Validation of a European Spanish-version of the University of California performance Skills Assessment (Sp-UPSA) in patients with schizophrenia and bipolar disorder

Maria P. Garcia-Portilla, Jesus J. Gomara, Maria T. Bobes-Bascaranb, Isabel Menendez-Miranda, Pilar A. Saiz, Jose Muniz, Celso Arango, Thomas L. Patterson, Philip D. Harvey, Julio Bobs, Terry E. Goldberg

Aims: To validate the Spanish version of the University of California Performance Skills Assessment (UPSA) in patients with severe mental disorders.

Methods: Naturalistic, 6 month follow-up, multicentre, validation study. 139 patients with schizophrenia, 57 bipolar patients and 31 controls were evaluated using the following scales: Spanish UPSA (Sp-UPSA), Clinical Global Impression, Severity (CGI-S), Global Assessment of Functioning (GAF), and Personal and Social Performance (PSP).

Results: Reliability: Internal consistency (Cronbach’s alpha) was 0.81 in schizophrenia and 0.58 in bipolar patients. Test–retest was 0.74 and 0.65 (p < 0.0001) respectively. Construct validity: Pearson correlation coefficients between Sp-UPSA and PSP total scores were 0.42 (p < 0.0001) for schizophrenia and 0.44 (p = 0.001) for bipolar patients. For Sp-UPSA and GAF scores correlation coefficients were 0.43 and 0.52 (p < 0.0001) respectively. Discriminant validity: The Sp-UPSA discriminated between patients and controls. In schizophrenia patients it also discriminated among different levels of illness severity according to CGI-S scores. In control versus patients with schizophrenia contrasts, the area under the curve was 0.89 and a cut-off point of 85 provided a sensitivity of 82.7% and a specificity of 77.4%. In bipolar patients, the area under the curve was 0.85 and a cut-off point of 90 provided a sensitivity of 82.5% and a specificity of 64.5%.

Conclusion: The Spanish UPSA is a reliable and valid instrument for assessing functional capacity in severe mentally ill patients. It seems to be appropriate for use in clinical trials and in everyday clinical practice as a means of monitoring functional outcomes.

1. Introduction

Severe mental illnesses have a great impact in the patient’s lives in terms of years of life lost due to disability. Indeed, in high-income countries in people under 60 years bipolar disorder and schizophrenia are 5th and 7th respectively in the list of leading health conditions associated with moderate and severe disability, with prevalences above 2 millions (WHO, 2011).

Although the relevance of functional disability to severe mental disorders is clear, its measurement has several challenges that need to be addressed. Approaches to its measurement vary according to the concept utilized, the domains included, and the method employed for obtaining the information.

Currently the measurement of the disability separates functional capacity from real-world functioning, thus improving the outcome predictive capacity (Harvey et al., 2007). Furthermore, it is generally considered that performance-based measures avoid biases of social desirability and respondent’s impairments, as well as situational influences or reports based on limited behavioural observations (Patterson et al., 2001; Harvey et al., 2007; Mausbach et al., 2007, 2010).
According to FDA any clinical trial of cognitive enhancement in schizophrenia will require a co-primary measure with face validity for clinicians and consumers (Keefe et al., 2006) and able to provide evidence referring to the potential functional impact of cognitive changes (Green et al., 2008). In this sense, the University of California, San Diego (UCSD), Performance-Based Skills Assessment (UPSA) scale has been included in the Consensus Battery for Clinical Trials in Schizophrenia (NIMH-MATRICS initiative) (Mausbach et al., 2008) and has been identified as a leading co-primary measure for schizophrenia cognition trials (Keefe et al., 2013).

The UPSA is a performance-based instrument developed for assessing functional capacity, i.e., the capacity of persons with schizophrenia to adequately perform the skills necessary for daily functioning such as paying bills or rescheduling medical appointments under optimal conditions (Patterson et al., 2001; Heinrichs et al., 2006). Thus it provides an important new perspective on functional outcome measures in schizophrenia and severe mental illnesses (Heinrichs et al., 2006) since it minimizes the potential influence of environmental and social factors as well as informant biases and lacunae in knowledge in measuring real-world functioning (Goldberg et al., 2010).

Both UPSA and its brief version (UPSA-B) are well-validated, efficient and promising measures of functional capacity in patients with schizophrenia (Heinrichs et al., 2006; Keefe et al., 2006; Leifker et al., 2010; Mausbach et al., 2010), and with bipolar disorder (Mausbach et al., 2010; Leifker et al., 2011).

Although available data indicate that modified versions of co-primary tests perform similarly across different Western cultures (Harvey et al., 2009), a RAND panel emphasized the need for their validation into other cultures/languages. To our knowledge the UPSA has not yet been validated in Spanish, although it is the third most commonly spoken language in the world. The MATRICS-CT initiative developed three Spanish versions of MATRICS Consensus cognitive battery (MCCB), creating South American, Caribbean, and European Spanish versions.

The aims of this study were to translate and culturally adapt the UPSA instrument into European Spanish and assess its psychometrics (including floor and ceiling effects, reliability, and construct and discriminant validity) in Spanish patients with schizophrenia and bipolar disorder under standard treatment. This validation was aimed at relationship with real-world outcomes, rather than cognitive test performance, because our interest was in the extent to which the Sp-UPSA related to real-world functioning, similarly to previous reports conducted the US and other European countries.

2. Material and methods

2.1. Study design

This is a naturalistic, 6 month follow-up validation study conducted at 7 centres in Spain. It was approved by the Ethics Committee for Clinical Research of one of the centres, Hospital Universitario Central de Asturias, Oviedo, Spain and it is in accordance with 1975 Declaration of Helsinki, as revised in 1983. Written informed consent was obtained from all subjects prior to enrolment.

2.2. Subjects

Participants included 139 patients with stable schizophrenia, 57 patients with stable bipolar disorder and 31 healthy controls. Stability was defined as those patients who were clinically stable and did not require any change in their current pharmacological treatment during the 3 past months.

Patients’ inclusion criteria were (1) age ≥ 18 years; (2) ICD-10 diagnosis of schizophrenia or bipolar disorder; (3) currently on treatment for his/her illness; and (4) written informed consent to participate in the study. Controls’ inclusion criteria were (1) age ≥ 18 years; (2) without mental or relevant physical disorder; (3) CGI-S score = 1; and (4) written informed consent to participate in the study.

Exclusion criteria were designed to be minimal, due to the nature of the study, and only persons with intellectual developmental disorder, acquired brain injury, or who refused to participate in the study were excluded.

2.3. Clinical measures

For all subjects demographic and clinical data collection and assessments were made at baseline and 6-month follow up. For each patient, both assessments were made by the same clinician. Severity of the illness was assessed employing the Clinical Global Impression, Severity scale (CGI-S) (Guy, 1976). The level of functioning was assessed using the Global Assessment of Functioning (GAF) (APA, 1987) and the Spanish version of the Personal and Social Performance scale (PSP) (Garcia-Portilla et al., 2011). Information was collected from the patients themselves and, when possible, from the main caregiver. Finally, the Spanish adaptation of the University of California San Diego Performance-Based Skills Assessment (Sp-UPSA) was employed.

2.3.1. Instrument

The University of California San Diego Performance-Based Skills Assessment (UPSA) (Patterson et al., 2001) is a performance-based measure of everyday function task. It measures functional capacity, i.e. a person’s repertoire of daily living skills performed under optimal condition, through a variety of proxy scenarios that involve role-play and skills demonstrations (Heinrichs et al., 2006).

The original instrument was first translated into Spanish by two Spanish clinical psychologists who were fluent in the English language. They also culturally adapted the Sp-UPSA by removing or changing those domains/items that do not apply for Spain. In the finances domain dollars were converted to euros, and the check writing was replaced by how to pay an electricity bill. In the communication domain dialing home was substituted by dialing 112 (emergency number). A beach scenario was used instead of a zoo in the planning recreational activities domain and, in the transportation domain the metro of Valencia was used instead of the bus. Finally, the household domain was removed because it requires an analogue kitchen and many of our previous studies excluded this test because of difficulties in administering this test at field sites. Then, the Spanish version was back-translated by a Mexican psychologist fluent in English and it was finally approved by the original authors.

The Sp-UPSA assesses the functional capacity in the following 4 domains: finances, communication, planning recreational activities, and transportation. Financial ability is assessed by a counting change task as well as by having the patient identify important features of a household gas and electricity bill. To evaluate communication abilities, role-playing tasks utilizing a telephone, including rescheduling a medical appointment, are used. Recreation planning is evaluated by having the participant read an article in a newspaper about a beach, and then planning an outing. The transportation domain involves reading and interpreting a subway route map and schedule.

Each of the 4 Sp-UPSA domains generates a raw score that is converted to a domain score ranging from 0 to 25 points. The sum of the 4 domain scores yields a total score, potentially ranging from 0 to 100 points. In all cases higher scores indicate better performance.

2.4. Statistical analysis

The statistical analysis was done using the SPSS 17.0. The two-tailed level of significance used was 0.05. Chi-square, Student’s t test, and ANOVA (Duncan post hoc) were used to determine statistically significant differences according to demographic and clinical status.
Skewness and kurtosis were calculated to measure the shape of the distributions (values of skewness and kurtosis $+/−1$ were considered as good). Coefficient of variation (standard deviation/mean) and ceiling and floor effects (in the control group and in the patients' groups respectively) were also determined (number of controls with scores greater than 95% and number of patients with scores smaller than 5%).

The internal consistency of the Sp-UPSA and its domains was calculated using the Cronbach's alpha coefficient at item level. To calculate test–retest reliability we used the Pearson correlation coefficient between Sp-UPSA scores at baseline and at 6-month follow-up. The effect of the practice on the retest performance of the Sp-UPSA was determined calculating the effect sizes for the changes in performance at month 6 (Cohen's $d$).

Construct validity was calculated using the Pearson correlation coefficient between the total Sp-UPSA score and total scores on the PSP and GAF with the hypothesis that a moderate $r$ coefficient will be found as they are related but different constructs.

For analysing the discriminant validity, patients with schizophrenia were classified in four groups based on their CGI-S scores: doubtfully ill (CGI-S = 1–2), mildly (CGI-S = 3), moderately (CGI-S = 4) and severely ill (CGI-S = 5–7). Patients with bipolar disorder were classified in three groups based on their CGI-S scores: doubtfully ill (CGI-S = 1–2), mildly (CGI-S = 3) and moderately–markedly ill (CGI-S = 4–5). An ANOVA test (Duncan post hoc) was used to identify statistically significant differences in the Sp-UPSA scores according to clinical status (schizophrenia, bipolar or healthy control) and severity groups. The diagnostic performance of the Sp-UPSA to discriminate between patients and controls was analysed using the receiver operating characteristic (ROC) curve analysis.

3. Results

3.1. Sample

Table 1 shows demographic and clinical characteristics for patients and controls separately. Bipolar patients were older than patients with schizophrenia and healthy controls (46.8 versus 39.9 and 39.2 respectively) and a smaller proportion were males (56.1% versus 73.4% and 65.6%). Patients with schizophrenia and bipolar disorder were more likely to be permanent disabled due to mental illness than healthy controls (61.2% and 36.8% versus 0.0%).

3.2. Psychometric properties of the Sp-UPSA

3.2.1. Distribution characteristics of Sp-UPSA scores

Patients with schizophrenia or bipolar disorder scored significantly lower than healthy controls [69.8 (95% CI = 66.8 - 72.8) versus 75.5 (95% CI = 72.1 - 78.9) versus 89.9 (95% CI = 87.8 - 92.1), $F = 21.111, p < 0.0001$]. There were no differences between patients with schizophrenia and bipolar disorder.

The psychometric characteristics of the total Sp-UPSA score and for the 4 Sp-UPSA domains were shown in Table 2.

In healthy controls and patients with bipolar disorder total Sp-UPSA scores exhibited symmetrical and mesokurtic distributions, while in patients with schizophrenia the distribution was slightly left skewed and leptokurtic compared to the normal distribution. The Sp-UPSA showed a mild ceiling effect in healthy controls (5 controls scored ≥95%), but not a floor effect (only 1 patient with schizophrenia and 0 patients with bipolar disorder scored <5%).

3.2.2. Reliability

The Sp-UPSA scale had good internal consistency for patients with schizophrenia and bipolar disorder (Cronbach’s $\alpha$ of 0.90 and 0.80 respectively). In controls the internal consistency was moderate (Cronbach’s $\alpha = 0.44$). For both, patients with schizophrenia and with bipolar disorder, Cronbach’s alphas for the Sp-UPSA domains were good with the exception of the planning recreational activities domain in patients with bipolar disorder (Cronbach’s $\alpha$ in patients with schizophrenia: Finances = 0.74; Communication = 0.73; Planning recreational activities = 0.75; Transportation = 0.80 - Cronbach’s $\alpha$ in patients with bipolar disorder: Finances = 0.66; Communication = 0.75; Planning recreational activities = 0.50; Transportation = 0.75). In control subjects, Cronbach’s alphas were low approaching to moderate, ranging from 0.105 (Communication domain) to 0.490 (Transportation domain).

Test–retest reliability was adequate with Pearson correlation coefficients of 0.74 and 0.65 ($p < 0.0001$) for patients with schizophrenia and bipolar disorder respectively. In both schizophrenia and bipolar patients the highest retest correlation coefficient was found in the communication domain (0.66 and 0.72 respectively, $p < 0.0001$) while the lowest corresponded to the transportation domain in the case of patients with schizophrenia ($r = 0.44, p = 0.001$) and to the planning recreational activities in patients with bipolar disorder ($r = 0.17, p = 0.392$).

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Schizophrenia</th>
<th>Bipolar</th>
<th>Controls</th>
<th>Statistical test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (sd)</td>
<td>39.9 (10.4)</td>
<td>46.8 (10.6)</td>
<td>39.2 (11.3)</td>
<td>9.485$^a$</td>
</tr>
<tr>
<td>Gender, males [n (%)]</td>
<td>102 (73.4)</td>
<td>32 (56.1)</td>
<td>15 (48.4)</td>
<td>10.065$^b$</td>
</tr>
<tr>
<td>Educational level [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>78 (56.1)</td>
<td>28 (49.1)</td>
<td>0 (0.0)</td>
<td>55.251$^b$</td>
</tr>
<tr>
<td>Secondary school</td>
<td>48 (34.5)</td>
<td>19 (33.3)</td>
<td>6 (27.3)</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>University</td>
<td>13 (9.4)</td>
<td>10 (17.5)</td>
<td>16 (72.7)</td>
<td></td>
</tr>
<tr>
<td>Work status [n (%)]</td>
<td></td>
<td></td>
<td></td>
<td>102.369$^c$</td>
</tr>
<tr>
<td>Working</td>
<td>12 (8.6)</td>
<td>10 (17.5)</td>
<td>21 (95.5)</td>
<td>$&lt;0.0001$</td>
</tr>
<tr>
<td>Permanent disabled due to mental illness</td>
<td>85 (61.2)</td>
<td>21 (36.8)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Not working</td>
<td>37 (26.6)</td>
<td>19 (33.3)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Housewife</td>
<td>2 (1.4)</td>
<td>3 (5.3)</td>
<td>0 (0.0)</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>3 (2.2)</td>
<td>4 (7.0)</td>
<td>1 (4.5)</td>
<td></td>
</tr>
<tr>
<td>CGI-S (Mean (sd))</td>
<td>3.6 (1.3)</td>
<td>3.3 (1.5)</td>
<td>3.1 (1.5)</td>
<td></td>
</tr>
<tr>
<td>GAF [Mean (sd)]</td>
<td>57.4 (17.2)</td>
<td>70.2 (14.0)</td>
<td>99.3 (3.6)</td>
<td>98.075$^b$</td>
</tr>
<tr>
<td>PSP total score [Mean (sd)]</td>
<td>61.9 (19.6)</td>
<td>73.0 (14.6)</td>
<td>98.7 (7.2)</td>
<td>59.364$^b$</td>
</tr>
</tbody>
</table>


1. ANOVA F value. 2. Chi-square test. 3. Student’s $t$ test. a. Duncan test showed that schizophrenia patients and healthy controls were equal between them and different from bipolar patients. b. Duncan test showed that the 3 groups were different among them.
Practice effects on the Sp-UPSA were small in patients with schizophrenia while in bipolar patients they were medium (Cohen’s $d = 0.15$ and 0.34 respectively).

### 3.2.3. Construct validity

The Pearson correlation coefficient between the Sp-UPSA total score and the total score of the PSP was 0.42 ($p < 0.0001$) for patients with schizophrenia, 0.44 ($p = 0.001$) for bipolar patients and 0.33 ($p = 0.070$) for controls. In the GAF, the Pearson coefficients were 0.43 ($p < 0.0001$), 0.52 ($p < 0.0001$), and 0.33 ($p = 0.070$) respectively.

### 3.2.4. Discriminant validity

As stated above (see 3.2.1. and Table 2) the Sp-UPSA was able to discriminate between patients (schizophrenia or bipolar) and healthy controls, but not between patients with schizophrenia or bipolar disorder. In addition, it was also able to discriminate between the different levels of illness severity according to CGI-S scores in patients with schizophrenia (see Table 3).

### 4. Discussion

The aim of this study was to validate the Spanish version of the UPSA in a sample of patients with severe mental disorders receiving standard maintenance treatment in Spain. First, we translated and culturally adapted the Sp-UPSA. This is the appropriate method for developing an equivalent version of the scale that allows clinicians to assess patients’ in their own cultural context and also make comparisons.

### Table 2

Distribution characteristics of Sp-UPSA scores for patients with schizophrenia, bipolar disorder and healthy controls separately.

<table>
<thead>
<tr>
<th>Sp-UPSA domain</th>
<th>Schizophrenia</th>
<th>Bipolar</th>
<th>Controls</th>
<th>ANOVA, F value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finances</strong> [Mean (sd)]</td>
<td>17.3 (5.3)</td>
<td>19.3 (4.2)</td>
<td>22.3 (2.6)</td>
<td>15.133a</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Communication</strong> [Mean (sd)]</td>
<td>15.3 (5.6)</td>
<td>18.6 (5.3)</td>
<td>22.1 (2.1)</td>
<td>24.933b</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Planning recreational activities</strong> [Mean (sd)]</td>
<td>20.3 (5.2)</td>
<td>20.6 (4.1)</td>
<td>22.9 (2.0)</td>
<td>4.045b</td>
<td>0.019</td>
</tr>
<tr>
<td><strong>Transportation</strong> [Mean (sd)]</td>
<td>16.9 (6.4)</td>
<td>16.9 (5.4)</td>
<td>22.6 (2.9)</td>
<td>12.721b</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

### Table 3

Sp-UPSA scores according to severity of the disorder as determined by CGI-S scores.

<table>
<thead>
<tr>
<th>Sp-UPSA domain</th>
<th>Doubtfully ill (CGI-S = 1–2)</th>
<th>Mildly ill (CGI-S = 3)</th>
<th>Moderately ill (CGI-S = 4)</th>
<th>Severely ill (CGI-S = 5–7)</th>
<th>ANOVA, F value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finances</strong> [Mean (sd)]</td>
<td>19.3 (3.4)</td>
<td>18.4 (4.4)</td>
<td>16.3 (5.5)</td>
<td>15.4 (6.3)</td>
<td>4.152a</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>Communication</strong> [Mean (sd)]</td>
<td>18.9 (3.4)</td>
<td>16.8 (4.1)</td>
<td>13.1 (5.6)</td>
<td>13.2 (6.6)</td>
<td>9.815b</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Planning recreational activities</strong> [Mean (sd)]</td>
<td>22.4 (4.4)</td>
<td>20.1 (5.0)</td>
<td>20.6 (4.4)</td>
<td>18.3 (6.4)</td>
<td>3.363c</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Transportation</strong> [Mean (sd)]</td>
<td>19.2 (5.7)</td>
<td>18.2 (5.3)</td>
<td>15.0 (7.4)</td>
<td>15.3 (6.3)</td>
<td>3.828d</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Total score</strong> [Mean (sd)]</td>
<td>79.8 (12.6)</td>
<td>73.6 (13.1)</td>
<td>65.0 (17.8)</td>
<td>62.3 (22.0)</td>
<td>7.226b</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

### Table 4

### 4. Discussion

The aim of this study was to validate the Spanish version of the UPSA in a sample of patients with severe mental disorders receiving standard maintenance treatment in Spain. First, we translated and culturally adapted the Sp-UPSA. This is the appropriate method for developing an equivalent version of the scale that allows clinicians to assess patients’ in their own cultural context and also make comparisons.

### Table 3

Sp-UPSA scores according to severity of the disorder as determined by CGI-S scores.

<table>
<thead>
<tr>
<th>Sp-UPSA domain</th>
<th>Doubtfully ill (CGI-S = 1–2)</th>
<th>Mildly ill (CGI-S = 3)</th>
<th>Moderately - markedly ill (CGI = 4–5)</th>
<th>ANOVA, F value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finances</strong> [Mean (sd)]</td>
<td>20.3 (4.3)</td>
<td>18.9 (2.6)</td>
<td>18.9 (5.3)</td>
<td>0.624</td>
<td>0.540</td>
</tr>
<tr>
<td><strong>Communication</strong> [Mean (sd)]</td>
<td>20.0 (6.6)</td>
<td>17.1 (5.3)</td>
<td>18.9 (4.1)</td>
<td>1.419</td>
<td>0.251</td>
</tr>
<tr>
<td><strong>Planning recreational activities</strong> [Mean (sd)]</td>
<td>21.1 (4.6)</td>
<td>21.3 (3.0)</td>
<td>19.8 (4.5)</td>
<td>0.845</td>
<td>0.435</td>
</tr>
<tr>
<td><strong>Transportation</strong> [Mean (sd)]</td>
<td>16.7 (6.7)</td>
<td>17.5 (4.4)</td>
<td>16.7 (5.4)</td>
<td>0.139</td>
<td>0.853</td>
</tr>
<tr>
<td><strong>Total score</strong> [Mean (sd)]</td>
<td>78.2 (16.8)</td>
<td>74.8 (10.2)</td>
<td>74.3 (11.8)</td>
<td>0.471</td>
<td>0.627</td>
</tr>
</tbody>
</table>

sd: standard deviation. CGI-S: Clinical Global Impression, Severity.

a Duncan test showed that in the Finances domain severely ill patients were different from those mildly and doubtfully ill, and moderately ill patients were different from those doubtfully ill.
b Duncan test showed that in the Communication domain and in the Total score severely and moderately ill patients were different from those mildly and doubtfully ill.
c Duncan test showed that in the Planning recreational activities domain severely ill patients were different from those doubtfully ill.
d Duncan test showed that in the Transportation domain severely ill patients were different from those doubtfully ill, and moderately ill patients were different from those mildly and doubtfully ill.

---

**Practice effects on the Sp-UPSA were small in patients with schizophrenia while in bipolar patients they were medium (Cohen’s $d = 0.15$ and 0.34 respectively).**

In subjects with schizophrenia, the area under the ROC curve was 0.89 (95% CI = 0.83 to 0.94), indicating good accuracy of the test. A cut-off point of 85 provided good sensitivity (82.7%) and moderate specificity (77.4%) for separating HC and patients (see Fig. 1). Predictive values were: positive 94.3% and negative 50.0%.

In subjects with bipolar disorder, we found similar results. The area under the curve was 0.85 (95% CI = 0.77 to 0.93). A cut-off point of 90 provided good sensitivity (82.5%) and moderate specificity (64.5%) (see Fig. 1). Predictive values were: positive 81.0% and negative 66.6%.

**4. Discussion**

The aim of this study was to validate the Spanish version of the UPSA in a sample of patients with severe mental disorders receiving standard maintenance treatment in Spain. First, we translated and culturally adapted the Sp-UPSA. This is the appropriate method for developing an equivalent version of the scale that allows clinicians to assess patients’ in their own cultural context and also make comparisons.
Patients with bipolar disorder

![ROC curves for schizophrenia and bipolar patients](image)

**Fig. 1.** Total Sp-UPSA score ROC curves for schizophrenia and bipolar patients.

among studies from different countries/cultures. Second, it is essential to demonstrate that the new adapted scale has adequate psychometric properties in order to know its utility. In this regard, we found adequate psychometric properties, particularly in patients with schizophrenia, which introduces the Spanish version of the UPSA as a reliable and valid instrument for assessing functional capacity in patients with severe mental disorders in daily clinical practice. The relatively poor psychometric properties of the Sp-UPSA in patients with bipolar disorder may be due, at least partially, to the small sample size of this population. However, it should be kept in mind that there is a lack of psychometric research of the UPSA in these types of patients, although it has been employed in different studies (Henry et al., 2013; McIntosh et al., 2011; Mausbach et al., 2010). Thus, further research on the psychometric properties of the UPSA in patients with bipolar disorder should be conducted in order to ascertain if this scale, initially developed for its use in patients with schizophrenia, could also be used in patients with bipolar disorder.

Our results demonstrate that both Spanish patients with schizophrenia and bipolar disorder have impaired functional capacity compared with healthy controls. These results are in agreement with the well-known evidence that patients with severe mental disorders have deficits in their everyday functioning.

Previously reported UPSA total scores for patients with schizophrenia ranged from 58.8 (Patterson et al., 2001) to 83.2 (Heinrichs et al., 2006). Our schizophrenia and bipolar patients’ total scores fall between them and they were very similar to the scores reported by Leifker et al. (2010) in patients with schizophrenia (69.8, 75.5 and 72.2 respectively). Interestingly our total Sp-UPSA score for patients with schizophrenia was also close to that found by Goldberg et al. (2010) in patients with mild cognitive impairment (63.6), thus suggesting a similar level of functional capacity impairment across both diagnoses.

In patients with schizophrenia, the Sp-UPSA total score showed a slightly left skewed distribution in agreement with Heinrichs et al. (2006) but, contrary to them, our distribution for healthy controls (and bipolar patients) was symmetrical and similar to the reported by Goldberg et al. (2010) for healthy control and patients with mild cognitive impairment and Alzheimer disease.

Internal reliability for the total Sp-UPSA score exceeded 0.80 in both groups of patients. Its good internal reliability demonstrates that the four “finances, communication, planning recreational activities, and transportation” domain scores can aggregate to build an overall functional capacity score. However, in controls the internal reliability was only moderate (0.44), thus requiring further research in this population.

Test–retest reliability was moderate for both patients with schizophrenia (0.74) and bipolar disorders (0.65) and similar to the previously reported by Leifker et al., 2010 (0.75 and 0.77). A previous study reported a test–retest reliability of 0.93 (Harvey et al., 2007). However, that very high reliability could be due to the effect size of practice since the interval period of time between test and retest was only 2 weeks. In our study the interval period was 6 months; thus it seems appropriate to expect a lower value due to the larger follow-up. With respect to the impact of the practice on the Sp-UPSA performance the low effect size for our patients with schizophrenia (d = 0.15) was almost identical to the reported by Leifker et al. (2010) in the UCSD sample (0.16), while practice effects were larger in the bipolar sample.

The correlation coefficients between the total Sp-UPSA score and the scores on the PSP and GAF were as expected, positive and weak in magnitude, thus confirming the hypothesis that these instruments measure related but not identical constructs. This may reflect differences between functional capacity under ideal conditions versus functional performance in real world, the former suffers from a limited set of proxies, the latter from lack of control over environment and informant lacunae.

As previously reported (Patterson et al., 2001; Heinrichs et al., 2006) we found that the Sp-UPSA discriminated between patients and controls in the expected direction—patients obtained significantly lower Sp-UPSA scores than controls. In addition, in patients with schizophrenia it discriminated among degrees of severity illness; those patients more severely ill scored significantly lower than mildly and doubtfully ill patients.

Finally, we established the sensitivity and specificity of the Sp-UPSA for identifying patients and controls based on their level of functional capacity. Although the specificity of the Sp-UPSA was moderate both in patients with schizophrenia and bipolar disorder, it was sufficiently sensitive to be of clinical value. A total score above 90 was indicative of a level of functional capacity compatible with a Spanish person older than 17 years without mental disorder.

The generalizability of this study can be considered an advantageous feature of the study since the patients enrolled in this study were similar to patients on maintenance treatment for a severe mental disorder seen in daily clinical practice throughout Spain. The study inclusion and exclusion criteria were very non-restrictive and it was a multicentre study that included patients from seven different cities in Spain. In addition, the
follow-up design and the inclusion of a control group are other two strengths of the study. However, some limitations in the design of the study should be acknowledged. First, the small sample size of patients with bipolar disorder, could have influenced the relatively poor psychometric properties of the Sp-UPSA in this population. Second, is the lack of specific training for psychometricians on the application of the Sp-UPSA scale. However, the potential negative impact of this lack of training on reliability, is minimized by the fact that the instruments used for assessing both types of functioning, i.e., capacity (Sp-UPSA) and real-world functioning (PSP), are structured scales with well-defined operational criteria.

In conclusion, we were able to demonstrate that the Spanish version of the UPSA is an instrument that is reliable and valid for measuring functional capacity in patients with severe mental disorders. As a performance-based clinician-rated instrument, it seems to be appropriate for use in clinical trials as a co-primary endpoint and in everyday clinical practice as a means of identifying and monitoring changes in functional capacity in this population.

Role of funding source
This study was supported by the Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM.

Contributors
MP Garcia-Portilla, J Bobes, and T. Goldberg conceptualized the project, designed the tasks and wrote the first draft of the manuscript.
MP Garcia-Portilla, J Gomar, and T. Goldberg analyzed the data.
All the authors reviewed the manuscript and contributed to its final version.

Conflict of interest
The authors declare that they have no competing interests.

Acknowledgment
We would like to thank Manuel Arrojo, Gonzalo Galvan, Natalia Ojeda, Fernando Sarramea and Pilar Sierra for their participation as raters.

References