

**Construct validity, sensitivity and specificity of the USCD Performance-based Skill
Assessment 2 in a mixed Portuguese sample**

Carolina da Motta^{a)}, Célia Barreto Carvalho^{a), b)}, Paula Castilho^{a)}, Michele T. Pato^{c), d)}

a) Center for Research in Neuropsychology and Cognitive and Behavioral Intervention
(CINEICC), College of Psychology and Education Sciences, University of Coimbra,
Portugal

b) Division of Psychology, Department of Educational Sciences, University of Azores,
Azores, Portugal

c) Director & Professor Institute for Genomic Health, SUNY Downstate Medical Center,
Brooklyn, NY

d) Keck School of Medicine, Department of Psychiatry and Behavioral Sciences, University
of Southern California

Correspondence concerning this article should be addressed to: Carolina da Motta, CINEICC,
Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra, Rua do
Colégio Novo, Apartado 6153, 3001-802, Coimbra, Portugal. E-mail:
carolina.d.motta@fpce.uc.pt

Abstract

Background: Assessment batteries of functional capacity provide robust indicators of real-world functioning in major psychiatric illnesses and important information on individual's ability to live autonomously and pursue relevant psychosocial goals.

Objectives: The current study explores the psychometric properties of the Portuguese USCD Performance-based Skill Assessment 2 (UPSA-2-PT) in a mixed sample of

Portuguese participants. **Method:** A sample of 110 participants, consisting of 37 patients diagnosed with schizophrenia, 27 first-degree relatives of patients and 46 controls were administered the UPSA-2-PT and self-report questionnaires. The UPSA-2-PT reliability was assessed through Inter-rater Reliability and internal consistency, convergent validity with community integration and a Receiver Operating Curve analysis was conducted to establish sensitivity and specificity of scores, and Youden's index was used to determine an optimal cutoff value of functional capacity. **Results:** Findings show an excellent inter-rater reliability, good internal consistency and construct validity, consistent with previous studies in several Western countries. The UPSA-2-PT also showed good discriminant ability between patients and controls, and an overall percentage of correct classification of 86.7% for the 81.75 cutoff point. **Discussion:** Findings are congruous with previous versions, strengthening the body of evidence supporting the construct validity and providing a useful tool for research and clinical purposes to practitioners of several fields.

Keywords: Functional assessment, performance-based, schizophrenia, UPSA-2, validation

Construct validity, sensitivity and specificity of the USCD Performance-based Skill Assessment 2 in a mixed Portuguese sample

Functional impairment is responsible low schooling or employment verified in people diagnosed with chronic psychiatric illnesses, in addition to the increased difficulties in autonomy and performing everyday tasks. Special attention has been given to the impairment of patients suffering from severe mental illnesses, such as schizophrenia (Bowie et al., 2008; Bowie & Harvey, 2006), as the burden of the disease affects the capacity to autonomous living and psychosocial integration in the community. Moreover, functional capacity has been referred as an important output of interventions to promote recovery through cognitive changes and rehabilitation (Keefe et al., 2013; Nuechterlein et al., 2008).

Performance-based assessment can provide information from manifest behavior, thus avoiding subjective biases and social desirability, because it is based in specific scoring of manifest behaviors and is less dependent of the participants' metacognitive processes that can influence evaluation of one's own behavior (Harvey, Velligan, & Bellack, 2007; Patterson, Goldman, McKibbin, Hughs, & Jeste, 2001). This methodology allows the objective definition of the set of steps or routines necessary for each task completion, also facilitating the scoring procedure for the test administrator. Currently, one of the preferred approaches to the assessment of patient's functionality are performance-based task, and test batteries that can simulate more closely several kinds daily tasks are thought to be a more valid assessment methodology to predict real-world functioning (Bowie et al., 2008; Harvey et al., 2007). This approach provides significant advantages, as increased ecological validity and objectivity, and the flexibility of adaptation to different cultural contexts, while maintaining task and scoring integrity. For this reason, performance-based assessment has been an important approach to the evaluation of the

functional capacity in patients suffering with all kinds of impairing diseases, like Dementia, Mild Cognitive Impairment, Bipolar Disorder and Schizophrenia and Schizoaffective disorders (Becattini-Oliveira, Dutra, Spenciere de Oliveira Campos, de Araujo, & Charchat-Fichman, 2018; Garcia-Portilla et al., 2013; Mausbach et al., 2016).

Among the existing measures, the USCD Performance-based Skill Assessment (UPSA) has been developed specifically for schizophrenia and translated and adapted to several languages and cultural backgrounds and is a cost-effective instrument that is easily administered and made accessible to many health professionals worldwide (Becattini-Oliveira et al., 2018; Patterson et al., 2001). This assessment is carried out in 5 different domains that reflect common areas of functioning in everyday life of western and eastern societies (organization/planning, finances, communication, transportation and household tasks), with tasks that can easily be adapted to the local specificities, such as local currency, local transportation systems and/or organization and planning of leisure activities, while maintaining the execution routines, task difficulty and scoring procedures.

The USCD Performance-based Skill Assessment (UPSA) was first developed by Patterson and colleagues (2001) and has different versions (UPSA 1, UPSA 2, and brief versions) available for paper-and-pencil, computer or mobile administration. In a recent systematic review (Becattini-Oliveira et al., 2018), the UPSA tests have been found to be translated to 8 languages and is administered in 17 countries, while maintaining the original properties after cultural adaptation. Overall findings indicate its good psychometric properties (internal consistency, concurrent validity with cognitive ability and other functional measures, sensibility and specificity) and utility discriminating patients from controls across several culturally-adapted versions of the UPSA. However, few studies have analyzed the UPSA's specificity and

sensitivity, and cutoff points established across different countries tend to vary and may not be generalized (Becattini-Oliveira et al., 2018).

Despite the UPSA-1 being validated and made available in Brazilian Portuguese (Mantovani, Machado-de-Sousa, & Salgado, 2015), to the best of our knowledge, the UPSA-2 was never translated and adapted to Portuguese. In addition, more important than the language nuances between Portuguese native-speakers, it is important to emphasize the need to adapt the tasks as much as possible to the cultural specificities and daily routines, which can differ considerably between European and American countries. This was verified in the English and Spanish versions of the UPSA, when administered to Hispanic populations (Garcia-Portilla et al., 2013; Mausbach et al., 2016). Providing equivalent but ecologically valid measures is also of utmost importance considering the increased mobility of European, American and African Portuguese-native speakers worldwide.

Previous studies have shown the importance of using the translated and adapted tasks when assessing participants from different cultural backgrounds or in their native language (Garcia-Portilla et al., 2013; Mantovani et al., 2015; Mausbach et al., 2016), and studying their psychometric properties in different populations is an important step in assuring scores comparability and test integrity across versions. Most studies on the validity of the UPSA did not include inter-rater reliability, an important aspect that should be excluded from the factors that may influence variability in scores across different versions of the UPSA and few have analyzed the sensitivity and specificity of scores and suggested possible cutoff points to functional capacity (Becattini-Oliveira et al., 2018). Thus, the current study aimed to adapt the UPSA-2 to Portuguese (UPSA-2-PT) and to the European context, and to study the validity of this measure

in Portuguese samples. Additional goals included exploring the UPSA-2-PT sensitivity and specificity, establishing the discriminant power and establishing a cutoff point for the measure.

Method

Participants and Procedures

The current study is a part of a broader research project registered in the National Data Protection Committee and approved by the Ethical Boards of the University of Coimbra and Azores University, and data collection was also evaluated and approved by the boards or directorship from 4 mental health institutions from the Azores and Portuguese mainland. Participants were patients diagnosed with schizophrenia, first-degree relatives of a patient affected with schizophrenia and controls (participants with no prior history of psychosis nor a first-degree relative of a person diagnosed with a psychotic disorder). Inclusion criteria for the patients' sample included being diagnosed with schizophrenia over a period longer than 6 months, confirmed by the psychiatrist or consultation of medical records. Patients and first-degree relatives were recruited at local hospitals, mental health institutions or were referred by private practitioners. Participants from the healthy control sample were recruited by snowball convenience sampling method, in which participants would refer other potential participants (Biernacki & Waldorf, 1981). Participation in the current study was voluntary and all participants were informed on the study goals, anonymity and confidentiality, and have signed an informed consent form prior to participation.

The adaption of the UPSA-2 administration manual was based in the UPSA-2 (English and Spanish versions) and also compared with the UPSA-1-BR to maintain the structure and format as close as possible across all versions, and proximity of vocabulary between the

European and Brazilian Portuguese. Adaptions to the tasks included changing financial tasks to the Euro currency, creating an alternative ATM payment method (*Multibanco*) with an equal number of steps to replace the check-writing task, adjusting transportation tasks materials and questions to local transportation lines, and household products pictures were changed to more closely match to locally available products.

Statistical Analyses

Descriptive and inferential statistics were calculated using SPSS v. 22 (IBM Corp. Released, 2013). Inter-rater Reliability (IRR) was calculated through Cohen's kappa coefficient from scores attributed by two independent raters. Missing responses to items in self-report questionnaires have not surpassed 5% of cases and conformed to a MCAR pattern in all variables, thus, missing values were estimated and replaced by regression method. Pearson product-moment correlation coefficients were calculated to examine the relationship between the UPSA-2-PT and Community Integration scores. Chi-square tests and One-way Analyses of Variance (ANOVA with Welch's robust test) were used to characterize the samples partaking in this study and to determine observed differences between groups. *Post-hoc* tests were used to further compare groups, using the Games-Howell estimation whenever the homogeneity of variance could not be assumed. A Receiver Operating Curves (ROC) was plotted to assess the discriminative power of the UPSA-2-PT scores, and the Area Under the Curve (AUC) with a 95% Confidence Interval was used as an index of the ability of the UPSA scores to differentiate participants who were healthy controls and patients diagnosed with schizophrenia. An examination of the combination sensitivity and specificity estimation has allowed the definition of an optimal *cutoff* score with Youden Index (*J*) (Youden, 1950) and the percentage of

participants correctly classified as having/not having the condition were provided to clarify the cut-off value's potential utility.

Measures

UPSA = USCD Performance-based Skills Assessment (Patterson et al., 2001) evaluates performance in 5 domains in everyday functioning and community living (household chores, comprehension/planning a recreational activity, communication, finance, transportation) through a series of standardized skills performance situations. The measure was developed for patients with schizophrenia, and can be administered by clinicians, occupational therapists, nurses and social workers. The UPSA tasks consist of a series of role-play situations of similar complexities and adequate to most participants who live in western communities (e.g. phone-calling, counting change, making a shopping list, using public transportation, planning an outing). Results are computed from raw-scores to a scale from 0 to 20, where functioning in different domains can be compared, and produces a global score ranging from 0 to 100. The test is completed in 20 to 30 minutes, and a briefer version (UPSA-brief) consisting of the financial and communication skills can be used in cases of severe impairments (Mausbach, Harvey, Goldman, Jeste, & Patterson, 2007). Both the UPSD and the UPSA-brief have good psychometric properties, sensitivity to change and discriminative ability between patients and healthy controls.

Community Integration Scale for Adults with Psychiatric Problems – Brief Version (CIS-APP; Cabral & Barreto Carvalho 2013). The CIS-APP is a self-report measure devised for adults with psychiatric problems, but also applicable for people without mental illnesses. The 12-item version was obtained through CFA based on the 34 item version of the CIS-APP (Cabral, Barreto

Carvalho, da Motta, & Silva, 2014). Instructions include a brief definition of community and items are rated in a 5-point Likert-type a scale ranging from 1 (“Completely disagree”) to 5 (“Completely agree”). Scores can range from 0 to 60 points; thus, higher scores indicate increased degrees of community integration. Internal consistency in the current sample was .87.

Results

Sample characteristics

Participants from the three samples differed in terms of gender distribution, in the expected sense: due to the higher schizophrenia male/female ratio, most participants from the clinical sample (SZ) were males and most participants from the patient’s first-degree relatives (PR) were females. The gender of participants from the healthy controls (HC) sample was distributed evenly. Most participants from the clinical sample were single, receiving disability pension, and belonged to medium-high household income. Ten patients (27%) were living full-time in an institution, and participants from the clinical sample had between an average duration of diagnosed illness of 19.28 years ($SD = 13.11$). Participants from the non-clinical samples tended to have a more favorable occupational status and household income. Participants from the three groups presented statistically significant differences regarding age and years of education ($F_{age}=4.854, p = .010$; and $F_{education}=17.454, p < .001$). *Post hoc* tests revealed that relatives of patients (PR) were significantly older than participants from the healthy controls and patients’ samples, which in turn were equivalent in terms of age. Regarding years of education, participants from the healthy control groups had significantly more years of education (e.g. college) in comparison with the patients and relatives’ samples, who had on average years close

to mandatory education (currently 12 years, in Portugal). More detailed sample characteristics are presented in table 1.

Table 1.

Sample characteristics (N = 110)

	Non-Clinical Samples		Clinical Sample		
	Controls (n = 46)	Patients' Relatives (n = 27)	Patients diagnosed with Schizophrenia (n = 37)		
	N (%)	N (%)	N (%)	X²	p
Gender				25.261	.000
Male	23 (50%)	4 (14.8%)	29 (78.4%)		
Female	23 (50%)	23 (85.2%)	8 (21.6%)		
Marital Status				29.063	.000
Single	24(52.2%)	5 (19.2%)	28 (82.4%)		
Married	17 (46%)	18 (9.6%)	4 (11.8%)		
Divorced	2 (4.3%)	0	1 (2.9%)		
Widowed	0	1 (3.8%)	0		
Civil Union	3 (6.5%)	2 (7.7%)	1 (2.9%)		
Employment Status				72.561	.000
Employed	35(76.1%)	21(80.7%)	8(21.6%)		
Protected employment	0	0	5 (13.5%)		
Voluntary work	0	0	1 (2.7%)		
Student	7(15.2%)	0	2(5.4%)		
Unemployed	2 (4.3%)	1 (3.8%)	3 (8.1%)		
Not seeking for job/not employed	0	2 (7.7%)	1 (2.7%)		
Retirement	2(4.3%)	2(7.7%)	0		
Disability Pension	0	0	17 (45.9%)		
Monthly household income				15.900	.014
Under 500€	1 (2.2%)	1 (4.2%)	8 (28.6%)		
500-1300€	18 (40%)	11 (45.8%)	9 (32.1%)		
1300-2500€	17 (37.8%)	7 (29.2%)	9 (32.1%)		
2500€ or more	9 (20%)	5 (20.8%)	2 (7.1%)		
	M (SD)	M (SD)	M (SD)	F	p
Age (in years)	39.70 (10.80)	48.04 (11.11)	40.92 (12.40)	4.854	.010
Min.-Max.	23-64	25-63	19-68		

Years of education	15.49 (2.74)	11.08 (4.68)	11.35 (3.74)	17.454	.000
Min.-Max.	10-20	4-19	3-19		

Inter-Rater Reliability

The UPSA-2-PT was administered to 30 participants in the presence of a second independent rater to assess inter-rater agreement. Inter-rater Reliability (IRR) of categorical responses coded by the two raters of each UPSA item was calculated through Cohen's kappa coefficient. The agreement coefficient was very high (Cohen's $k = .966$; $t = 23.324$; $p < .001$), where only 3 scores out of 583 observations have diverged between raters. Due to the small number of disagreements, the authors have inspected the reasons underlying the different scores attributed by each rater. Two of these coding disagreements have been due to the misapplication of the scoring rule for the shopping list task (Household chores skills) and one phone number being mistakenly coded as correct in the communication skill task.

Task equivalence of financial task

To ensure the ecological validity of the Portuguese version and a direct comparability of the financial task requiring participants to pay an electric bill through ATM and check writing methods, the participants were asked to perform both tasks. The association of performance scores in the 6 steps of both tasks was only moderate ($r = .512$; $p < .001$) and participants scored significantly worse, on average, in the check writing task ($M = 5.16$; $SD = 1.37$), in comparison to the ATM method ($M = 5.58$; $SD = 1.02$; $t_{(116)} = -3.715$; $p < .001$). When asked about what their preferred method would be, all the participants answered the ATM method was their preferred method of payment, to which they were more familiar. For this reason, the UPSA-2-PT should

solely include the ATM test scores, and results presented in the following sections will always refer to calculations using the ATM task instead of check-writing.

Internal consistency

Internal consistency values for the complete UPSA-2-PT was good, $\alpha = .860$, and adequate values were found in each subscale (Planning $\alpha = .664$; Finance $\alpha = .711$; Transportation $\alpha = .823$; Household chore $\alpha = .745$), except for the Communication skill module that presented the lowest internal consistency value: $\alpha = .566$.

Convergent and discriminant validity

The correlation between UPSA-2-PT scores and Community Integration Scale (CIS) was statistically significant for the total scale ($r = .305, p = .001$). Out of 5 domains, the household chores, financial and communication skills presented a weak but statistically significant correlation with Community Integration (ranging from $r = .201$ to $r = .299; p < .050$) and the tasks pertaining to the use of public transportation and planning skills did not present a statistically significant correlation with CIS.

As presented in table 2, participants from the three samples differed in the overall performance of the UPSA-2-PT tasks and across all domains. The HC group performed significantly better than the SZ group in all 5 domains and total scores. However, the HC group scored significantly higher than the PR group only in the transportation and total score, while the PR performance did not statistically differ from SZ group in activity planning, finance, transportation and household chores skills scores.

Table 2.

Group comparisons of UPSA-2- PT domains and total scores (N = 110)

	Non-Clinical Samples				Clinical Sample				
	Controls (n = 46)		Patient's Relatives (n = 27)		Patients diagnosed with Schizophrenia (n = 37)				
	Mean	SD	Mean	SD	Mean	SD	<i>F_{Welch}</i>	<i>p</i>	Post-Hoc
Planning	19.29	1.08	17.62	1.68	16.53	3.27	20.029	.000	HC>PR≥SZ
Finances	18.66	1.55	17.64	3.66	3.79	.62	10.016	.000	HC≥PR≥SZ
Communication	13.19	2.89	13.33	2.73	4.29	.70	8.623	.000	HC,PR>SZ
Transportation	19.28	1.56	14.73	5.55	5.24	.86	19.990	.000	HC>PR≥SZ
Domestic	18.59	2.28	17.59	3.21	7.57	1.24	8.900	.000	HC.PR>SZ
Total	88.99	4.24	80.92	9.32	70.27	16.88	28.422	.000	HC>PR>SZ
	n	%	n	%	n	%			
Floor effect (<5%)	-	-	1	.9	4	3.6			
Ceiling effect (>95%)	4	3.6	-		1	.9			

Because the base rate of participants from the clinical sample living independently or employed was too low and age was equivalent between the HC and SZ samples, ROC curve analysis examined the UPSA-2-PT sensitivity and specificity for identifying patients diagnosed with schizophrenia versus non-affected participants. Thus, the analysis was carried out in the patients and healthy control samples, totalizing 83 participants. ROC analysis predicting patient's status shows a good discriminant power, with an AUC = .885; $p < .001$ (95% CI: .81-.97, Figure 1).

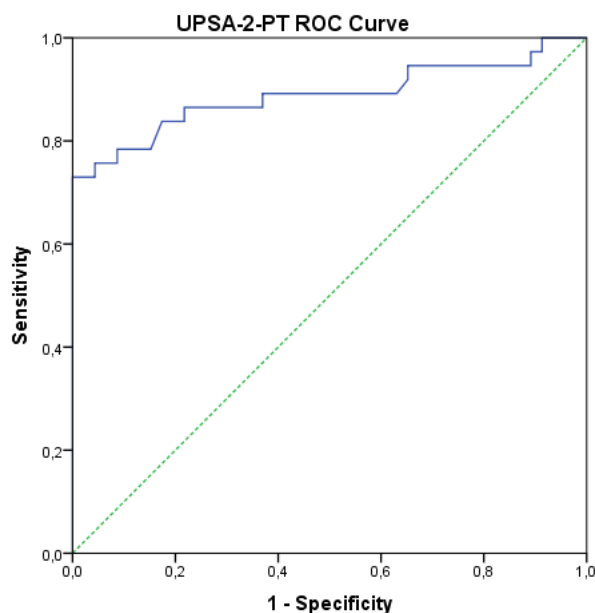


Figure 1 – ROC analysis predicting patient's status; AUC = .888 (95% CI: .81-.97)

Table 3.

Partial output of Sensitivity, Specificity & Youden's Index cutoff scores (Youden's $J > .45$)

Positive if Less Than or Equal To	Sensitivity	Specificity	Youden J
68.87	.459	1.000	.459
70.40	.486	1.000	.486
71.92	.514	1.000	.514
71.93	.541	1.000	.541
73.47	.568	1.000	.568
76.33	.595	1.000	.595
78.47	.622	1.000	.622
79.85	.649	1.000	.649
80.44	.676	1.000	.676
80.64	.703	1.000	.703
80.93	.703	.978	.681
81.20	.730	.978	.708
81.48	.730	.957	.686
81.75	.757	.957	.713
82.32	.757	.935	.692

82.95	.757	.913	.670
83.46	.757	.891	.648
83.94	.784	.891	.675
84.41	.784	.870	.653
84.74	.784	.848	.632
84.94	.784	.826	.610
85.28	.838	.804	.642
85.60	.838	.783	.620
85.76	.838	.761	.599
86.09	.865	.761	.626
86.44	.865	.717	.582
86.56	.865	.674	.539
86.91	.865	.652	.517
87.34	.865	.630	.495
87.54	.892	.630	.522
87.82	.892	.609	.501
88.02	.892	.587	.479
388.11	.892	.565	.457

The threshold between 71.91 and 86.91 points revealed the combinations of best sensitivity and specificity values. The optimal tradeoff between sensitivity and specificity was reached at the cutoff score of 81.75, as indicated by the highest Youden index value (table 3). Of all individuals that scored below that cutoff value, only 2 (4.3%) were from the healthy controls sample and 28 (75.7% - positive predictive value) were patients, while 44 participants scoring above 81.75 were healthy controls (95.7% - negative predictive value) and 9 were patients diagnosed with schizophrenia (24.3%). The overall percentage of correct classification is 86.7%.

Discussion

The current study sought to examine the psychometric properties of the Portuguese version of the UPSA-2 (UPSA-2-PT), adapted to the European context, in clinical and non-

clinical samples, and to establish a cutoff point with enough sensitivity and specificity to distinguish the functional capacity of participants.

Due to the pertinence of the financial tasks, specifically the payment of an electric bill, it was necessary to adapt the task so it could match the more disseminated electronic payment practices, as the use of check has markedly declined over the past years. Because maintaining the original task might artificially raise the task difficulty and lose ecological validity, as demonstrated by the significantly higher performance in task when the payment method corresponded to the participant's preferred method, the equivalent ATM payment method (consisting of 6 steps of similar difficulty) should be preferred when administering the UPSA-2-PT.

Inter-rater Reliability analysis showed an excellent degree of agreement for the categorical ratings in UPSA-2-PT. Internal consistency was good for all scales and subscales, except for the communication module. This may be because the computations were carried out for the complete sample, and communication skills scores might be less consistent due to the greater impairment some patients diagnosed with schizophrenia suffer in this domain.

Consistent with previous findings (Garcia-Portilla et al., 2013; Mausbach et al., 2007; Patterson et al., 2001; Vella et al., 2017), the UPSA-2-PT results showed patients have a significant in functional capacity comparing to controls or their first-degree relatives, who were generally older participants.

In a relatively functional sample of participants, the UPSA-2-PT score was able to better discriminate participants from the healthy controls and an heterogeneous sample of patients at a cutoff score of 81.75, with an overall 86.7% percentage of correct classifications. The presence of participants institutionalized for a variable time period may be regarded a limitation to the

current study, as long-term institutionalizations can also preclude participants from performing certain tasks (e.g. paying bills, stricter management of money, appointments or external activities). However, not only some of the participants have been committed to short-term institutionalization (6 months or less), but also community-dwelling patients often live with their families and receive help or delegate the same tasks to their caretakers.

The establishment of cutoff scores provides several advantages and may be an effective way to estimate the ability of patients to reside or secure a professional activity more autonomously. The functional capacity scores provided by UPSA-2-PT presented a positive, but weak, association with subjective assessment of Community Integration. Therefore, a cautionary note should be added: findings suggest an adequate performance in skills assessment is a necessary but not sufficient condition for independent living to occur. Other individual and societal factors (e.g. stigma, job opportunities, motivation, symptoms and comorbidity) may contribute to foster or to hinder patient's integration in the community, and access to education or employment, consequently thwarting or facilitating the degree of success in these pursuits.

An important limitation of the current study is the absence of test-retest reliability testing. Moreover, while the convenience sampling method used in the current study requires caution in the generalization of findings. Nevertheless, it is important to highlight that the general psychometric characteristics and functional milestones in the current study is similar to those found in several UPSA studies in Western countries, particularly to those using samples of in- and outpatients with schizophrenia and controls in studies (Garcia-Portilla et al., 2013; Keefe et al., 2016; Mantovani et al., 2015; Mausbach et al., 2016; Patterson et al., 2001). Overall findings suggest a good integrity and robustness of the UPSA set measures in general, and the UPSA-2-PT. Future studies should aim to explore the evolution of UPSA-2-PT scores longitudinally, not

only to assess test-retest reliability, but also to address the UPSA-2-PT's potential to detect functional changes resulting from interventions in clinical samples. Additionally, studies should further explore the relationships between functional capacity across different domains and neurocognitive impairments that have a significant impact task performance (Ahmed et al., 2016; Ventura, Helleman, Thames, Koellner, & Nuechterlein, 2009)

Overall, the use of the UPSA-2-PT is of increased utility to practitioners aiming to screen participants for rehabilitation programs, to assess the effectiveness of recovery-based interventions and increasing the likelihood of success of interventions aiming to foster the integration of patients with severe psychiatric illnesses.

Acknowledgments

This research was supported by the first author's Ph.D. Grant (SFRH/BD/110308/2015), sponsored by FCT (Foundation for Science and Technology), Portugal, and co-sponsored by ESF (European Social Fund), Brussels, through Portuguese POPH (Human Potential Operational Program). The authors would like to thank the institutional review boards, directors and staff from the *Associação para a Recuperação de Cidadãos Inadaptados da Lousã* (ARCIL), *Associação Quinta das Pontes* and *Unidade de Cuidados Avançados de Esquizofrenia Resistente ao Tratamento* (UCAERe-T) for their collaboration in this study.

References

- Ahmed, A. O., Murphy, C. F., Latoussakis, V., McGovern, K. E., English, J., Bloch, A., ... Savitz, A. J. (2016). An Examination of Neurocognition and Symptoms as Predictors of Post-Hospital Community Tenure in Treatment Resistant Schizophrenia. *Psychiatry Research*. <https://doi.org/10.1016/j.psychres.2016.01.001>
- Becattini-Oliveira, A. C., Dutra, D. de F., Spenciere de Oliveira Campos, B., de Araujo, V. C., &

- Charchat-Fichman, H. (2018). A systematic review of a functional assessment Tool: UCSD Performance-based skill assessment (UPSA). *Psychiatry Research*, 267(January), 12–18.
<https://doi.org/10.1016/j.psychres.2018.05.005>
- Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, 10(2), 141–163.
<https://doi.org/10.1177/004912418101000205>
- Bowie, C. R., & Harvey, P. D. (2006). Cognitive deficits and functional outcome in schizophrenia. *Neuropsychiatric Disease and Treatment*, 2(4), 531–6. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2671937&tool=pmcentrez&rendertype=abstract>
- Bowie, C. R., Leung, W. W., Reichenberg, A., McClure, M. M., Patterson, T. L., Heaton, R. K., & Harvey, P. D. (2008). Predicting schizophrenia patients' real-world behavior with specific neuropsychological and functional capacity measures. *Biological Psychiatry*, 63(5), 505–11.
<https://doi.org/10.1016/j.biopsych.2007.05.022>
- Cabral, J. M., Barreto Carvalho, C., da Motta, C., & Silva, O. D. L. da. (2014). Development and Psychometric Properties of the Community Integration Scale of Adults with Psychiatric Disorders. *European Scientific Journal*, 10(29), 194–208. Retrieved from <http://ejournal.org/index.php/esj/article/view/4426>
- Garcia-Portilla, M. P., Gomar, J. J., Bobes-Bascaran, M. T., Menendez-Miranda, I., Saiz, P. A., Muñoz, J., ... Goldberg, T. E. (2013). Validation of a European Spanish-version of the University of California performance Skills Assessment (Sp-UPSA) in patients with schizophrenia and bipolar disorder. *Schizophrenia Research*, 150(2–3), 421–426.
<https://doi.org/10.1016/j.schres.2013.07.049>

- Harvey, P. D., Velligan, D. I., & Bellack, A. S. (2007). Performance-based measures of functional skills: usefulness in clinical treatment studies. *Schizophrenia Bulletin*, 33(5), 1138–48.
<https://doi.org/10.1093/schbul/sbm040>
- IBM Corp. Released. (2013). IBM SPSS Statistics for Windows, Version 22.0. 2013.
- Keefe, R. S. E., Buchanan, R. W., Marder, S. R., Schooler, N. R., Dugar, A., Zivkov, M., & Stewart, M. (2013). Clinical Trials of Potential Cognitive-Enhancing Drugs in Schizophrenia: What Have We Learned So Far? *Schizophrenia Bulletin*, 39(2), 417–435.
<https://doi.org/10.1093/schbul/sbr153>
- Keefe, R. S. E., Davis, V. G., Atkins, A. S., Vaughan, A., Patterson, T., Narasimhan, M., & Harvey, P. D. (2016). Validation of a Computerized test of Functional Capacity. *Schizophrenia Research*. <https://doi.org/10.1016/j.schres.2016.03.038>
- Mantovani, L. M., Machado-de-Sousa, J. P., & Salgado, J. V. (2015). UCSD Performance-Based Skills Assessment (UPSA): validation of a Brazilian version in patients with schizophrenia. *Schizophrenia Research: Cognition*, 2(1), 20–25. <https://doi.org/10.1016/j.scog.2014.12.002>
- Mausbach, B. T., Harvey, P. D., Goldman, S. R., Jeste, D. V., & Patterson, T. L. (2007). Development of a brief scale of everyday functioning in persons with serious mental illness. *Schizophrenia Bulletin*, 33(6), 1364–72. <https://doi.org/10.1093/schbul/sbm014>
- Mausbach, B. T., Tiznado, D., Cardenas, V., Jeste, D. V., Patterson, T. L., Acion, L., ... Aggarwal, J. (2016). Validation of the UCSD Performance-based Skills Assessment (UPSA) in Hispanics with and without schizophrenia. *Psychiatry Research*, 244, 388–393.
<https://doi.org/10.1016/j.psychres.2016.08.027>
- Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., ... Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, Part 1: Test Selection,

Reliability, and Validity. *American Journal of Psychiatry*, 165(2), 203–213.

<https://doi.org/10.1176/appi.ajp.2007.07010042>

Patterson, T. L., Goldman, S., McKibbin, C. L., Hughs, T., & Jeste, D. V. (2001). UCSD

Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin*, 27(2), 235–245.

<https://doi.org/10.1093/oxfordjournals.schbul.a006870>

Vella, L., Patterson, T. L., Harvey, P. D., McClure, M. M. N., Mausbach, B. T., Taylor, M. J., &

Twamley, E. W. (2017). Exploratory analysis of normative performance on the UCSD Performance-Based Skills Assessment-Brief. *Psychiatry Research*, 256, 150–155.

<https://doi.org/10.1016/j.psychres.2017.06.025>

Ventura, J., Helleman, G. S., Thames, A. D., Koellner, V., & Nuechterlein, K. H. (2009).

Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: A meta-analysis. *Schizophrenia Research*, 113(2–3), 189–199.

<https://doi.org/10.1016/j.schres.2009.03.035>

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35.

[https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)