Spanish Cultural Adaptation and Inter-Rater Reliability of the Test of Playfulness

Sergio Serrada-Tejeda, Sergio Santos-del-Riego, Anita Bundy, and Marta Pérez-de-Heredia-Torres

a Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Rey Juan Carlos University, Avenida de Atenas s/n. CP.28922, Alcorcón, Madrid, Spain;
b Department of Physical Therapy, Medicine and Biomedical Sciences, Universidade da Coruña, A Coruña, Spain;
c Department of Occupational Therapy, Colorado State University, Colorado, USA;
d Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Rey Juan Carlos University, Avenida de Atenas s/n. CP.28922, Alcorcón, Madrid, Spain

Abstract

Aim

To adapt the Test of Playfulness (ToP) to the Spanish context (culturally and linguistically) and analyze the inter-rater reliability of the translated version.

Methods

Following a specific methodology for cultural adaptation, direct and back translations were carried out, together with cognitive interviews with pediatric occupational therapists, to analyze the comprehensibility of the translated version. The resulting version was analyzed by a linguistic expert to determine its grammatical and linguistic adequacy. Finally, inter-rater reliability was analyzed in a sample of typically developing children aged between 4 years old and 6 years and 11 months (n = 12).

Results

Direct and back translation phases showed a good conceptual equivalence with the original version. The cognitive interviews enabled the identification of difficulties surrounding the understanding and the terminology used in several items. The linguistic expert and the research team revision ensured that final grammatical and conceptual improvements were made. Inter-rater reliability analysis showed an almost perfect degree of agreement (0.82 ≤ κ ≤ 1.00; p < 0.001), as well as stability and a strong positive linear correlation (0.81; p < 0.001).
Conclusions

The translated version of the ToP fits the Spanish context and can be used in the clinical setting of pediatric occupational therapy in Spain.

Keywords

Children; cultural adaptation; playfulness; assessment; test of Playfulness

Among children, play is considered one of the primary occupations, representing one of the main areas of interest for Occupational Therapy (OT) professionals working with children (Bundy, 1992; Parham & Fazio, 2008; American Occupational Therapy Association (AOTA) 2017). Although play is considered the natural setting in which children develop motor, cognitive and social skills, its multiple dimensions and conceptualizations hinder and limit the development of assessments that can be used internationally and are currently considered the gold standard (Stagnitti, 2004). To address play as an occupation, it is necessary to understand how the child approaches the activity (Bundy, 1993; Ferland, 2005). This perspective, which focuses more on the observation of the process and disposition toward play or playfulness, may encourage occupational therapists to consider play more broadly by facilitating the use of assessments based on occupational observation.

Historically, play-based assessments have been approached by professionals from a functional perspective, considering play as an indicator that enables the observation and quantification of changes in developmental skills (Bundy, 1993; Kaugars, 2011; Miller Kuhaneck et al., 2013). However, for occupational therapists, these assessments, based on the observation of functional skills, offer a limited view that contrasts with the comprehensive nature of play, thus providing scant information about the adjustments that occur between the experience of the performance, the child's skills, and the opportunities that are inherent to the environment (Law et al., 2005). Although there is a large amount of literature that emphasizes the importance of addressing play from an occupational perspective (Bundy, 2011a; Miller Kuhaneck et al., 2013; Lynch et al., 2016; Lynch et al., 2016; Lynch et al., 2018), assessment tools are limited (Miller Kuhaneck et al., 2013). It is therefore necessary to have assessments that consider play as an occupation that is meaningful and purposeful for children (Kielhofner, 2008).

One of the assessments that offers a conceptualization of play as an occupation and as a goal in OT is the Test of Playfulness (ToP) (Bundy et al., 2001; Parham, 2008; Skard & Bundy, 2008). The ToP enables observation of the interactions that occur during a play situation between the child and the environment, providing information about the style and disposition toward playing. The exchanges and transactions that are observed allow one to distinguish a play situation from one that is not, due to the presence of four essential elements: intrinsic motivation, internal control, freedom to suspend reality, and framing (Parham, 2008; Skard & Bundy, 2008).

The ToP was developed in an English-speaking country and there are no versions available that are adapted to the Spanish population, and which could facilitate their use in a different linguistic context. Furthermore, for the ToP to provide valid information during clinical assessments, it is necessary to have culturally adapted versions, with appropriate metrics that enable their extended use, while maintaining the original meaning of the test. Thus, the aim of this study was to conduct a process of cultural adaptation of the ToP into Spanish, maintaining the semantic and conceptual equivalence of the original version, in addition to analyzing the inter-rater reliability of the tool.
Methods

This study was conducted in Spain as part of a larger research project on the role of play in children aged 4–6 with autism spectrum disorder (ASD). For the cultural adaptation, the linguistic criteria developed by The International Test Commission (ITC, 2005) and the methodology proposed by Ramada-Rodilla et al. (2013) were taken into consideration. This methodology includes the phases of direct and reverse translation, cognitive interviews, and review by a linguistic expert. The study was approved by the Clinical Research Ethics Committee of Universidad Rey Juan Carlos.

Participants

Five groups of participants were formed (Figure 1). A first group formed by bilingual experts (n = 5), was divided in two different teams, who developed the translation of the test: the first team of translators (n = 3) consisting of two expert bilingual translators and one bilingual occupational therapist, carried out the direct translation. The second translation team (n = 2), who accomplished the reverse translation, was formed by expert bilingual translators. A second group constituted a convenience sample of occupational therapists who participated in the cognitive interviewing process (n = 8). A third group comprised the linguistic expert, who oversaw the Spanish version of the test (n = 1). Finally, the fifth group, was a panel of experts (n = 3) in charge of reviewing and accepting the versions resulting from the direct and reverse translation processes of the test.

For the cultural adaptation phase, participants had to be residents in Spain. The criteria for the inclusion of the group of occupational therapists who carried out the cognitive interview phase were that they had to be OT graduates, have clinical experience working with children aged three or older, and hold Spanish nationality. The group of experts included an occupational therapist, a physiatrist, and a pediatric occupational therapist with over six years of experience in clinical treatment and evaluation.
Two occupational therapists, with more than six years of experience in the field of pediatrics, were involved in the inter-rater reliability analysis phase. Both participants used the final Spanish version of the ToP to evaluate and score the play recordings in a sample of twelve typically developing children aged between four years old to six years and 11 months.

**Measures**

Test of playfulness (ToP) (Skard & Bundy, 2008): The ToP is an observational tool that evaluates the child's disposition to play, from six months to 18 years, through the operationalization of the four basic elements of play: intrinsic motivation, internal control, suspension of reality, and framing or ability to understand and provide clues during play interactions. For this study, version 4.4 (Bundy, 2017) was used, which includes a total of 28 items, scored from 0 to 3 depending on the extent (3: almost always; 2: much of the time; 1: some of the time; 0: rarely or never) the intensity (3: highly; 2: moderately; 1: mildly; 0: not) and the skillfulness (3: highly skilled; 2: moderately skilled; 1: slightly skilled; 0: unskilled) demonstrated by the child during play. The ToP has good clinical utility, and adequate psychometric reliability and validity properties (Brentnall et al., 2008; Bundy et al., 2001; Harkness & Bundy, 2001).

**Procedure**

**Direct Translation**

In this first phase, the principal investigator sent the original test to the first team of translators, formed by two bilingual experts and one bilingual occupational therapist, for direct translation from English to Spanish. Each of the translators independently completed and submitted their translated version. After obtaining the three versions, the group of experts reviewed them and elaborated the first draft of the test in Spanish. This document was sent to each of the three members of the direct translation team to evaluate the degree of conceptual equivalence with the original version (Román-Oyola & Reynolds, 2010; Gándara-Gafo et al., 2019) identifying the following options: (A) equivalent translation; (B) moderate conceptual equivalence; and (C) no equivalence. Those translations scored with criteria B or C were sent back to the first team of translators to request an alternative translation, which was later reviewed and accepted by the panel of experts.

**Back Translation**

Once the Spanish version was developed, a procedure similar to direct translation was carried out. During this phase, the second team, formed by two bilingual translators was involved in independently performing a back translation of the text into English. As in the previous phase, the group of translators identified the degree of linguistic and semantic equivalence, and finally the panel of experts analyzed and compared both versions with the original English version, ultimately producing a draft with the translated version.
Cognitive Interviews

Once the translation process was completed, semi-structured cognitive interviews using retrospective verbal probing were carried out to analyze the comprehensibility and adequacy of the vocabulary of the translated version. Cognitive interviewing is a respondent-centered research technique that uses an in-depth approach to study how targeted audiences understand, mentally process, and respond to materials such as assessment items. The methodological cognitive interviews recommendations suggested by Willis (2005), were followed, which state that a sample of 5 to 10 participants should be available, in addition to audio recordings to facilitate the recording of the interviewees' comments (Willis, 2005; Beatty & Willis, 2007). The eight occupational therapists were asked the following questions: “How do you interpret what the item is asking?” and “What changes, if any, would you make to the item to improve its understanding?”. All answers were scripted and spontaneous for examining whether the participants understood the questionnaire’s items the way they were intended. For each item, and according to Willis (2005), participants were asked “thinking out loud” to generate verbal information and to interpret, reformulate, and prepare a modification proposal to improve understanding of the text if necessary. Following the recommendations of Román-Oyola and Reynolds (2010), when difficulties in understanding were observed in two or more participants, the group of experts reviewed each item to improve its comprehensibility.

Linguist Expert Review

After the process of cognitive interviews was concluded, the linguistic expert reviewed the adequacy of the comprehensibility difficulties encountered and their adaptation by the research team to the grammatical and cultural context of the population.

Inter-Rater Reliability

In order to participate in the inter-rater reliability process, and as a compulsory requirement for this procedure, signed consent was requested from each minor’s parent or legal guardian, by means of an informed consent form, for the purpose of filming a 15–20 min video recording of a play situation. The video recording was carried out at the participant's home and was subsequently evaluated and scored at the Cognitive Intervention Laboratory of the Faculty of Health Sciences of the Rey Juan Carlos University.

The sample (n = 12), was divided regarding their age range into three different groups: group 1 (n = 4), from 4–0 to 4–11 years old and a mean age of 4.5 ± 0.2 years; group 2 (n = 4), from 5–0 to 5–11 years and a mean age of 5.5 ± 0.3 years, and a third group (n = 4), from 6–0 to 6–11 years old and a mean age of 6.4 ± 0.3 years.

For the analysis of inter-rater reliability, the Cohen's Kappa index (κ) and Pearson's correlation coefficient (r) were calculated using the IBM SPSS Statistics statistical program for Windows software, version 22.0 (Copyright© 2013 IBM SPSS Corp.) Cohen's Kappa coefficient (κ), a coefficient that enables the evaluation and measurement of the chance corrected agreement between two evaluators (Cohen, 1960). Kappa takes values between −1 and 1, which can be interpreted in five different levels of agreement according to Landis and Koch (1977): < 0.00, zero degree of agreement; between 0.00 and 0.20, slight degree of agreement; between 0.21 and 0.40, acceptable degree of agreement; between 0.41 and 0.60, moderate degree of agreement; between 0.61 and 0.80, considerable degree of agreement; between 0.81 and 1.00, almost perfect degree of agreement. To complement the inter-rater reliability analysis, Pearson's correlation coefficient was calculated. This coefficient is a dimensionless measure of the covariance, which is scaled such that it ranges from −1 to +1 (Wackerly et al., 2008) and it is frequently used when the results of the evaluation process are expressed with measurements that involve variables of a continuous
quantitative nature, determining the degree of linear association between two variables and assessing their measurement’s stability.

Results

Direct and Back Translation

From the results obtained in the analysis of direct and reverse translation, it was concluded that, after the direct translation, 81% of the items showed similarities in the translation and wording of the items performed by the translators (A), whereas 19% showed a moderate conceptual equivalence (B). After the back-translation stage, 90% of the items showed adequate translation equivalence (A), and the remaining 10% showed a moderate conceptual equivalence (B). The group of experts unified the direct and back translations and made slight cultural adjustments due to the use of expressions and/or phrases that made the text difficult to understand. Some examples of the changes made in this phase are described in Table 1.

Table 1. Examples of direct/back translation procedures.

<table>
<thead>
<tr>
<th>Original test item</th>
<th>Spanish translation</th>
<th>Recommended modifications</th>
<th>Final translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tries to overcome barriers or obstacles to persist with an activity.</td>
<td>Trata de superar obstáculos o barreras para persistir con la actividad</td>
<td>Replace “Trata” with “intenta”</td>
<td>Trata de superar dificultades, barreras u obstáculos para persistir con la actividad.</td>
</tr>
<tr>
<td>Pretends (to be someone else; to do something else; that an object is something else; that something else is happening.</td>
<td>Simula (ser otra persona; como que hace otra cosa; que un objeto es otra cosa; que algo está sucediendo).</td>
<td>Replace “simula” with “finge”</td>
<td>Finge (ser otra persona; como que hace otra cosa; que un objeto es otra cosa; que algo está sucediendo).</td>
</tr>
<tr>
<td>Initiates play that others take up.</td>
<td>Inicia un juego que otros siguen.</td>
<td>Replace “siguen” with “son capaces de seguir”</td>
<td>Inicia un juego que otros son capaces de seguir</td>
</tr>
<tr>
<td>Transitions from one play activity to another.</td>
<td>Realiza transiciones de un juego a otro</td>
<td>Replace “realiza transiciones” with “cambia” Replace “de un juego a otro” with “de una actividad de juego a otra”</td>
<td>Cambia de una actividad de juego a otra.</td>
</tr>
</tbody>
</table>
Cognitive Interviews

The occupational therapists interviewed showed a good understanding of most of the translated items, being able to paraphrase and pose similar alternatives to the original item. However, in four of the items, comprehension difficulties arose for two or more participants, due to the use of certain words and terms, as well as phrasing. Thus, these items were finally reviewed and modified by the research team, as shown in Table 2.

Table 2. Examples of cognitive interviews process.

<table>
<thead>
<tr>
<th>Original English Item</th>
<th>Spanish translation</th>
<th>Recommended modification</th>
<th>Final Translation and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feels sufficiently safe to keep playing.</td>
<td>Se siente suficientemente seguro para continuar jugando.</td>
<td>No understanding difficulties are observed. However, four participants suggest replacing &quot;suficientemente&quot; to &quot;suficiente&quot; and &quot;seguro&quot; to &quot;seguridad&quot;</td>
<td>Siente la seguridad suficiente para seguir jugando</td>
</tr>
<tr>
<td>Pretends (to be someone else; to do something else; that an object is something else; that something else is happening.</td>
<td>Finge (ser otra persona; como que hace otra cosa; que un objeto es otra cosa; que algo está sucediendo).</td>
<td>Difficulties of understanding arise in more than two participants with the phrase &quot;que algo está sucediendo&quot;. They suggest adding the term &quot;diferente&quot; to the sentence, to avoid confusion with the option &quot;cómo que hace otra cosa&quot;.</td>
<td>Finge (ser otra persona; como que hace otra cosa; que un objeto es otra cosa; que algo diferente está sucediendo).</td>
</tr>
<tr>
<td>Supports play of others.</td>
<td>Apoya el juego de los demás.</td>
<td>Three participants suggest adding the term &quot;facilita&quot; to improve the comprehensibility of the item</td>
<td>Facilita y apoya el juego de los demás.</td>
</tr>
<tr>
<td>Demonstrates positive affect during play.</td>
<td>Muestra afecto positivo durante el juego.</td>
<td>Several doubts arise in five participants regarding the term &quot;afecto positivo&quot;. They propose to add the term &quot;estado&quot; to facilitate understanding of the item</td>
<td>Muestra un estado afectivo positivo durante el juego</td>
</tr>
</tbody>
</table>

Review by an Expert Linguist

The expert linguist identified and reviewed the proposed translations which were considered adequate, as well as including and suggesting various modifications to support the necessary grammatical and linguistic adjustment to the Spanish context. For example, the first version of the item Clowns or jokes, was translated as "Hace el payaso o bromea", and after the linguistic revision, it was suggested to change the expression "hace el payaso" for the term "payasadas", because it has a less negative connotation and adapts better to the original meaning of the item; in addition, the verb "bromea" was modified for the expression "gastar bromas", which is a more accurate meaning compared to the original version.
**Inter-Rater Reliability**

The ToP items are scored on duration, intensity, and skill observed during the playing situation; therefore, the inter-rater reliability and linear association results were grouped taking into consideration the different established categories (Table 3). Of the items analyzed, 10 showed a considerable degree of agreement (0.64 ≥ κ ≤ 0.80; p < 0.01). The remaining items, a total of 18, showed a near perfect degree of agreement (0.82 ≤ κ ≤ 1.00; p < 0.00).

Most items showed strong positive linear correlation and stability values, higher than 0.81 (p < 0.00), except items 2 and 5 which showed medium linear stability and linear correlation values (0.62 ≤ r ≤ 0.65; p < 0.01).

**Table 3. Inter-rater reliability.**

<table>
<thead>
<tr>
<th>No.</th>
<th>κ</th>
<th>r</th>
<th>No.</th>
<th>κ</th>
<th>r</th>
<th>No.</th>
<th>κ</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.68</td>
<td>0.81***</td>
<td>2</td>
<td>0.65</td>
<td>0.62*</td>
<td>3</td>
<td>0.71</td>
<td>0.89**</td>
</tr>
<tr>
<td>4</td>
<td>0.83</td>
<td>0.89**</td>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
<td>7</td>
<td>0.75</td>
<td>0.89**</td>
</tr>
<tr>
<td>5</td>
<td>0.63</td>
<td>0.63*</td>
<td>15</td>
<td>0.84</td>
<td>0.90**</td>
<td>8</td>
<td>0.84</td>
<td>0.90**</td>
</tr>
<tr>
<td>9</td>
<td>0.83</td>
<td>0.84**</td>
<td>25</td>
<td>0.81</td>
<td>0.87**</td>
<td>10</td>
<td>0.86</td>
<td>0.91**</td>
</tr>
<tr>
<td>11</td>
<td>0.83</td>
<td>0.84**</td>
<td>26</td>
<td>0.86</td>
<td>0.92**</td>
<td>12</td>
<td>0.80</td>
<td>0.81**</td>
</tr>
<tr>
<td>14</td>
<td>0.82</td>
<td>0.87**</td>
<td>13</td>
<td>0.76</td>
<td>0.91**</td>
<td>15</td>
<td>0.80</td>
<td>0.91**</td>
</tr>
<tr>
<td>22</td>
<td>0.85</td>
<td>0.91**</td>
<td>16</td>
<td>0.85</td>
<td>0.90**</td>
<td>17</td>
<td>0.86</td>
<td>0.91**</td>
</tr>
<tr>
<td>18</td>
<td>0.82</td>
<td>0.83**</td>
<td>19</td>
<td>0.80</td>
<td>0.81**</td>
<td>20</td>
<td>0.85</td>
<td>0.91**</td>
</tr>
<tr>
<td>21</td>
<td>0.96</td>
<td>0.95**</td>
<td>23</td>
<td>0.84</td>
<td>0.90**</td>
<td>24</td>
<td>0.85</td>
<td>0.90**</td>
</tr>
<tr>
<td>27</td>
<td>0.76</td>
<td>0.83**</td>
<td>28</td>
<td>0.80</td>
<td>0.81**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: κ (Cohen’s Kappa); r (Pearson).
*p < 0.01; **p < 0.001.

**Discussion**

The use of culturally adapted assessment tools is an essential element when an assessment is used in a different language, environment, and time period, to reduce the risk of introducing bias into a study (Gjersing et al., 2010). Due to the limited availability of assessments that evaluate play in Spain, their use is limited, and, as in other countries (Lynch et al., 2018), the evaluation of play has been reduced to the use of unstructured observations or lists in the clinical and research setting. This study represents the first version of the ToP translated and culturally adapted into Spanish, a play-occupation-centered tool, thus allowing its extensive use in Spanish-speaking communities.

This study has followed the methodological phases recommended internationally by the ITC (2005), ensuring that the procedure of cultural adaptation of the final translated version is conceptually and semantically adjusted to the context of its application. To this end, direct and back translation phases were followed by bilingual experts, comprehensibility interviews were conducted, and the final revision was overseen by an expert linguist. In addition, and following the methodological proposals indicated by Muñiz et al. (2013), to ensure the comprehensibility of the final version, a panel of experts was formed by the members of the research group, who had experience in carrying out cultural adaptation processes.
The results of the cultural adaptation process revealed that the Spanish version of the ToP displays a conceptual and semantic equivalence similar to the English version (Bundy et al., 2001). The main difficulties observed in the cognitive interviews were related to the infrequent use of some of the terms used, as well as the organization and structuring of some of the phrases, which promoted ambiguity in the understanding of the item (e.g., facilita y apoya for supports; or seguridad for safe).

This process of cultural adaptation makes it possible to satisfy the demands and needs that have been reported in previous studies, such as that of Lynch et al. (2018), in which it was observed that the current tools do not respond to the current needs of the professional community, and that despite evidence evaluating play from an occupational perspective, the lack of standardization studies limits their usefulness. This situation has been observed in other studies, such as the study conducted by Miller Kuhaneck et al. (2013), which indicate that, despite the importance of play assessment in occupational therapy, its deployment has decreased probably due to difficulties in accessing and using these tools, as well as the lack of specific training in their administration.

Of the total items analyzed in the inter-rater reliability phase, most showed a good degree of agreement and almost perfect correlation. These results were also observed in items with considerable inter-rater agreement, except in items 2 and 5, which showed medium linear association and stability. To avoid bias in the interpretation of these items, the indications for scoring the parameters of duration and specific intensity of each item were reviewed to avoid errors in the interpretation of their meaning.

**Limitations and Recommendations for Further Research**

A limited sample of participants without developmental condition was obtained through convenience sampling for the cultural adaptation and inter-rater reliability analysis phases, therefore some selection bias may have occurred. Although the version obtained in the study was developed in Spain and specifically for the Spanish population, Spanish-speaking countries can use it by making the necessary linguistic adjustments, ensuring that the corresponding version is better suited to the corresponding cultural context. We plan to conduct further research on the Spanish version of the ToP with a larger and more diverse sample including Rasch analysis.

**Conclusions**

Comprehensive translation procedure based on international standards for the adaptation of assessment tests, guarantees the linguistic and grammatical adjustment necessary to facilitate the comprehension. Moreover, the inclusion of occupational therapists specialized in the pediatric rehabilitation ensures that the grammatical adjustments of the scale are adapted to the scope of application and use of the test. Furthermore, adequate preliminary results have been obtained from the inter-rater reliability properties of the translated version, which, together with the results of the cultural adaptation, suggest that Spanish version of the ToP can be used in the clinical setting of OT applied to children in Spain.

**Disclosure statement**

The authors report no conflict of interest.
Author Contributions

Conceptualization, S.S-T; methodology, S.S-T, M.P-d-H-T and S.S-d-R; formal analysis, S.S-T; investigation, S.S-T, M.P-d-H-T and S.S-d-R; data curation and statistics, S.S-T; writing—original draft preparation, S.S-T; writing, review and editing, S.S-d-R, A-B and M.P-d-H-T; supervision, S.S-d-R, A-B and M.P-d-H-T. All authors have read and agreed to the published version of the manuscript.

References