzo-Seva & Ferrando, 2023) shows a fair root-mean-square error of approximation (RMSEA) of .064 and a bootstrap 95% confidence interval between .063 and .073. Regarding good RMSEA values, "a cut-off value close to .06 (Hu & Bentler, 1999) or a stringent upper limit of .07 (Steiger, 2007) seems to be the general consensus amongst authorities in this area." (Hooper et al., 2008, p. 54).

Multivariate Construct Validity

To test a possible multivariate construct measuring Words, Math, and Puzzles as independent constructs, a confirmatory factorial analysis (CFA) using AMOS 27.0 was conducted with unweighted least squares estimates. The CFA is a multivariate statistic that allows to estimate the internal structure of latent variables, verifying how well the hypothesized model represents the number of constructs in the statistical analysis. Moreover, despite the adoption of previous exploratory factor analysis (EFA) to test the internal structure of the measure, CFA is recommended to verify the alternative dimensionality of instruments within the same sample (Lance & Vandenberg, 2002). The literature suggests two broad groups of factors with verbal and nonverbal dimensions (Vernon, 1964). Moreover, task switching is a complex multivariate dimension (W. J. Schneider & McGrew, 2012). Therefore, models considering three independent and covariate latent variables (i.e., Words, Math, and Puzzle) were considered. Results suggest that the data did not fit the hypothesized independent model (χ^2 = 108.767, RMR = .075, SRMR = .325, GFI = .342). As expected, and in line with the previous EFA, the covariate model suggests good fit values (χ^2 = 1.987, RMR = .010, SRMR = .0499, GFI = .991).

Item Response Theory

Table 2 shows the person and item summary statistics for the rapid task-switching game. The findings reveal that the infit and outfit mean per person and item are adjusted to, or close to, one (a perfect fit). The average measure per person reveals proximity with Rasch measures for items, which suggests an adjustment between item difficulty and subject performance. We have two items: Words 1 and Puzzle 1 with an outfit above the corrected critical range, as well as item Puzzle 2 with an infit above the corrected critical range (INFIT = 1.61).

Table 2. Person and item summaries Rasch measurements

Items	Level of difficulty	Infit (MNSQ)	Outfit (MNSQ)	Proportion of correct answers (%)
WORDS 5	3.70	1.21	1.51	12.65
WORDS 4	−.71	.93	.78	42.77
WORDS 3	−.70	.75	.48	57.83
WORDS 2	−1.74	.93	.68	68.67
WORDS 1	−4.59	1.00	3.99	92.77
PUZZLE 5	1.73	.97	.75	31.52
PUZZLE 4	2.18	.99	.76	26.51
PUZZLE 3	.76	1.19	1.14	42.17
PUZZLE 2	.32	1.61	1.34	46.99
PUZZLE 1	−2.15	1.24	4.50	72.89
MATH 5	1.52	.78	.44	33.73
MATH 4	.49	.66	.40	45.18
MATH 3	.15	.49	.28	48.80
MATH 2	−.13	.71	.44	51.81
MATH 1	−2.27	1.50	1.75	74.10
Mean	.00	1.00	1.28	49.91
SD	1.99	.30	1.24	
Max.	3.70	1.61	4.50	92.77
Min.	−4.59	.49	.28	12.65

Note. MNSQ = mean-square standardized. *SD* = standard deviation.

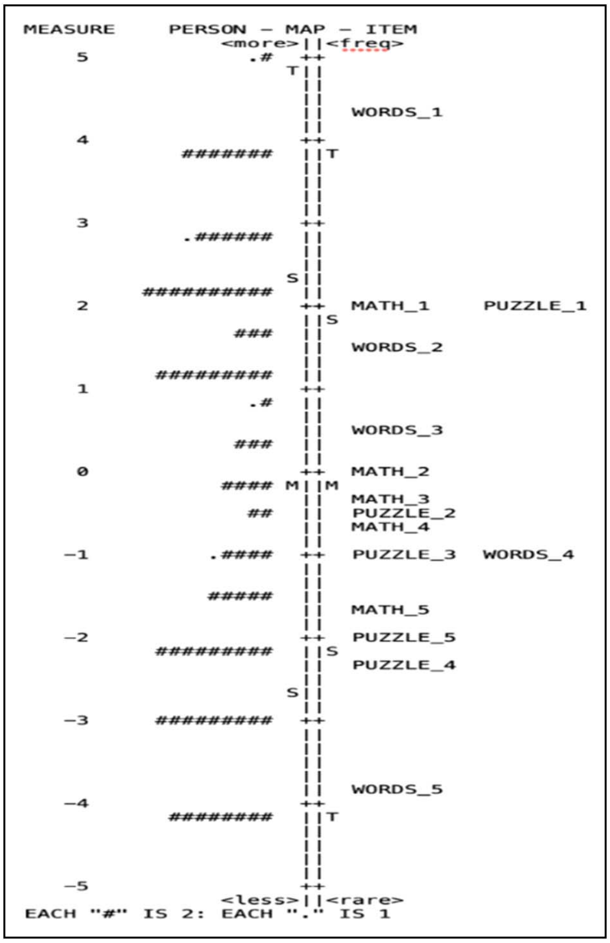

Figure 2. Person–item map for the rapid task-switching test.

The difficulty estimates for rapid task-switching range from −4.59 (WORDS 1) to 3.70 (WORDS 5), with WORDS 1 representing the easiest item and WORDS 5 the hardest. We also adopted person separation reliability (PSR) and item separation reliability (ISR) procedures as reliability indicators derived from Rasch measures. The PSR refers to the proportion of the sample variance that is not explained by the measure error. The ISR corresponds to the item percentage of the item variance that is not explained by the measure error (Smith, 2001).

Findings suggest high scores of PSR (0.85) and ISR (0.98) and a Cronbach's coefficient α of .91. The estimated reliability for the EAP distribution was 0.94. Figure 2 shows that items MATH 1 and PUZZLE 1 had quite similar levels of difficulty (difficulty of −2.27 log and −2.15 log, respectively). The same occurs for items PUZZLE 3 and WORDS 4 (difficulty of 0.76 log and 0.71 log, respectively).

Fifteen items were kept in the final tool (five per subtask). At the end of the exercise, a total score regarding participants' overall performance from 0 to 100 was calculated. The mean score was 53.72 (*SD* = 28.21, Mo = 80.00), following a non-normal distribution (kurt = −1.51, skew = −.001). Each individual subtask comprises five levels, with each coded as 0 when not completed or 1 when completed. This allowed the elaboration of a summed score for each subtask, varying between 0 and 5. With regard to the way in which participants performed on the subtasks, the mean of the summed scores of the subtask *Math* was 2.520 (*SD* = 1.967, kurt = −1.612, skew = .058), the mean of the subtask *Words* was 2.750 (*SD* = 1.528, kurt = −1.226, skew = −.228), and the mean of the subtask *Puzzle* was 2.180 (*SD* = 1.681, kurt = −1.285, skew = .108).

Predictive Validity

Almost half the sample (47.56%) was evaluated in terms of their job performance. Spearman's rank correlation was computed to assess the relationship between job performance and task-switching ability. There was a positive and significant correlation between the two variables ($r_{(76)}$ = .23, *p* = .039).

Discussion

Task-switching ability is required of some people with more demanding jobs (Vasilescu et al., 2016), like ICT staff. Current multitasking research focuses on task-switching performance (Lee & Taatgen, 2002; Oswald et al., 2017; Pashler, 2000), as does the current article. The present study, therefore, aimed to develop a rapid task-switching assessment for Portuguese-speaking communities.

Verbal components in culture-fair tests are required, making it necessary to consider and incorporate them when developing a psychological assessment tool. Thus, taking these into consideration may mean that participants will react better to the selection tool (Balcerak & Woźniak, 2021), which is beneficial not just to the candidate but also to the employer (J. Wright & Atkinson, 2019). This article seeks to address a very significant gap in the field of psychological assessment in Portuguese-speaking countries. Essentially, the existence of few computerized tests adapted to the new reality of human resources digital transformation (Brock & Buckley, 2013; Srivastava & Agarwal, 2012). On the other hand, companies are increasingly looking for tests with high predictive ability of performance and that can be applied quickly. Embodied in the CHC model of intelligence (W. J. Schneider & McGrew, 2012), task-switching seems to encompass several facets of human cognition into a single general factor, which could multiply the predictive value of cognitive skills if evaluated alone.

A sample (N = 164) was shown to be adequate. The one-factor structure observed, with 45.95% explained variance, suggests that although the subtasks used to assess overall rapid task-switching differed in their demands, a coherent construct specific to task-switching ability was identified. This corroborates the fact that rapid task-switching ability is a construct composed of verbal and nonverbal components. Moreover, the construct shows the variety and complex nature of the possible tasks and distractions that people can be exposed to, which they need to be able to handle to multitask. Cronbach's α of .91 showed excellent internal consistency reliability (Cronbach, 1951). Streiner (2003) sets the maximum value for this coefficient at .90. "If α is too high it may suggest that some items are redundant as they are testing the same question but in a different guise" (Tavakol & Dennick, 2011, p. 54).

Our findings reveal a perfect fit and an adjustment between item difficulty and subject performance. The item Puzzle 2 has an infit above the corrected critical range, presupposing its poor performance with regard to the people it is targeted at (B. D. Wright & Linacre, 1998; Institute for Objective Measurement, n.d.). Such a result may suggest the need to iterate Puzzle 2 to better match it to Portuguese-speaking IT workers. Nevertheless, a larger sample size that includes people from other companies may provide further insight into the performance of this item. No issues were detected regarding the size of our sample.

Furthermore, the items Words 1 and Puzzle 1 have an outfit above the corrected critical range, suggesting a misfit (B. D. Wright & Linacre, 1998). Wilson (2005) and Prieto and Delgado (2007) indicate that the outfit is very sensitive to extreme responses or values. Such a result may suggest participants' lack of attention, dedication/interest, or understanding of the exercise, causing them to make errors and produce unexpected answers. When revising these items, it would also be useful to check whether their verbal content is suited to the target population.

This is the first construction and validation study of a computerized rapid task-switching test; new investigations with other samples should continue in light of the results presented here. It is also relevant to mention the correlation between task-switching performance and job performance was positive and significant, which further shows the utility of such a test.

Answering the task-switching test took no more than 10 min per participant, but there was no control over how participants chose to tackle the exercises included in the DTE, which could take up to an hour to complete in one go. Time-consuming tests are not beneficial for the participants' performance (Burchell & Marsh, 1992; Donnellan et al., 2006; Fowler & Fowler, 1995), which could lead to poor sample compliance and biased data collection (McKnight et al., 2007; Moore et al., 2002).

Further work is needed on how applicants from different countries, including Portugal, react to preferences for certain selection tools. Anderson and Witvliet (2008) found "high levels of similarity between applicant reactions across studies covering countries as culturally diverse as the Netherlands, the United States, France, Spain, Portugal, and Singapore" (p. 9). Balcerak and Woźniak (2021) found candidates' reactions to traditional and ICT-based (including gaming-styled tools) selection tools differed, with the perceived fairness of the assessment results being lower for the latter than the former. Perceptions of procedural justice have an impact on the applicants' perceptions of selection methods and influence organizational attractiveness (Folger et al., 2022). These results should be interpreted considering the different social and psychological variables that characterize candidates (Balcerak & Woźniak, 2021).

Study Limitations and Suggestions for Future Studies

We advise that a larger sample be used in future studies when correlating job performance and task-switching ability. Although no sample size issues were detected, the correlation is based on a relatively small sample size and, thus, might lead to results that require further confirmation to ensure more accurate conclusions (e.g., Schönbrodt & Perugini, 2013). As suggested by Schönbrodt and Perugini (2013), in research contexts of convergent validity, expected correlations can be greater. Therefore, other criterion measures of job knowledge, or ratings from peers and direct supervisors combined (instead of relying on third-party ratings alone), should be considered in further studies.

Moreover, due to the technical specificities of the instrument (i.e., the subtests end if participants are unable to control the velocimeter), it becomes difficult to extrapolate this instrument beyond its use in a computerized context. However, we believe that some features of the tasks may be relevant to generate discussion of the importance of these tasks to assess the construct task-switching even in nontechnological contexts involving distractors other than the velocimeter adopted in this instrument.

Finally, in the present study, nationality was not an inclusion criterion if participants were able to communicate in Portuguese. Future studies should expand the sample to include other Portuguese-speaking populations, such as Brazil, Angola, and Mozambique.

Conclusion

The present study is a promising first step toward developing a Portuguese-speaking tool to assess ICT workers' task-switching ability. Statistical analyses showed evidence for the acceptable psychometric properties of the assessment tool.

References

Aagaard, J. (2019). Multitasking as distraction: A conceptual analysis of media multitasking research. *Theory & Psychology*, *29*(1), 87–99. https://doi.org/10.1177/0959354318815766

Aghvinian, M., Santoro, A. F., Gouse, H., Joska, J. A., Linda, T., Thomas, K. G., & Robbins, R. N. (2021). Taking the test: A qualitative analysis of cultural and contextual factors impacting neuropsychological assessment of Xhosa-speaking South Africans. *Archives of Clinical Neuropsychology*, *36*(6), 976–980. https://doi.org/10.1093/arclin/acaa115

Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between The Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, *16*(1), 1–13. https://doi.org/10.1111/j.1468-2389.2008.00404.x

Appelbaum, S. H., Marchionni, A., & Fernandez, A. (2008). The multi-tasking paradox: Perceptions, problems and strategies. *Management Decision*, *46*(9), 1313–1325. https://doi.org/10.1108/00251740810911966

Balcerak, A., & Woźniak, J. (2021). Reactions to some ICT-based personnel selection tools. *Economics and Sociology*, *14*(1), 214–231. https://doi.org/10.14254/2071-789X.2021/14-1/14

Bannister, F., & Remenyi, D. (2009). Multitasking: The uncertain impact of technology on knowledge workers and managers. *Electronic Journal of Information Systems Evaluation*, *12*(1), 1–12. https://academic-publishing.org/index.php/ejise/article/view/291

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, *117*(1), 339–376. https://doi.org/10.1162/003355302753399526

Brock, M. E., & Buckley, M. R. (2013). Human resource functioning in an information society: Practical suggestions and future implications. *Public Personnel Management*, *42*(2), 272–280. https://doi.org/10.1177/0091026013487047

Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality and Quantity*, *26*(3), 233–244. https://doi.org/10.1007/BF00172427

Carrier, L. M., Rosen, L. D., Cheever, N. A., & Lim, A. F. (2015). Causes, effects, and practicalities of everyday multitasking. *Developmental Review*, *35*, 64–78. https://doi.org/10.1016/j.dr.2014.12.005

Cattell, R. B. (1940). *Culture fair intelligence test (CFIT)* [Database record]. APA PsycTests. https://doi.org/10.1037/t14354-000

Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K. P. L. (2013). Improving multi-tasking ability through action videogames. *Applied Ergonomics*, *44*(2), 278–284. https://doi.org/10.1016/j.apergo.2012.08.002

Crews, D. E., & Russ, M. J. (2020). The impact of individual differences on multitasking ability. *International Journal of Productivity and Performance Management*, *69*(6), 1301–1319. https://doi.org/10.1108/IJPPM-04-2019-0191

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dzyubenko, I. B. (2021). *Fast-growing tech companies as a driver of regional and national sustainable economic development*. Proceedings of the International Scientific and Practical Conference on Sustainable Development of Regional Infrastructure (pp. 540–548). https://doi.org/10.5220/0010593605400548

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*(1), 5–18. https://doi.org/10.1007/s11136-007-9198-0

Ferrando, P. J., & Lorenzo-Seva, U. (2017). Program FACTOR at 10: Origins, development and future directions. *Psicothema*, *29*(2), 236–241. https://doi.org/10.7334/psicothema2016.304

Ferreira, A. I., Almeida, L. S., & Prieto, G. (2012). Construction of a memory battery for computerized administration, using item-response theory. *Psychological Reports*, *111*(2), 585–609. https://doi.org/10.2466/03.04.PR0.111.5.585-609

Fischer, R., & Plessow, F. (2015). Efficient multitasking: Parallel versus serial processing of multiple tasks. *Frontiers in Psychology*, *6*, Article 1366. https://doi.org/10.3389/fpsyg.2015.01366

Folger, N., Brosi, P., Stumpf-Wollersheim, J., & Welpe, I. M. (2022). Applicant reactions to digital selection methods: A signaling perspective on innovativeness and procedural justice. *Journal of Business and Psychology*, *37*(4), 735–757. https://doi.org/10.1007/s10869-021-09770-3

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-labAdvanced: An air traffic control simulator with realism and control. *Behavior Research Methods*, *41*(1), 118–127. https://doi.org/10.3758/BRM.41.1.118

Fowler, F. J., Jr, & Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Sage.

Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive psychology*, *44*(3), 297–337. https://doi.org/10.1006/cogp.2001.0770

Groth-Marnat, G. (2003). *Handbook of psychological assessment* (4th ed.). John Wiley & Sons.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167. https://doi.org/10.1002/acp.1624

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Institute for Objective Measures. (n.d.). *What do infit and outfit, mean-square and standardized mean?* https://www.rasch.org/rmt/rmt162f.htm

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–583. https://doi.org/10.1037/bul0000144

Kohl, K., Vasilescu, B., & Prikladnicki, R. (2020, June). *Multitasking across industry projects: A replication study*. Proceedings of the

IEEE/ACM 42nd International Conference on Software Engineering Workshops (pp. 93–100). https://doi.org/10.1145/3387940.3391495

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), Article 33267. https://doi.org/10.1525/collabra.33267

Lance, C. E., & Vandenberg, R. J. (2002). Confirmatory factor analysis. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221–254). Jossey-Bass/Wiley.

Lee, F. J., & Taatgen, N. A. (2002). Multitasking as skill acquisition. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (2nd ed., pp. 572–577). Routledge.

Liefooghe, B., Demanet, J., & Vandierendonck, A. (2010). Persisting activation in voluntary task switching: It all depends on the instructions. *Psychonomic Bulletin & Review*, *17*(3), 381–386. https://doi.org/10.3758/PBR.17.3.381

Linacre, J. M. (2005). *A user's guide to WINSTEPS*. Mesa Press.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2), Article 878. https://www.rasch.org/rmt/rmt162f.htm

Linden, W. J., van der Linden, W. J., & Glas, C. A. (Eds.), (2000) *Computerized adaptive testing: Theory and practice*. Springer Science & Business Media.

Lockwood, N. R. (2006). Talent management: Driver for organizational success. *HR Magazine*, *51*(6), S1–S11. https://www.proquest.com/docview/205016356

Lorenzo-Seva, U., & Ferrando, P. J. (2023). A simulation-based scaled test statistic for assessing model-data fit in least-squares unrestricted factor-analysis solutions. *Methodology*, *19*(2), 96–115. https://doi.org/10.5964/meth.9839

Lorenzo-Seva, U., & Ferrando, P. J. (2019). Robust promin: un método para la rotación de factores de diagonal ponderada [Robust promin: A method for diagonally weighted factor rotation]. *LIBERABIT, Revista Peruana de Psicologia*, *25*(1), 99–106. https://doi.org/10.24265/liberabit.2019.v25n1.08

Lorenzo-Seva, U., & Ferrando, P. J. (2021). MSA: The forgotten index for identifying inappropriate items before computing exploratory item factor analysis. *Methodology*, *17*(4), 296–306. https://doi.org/10.5964/meth.7185

Lorenzo-Seva, U., & Van Ginkel, J. R. (2016). Multiple imputation of missing values in exploratory factor analysis of multidimensional scales: Estimating latent trait scores. *Anales de Psicologia/Annals of Psychology*, *32*(2), 596–608. https://doi.org/10.6018/analesps.32.2.215161

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. Guilford Press.

Moore, K. A., Halle, T. G., Vandivere, S., & Mariner, C. L. (2002). Scaling back survey scales. How short is too short? *Sociological Methods & Research*, *30*(4), 530–567. https://doi.org/10.1177/0049124102030004003

Morgan, B., D'Mello, S., Abbott, R., Radvansky, G., Haass, M., & Tamplin, A. (2013). Individual differences in multitasking ability and adaptability. *Human Factors*, *55*(4), 776–788. https://doi.org/10.1177/0018720812470842

Naqvi, S. M. M. R., & Bashir, S. (2015). IT-expert retention through organizational commitment: A study of public sector information technology professionals in Pakistan. *Applied Computing and Informatics*, *11*(1), 60–75. https://doi.org/10.1016/j.aci.2011.11.001

Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory – Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, *106*(37), 15583–15587. https://doi.org/10.1073/pnas.0903620106

Oswald, F. L., Hambrick, D. Z., & Jones, L. A. (2017). Keeping all the plates spinning: Understanding and predicting multitasking performance. In D. H. Jonassen (Ed.), *Learning to solve complex scientific problems* (pp. 77–96). Routledge. https://doi.org/10.4324/9781315091938

Pashler, H. (2000). Task switching and multitask performance. In S. Monsell & J. Driver (Eds.), *Attention and performance, XVIII: Control of mental processes* (pp. 277–307). MIT Press.

Prieto, G., & Delgado, A. R. (2007). Measuring math anxiety (in Spanish) with the Rasch rating scale model. *Journal of Applied Measurement*, *8*(2), 149–160.

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, D. Z., & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473–1492. https://doi.org/10.1037/xge0000219.supp

Reynolds, C. R. (1998). 10.03—Cultural bias in testing of intelligence and personality. In A. S. Bellack & M. Hersen (Eds.), *Comprehensive clinical psychology* (pp. 53–92). Elsevier Science. https://doi.org/10.1016/B0080-4270(73)00105-X

Reynolds, C. R., & Suzuki, L. A. (2012). Bias in psychological assessment: An empirical review and recommendations. In I. B. Weiner (Ed.), *Handbook of psychology* (2nd ed., pp. 82–113). https://doi.org/10.1002/9781118133880.hop210004

Rock, D., & Price, I. R. (2019). Identifying culturally acceptable cognitive tests for use in remote northern Australia. *BMC Psychology*, *7*, Article 62. https://doi.org/10.1186/s40359-019-0335-7

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Russ, M., & Crews, D. E. (2014). A survey of multitasking behaviors in organizations. *International Journal of Human Resource Studies*, 4(1), 137–153. https://doi.org/10.5296/ijhrs.v4i1.5155

Schneider, D. W., & Logan, G. D. (2009). Task switching. *Encyclopedia of Neuroscience*, *9*, 869–874. https://doi.org/10.1016/B978-008045046-9.00426-5

Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). The Guilford Press.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measure perspective. *Journal of Applied Measurement*, *2*(3), 281–311.

Srivastava, E., & Agarwal, N. (2012). The emerging challenges in HRM. *International Journal of Scientific & Technology Research*, *1*(6), 46–48.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–98. https://doi.org/10.1016/j.paid.2006.09.017

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Strobach, T., Wendt, M., & Janczyk, M. (2018). Multitasking: Executive functioning in dual-task and task switching situations. *Frontiers in Psychology*, *9*, Article 108. https://doi.org/10.3389/fpsyg.2018.00108

Suija-Markova, I., Briede, L., Gaile-Sarkane, E., & Ozoliņa-Ozola, I. (2020). Multitasking in knowledge intensive business services. *Emerging Science Journal*, *4*(4), 305–318. https://doi.org/10.28991/esj-2020-01233

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Vasilescu, B., Blincoe, K., Xuan, Q., Casalnuovo, C., Damian, D., Devanbu, P., & Filkov, V. (2016, May). The sky is not the limit: Multitasking across GitHub projects. In *ICSE '16: Proceedings of the 38th International Conference on Software Engineering* (pp. 994–1005). ACM Digital Library. https://doi.org/10.1145/2884781.2884875

Vernon, P. E. (1964). *The structure of human abilities*. Methuen.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Erlbaum.

Wright, B. D., & Linacre, J. M. (1998). *Winsteps: A Rasch computer program*. Mesa Press.

Wright, J., & Atkinson, D. (2019). *The impact of artificial intelligence within the recruitment industry: Defining a new way of recruiting* (pp. 1–39). Carmichael Fisher.

Yang, F. M., & Kao, S. T. (2014). Item response theory for measurement validity. *Shanghai Archives of Psychiatry*, *26*(3), 171–177. https://doi.org/10.3969/j.issn.1002-0829.2014.03.010

Ziegler, M. (2020). Psychological test adaptation and development – How papers are structured and why [Editorial]. *Psychological Test Adaptation and Development*, *1*, 3–11. https://doi.org/10.1027/2698-1866/a000002

Ziegler, M. (2014). Stop and state your intentions! Let's not forget the ABC of test construction. *European Journal of Psychological Assessment*, *30*(4), 239–242. https://doi.org/10.1027/1015-5759/a000228

History
Received December 31, 2022
Revision received September 19, 2023
Accepted January 26, 2024
Published online March 6, 2024

Section: Methodological Topics in Assessment

Conflict of Interest
The authors declare no conflict of interest.

Open Data
All measures, conditions, data exclusions, and missing data have been reported.

Preregistration
The study was registered at the Center for Open Science (https://doi.org/10.17605/OSF.IO/N8YSM, preregistration: https://osf.io/2xjf6, databases: https://osf.io/uyw64/).

Funding
This work was supported by Fundação para a Ciência e Tecnologia (Grant UIDB/00315/2020) and the P2020 Digital Talent Ecosystem Project (DTE; LISBOA-01-0247-FEDER-045216).

ORCID
Beatriz Regina de Oliveira Trigo
https://orcid.org/0000-0002-0337-5428
Aristides I. Ferreira
https://orcid.org/0000-0002-8280-1623

Beatriz Regina de Oliveira Trigo
Product Quality & Research
Enhanced Fertility
EFP Clinic
Lisbon
Portugal
beatriz@efp.clinic