Development of an objective instrument to assess quality of life in social services: Reliability and validity in Spain

Miguel Ángel Verdugo (University of Salamanca, Spain), Benito Arias (University of Valladolid, Spain), Laura E. Gómez (University of Valladolid, Spain), and Robert L. Schalock (Hastings College, USA)

ABSTRACT. This study reports on the development and validation of a questionnaire that allows the objective assessment of the quality of life of social service users. Because of the increasing demand for applying the concept of quality of life in social services for handicapped people, and because of the need to have a sensitive tool to use as a basis for organizational changes, the development of GENCAT Scale was carried out. The first study entailed the construction of items based on our current understanding of quality of life domains and indicators. In the second study, the psychometric properties of the Catalonian version of the GENCAT Scale were analyzed after applying the instrument to a representative sample of 608 professionals and 3,029 users of 239 services from Catalonia. The results showed reliability in terms of internal consistency of .92. The scale’s dimensionality was studied using confirmatory factor analysis (CFA). Finally, the use of this scale is discussed as a useful and innovative tool to serve as the basis for evidence-based practices, the improvement of services and programs.


1 This work was supported by funding from the Catalonian Institute of Assistance and Social Services (Action and Citizenship Department, Generalitat of Catalonia, Spain) and developed under the framework of research activities related to 1+D SEJ2006-12575 project (Spanish Ministry of Science and Innovation) and the Excellence Research Group on Disabilities (GR 197; EDU/894/2009) (‘Junta de Castilla y León’, Spain). The authors are grateful to services, centers, entities and professionals for their collaboration in the field test of this study.

2 Correspondence: Institute on Community Integration (INICO). Facultad de Psicología. Universidad de Salamanca. Avda. de la Merced, 109-131. 37005 Salamanca (Spain). E-mail: verdugo@usal.es
RESUMEN. Este estudio instrumental describe el proceso de desarrollo y validación de un instrumento de evaluación objetiva de la calidad de vida de usuarios de servicios sociales. Ante la creciente demanda para aplicar el concepto de calidad de vida en los servicios dirigidos a personas en situación de desventaja social y ante la necesidad de contar con instrumentos sensibles a los cambios organizacionales, se planteó el desarrollo de la Escala GENCAT. El primer estudio consistió en la elaboración y evaluación teórica de los ítems. En el segundo estudio se analizaron las propiedades psicométricas de la versión catalana del instrumento tras su aplicación a una muestra representativa de Cataluña formada por 608 profesionales y 3.029 usuarios de 239 centros y organizaciones catalanas proveedoras de servicios sociales. Los resultados indicaron una consistencia interna de .92. La dimensionalidad de la escala se estudió mediante análisis factorial confirmatorio (AFC). Finalmente, se discute el uso de la escala como un instrumento útil y novedoso para la mejora continua de la calidad de los servicios y para el desarrollo de programas basados en evidencias empíricas.


This article describes an instrumental study (Montero and León, 2007) whose goals consisted in developing and validating an objective quality of life (QOL) questionnaire for users of social and human services: the GENCAT Scale3 (Verdugo, Arias, Gomez, and Schalock, 2008a, 2008b). During the last three decades, the concept of quality of life has evolved from a generic philosophical concept to a sensitizing notion that guides program practices, from an individual perspective to a social construct that allows assessing the core domains of quality of life and guides quality improvement, and to a research construct that serves as a systematic structure to develop politics and practices to enhance peoples’ quality of life. It is widely accepted that this concept is important in social, health, human and educative services since it is currently being used to implement person-centered programmes and practices, to assess and report personal outcomes, to guide quality improvement strategies, and to evaluate the effectiveness of those practices and strategies (Martin-Rodríguez and Pérez-San Gregorio, 2007; McCabe, Firth, and O’Connor, 2009; Schalock, Bonham, and Verdugo, 2008). The QOL concept has recently begun to be applied in social policies since it has become a reference model for service provision, a basis for evidence-based practices, and a tool to develop quality improvement strategies (Schalock and Verdugo, 2002, 2007, 2008).

The construction of an assessment tool must be supported by a theoretical framework. To that end, the GENCAT Scale has been developed on the basis of the eight-domain model proposed by Schalock and Verdugo (2002) and subsequent work regarding the model’s validation and cross-cultural use (Jenaro et al., 2005; Schalock, Verdugo, Bonham, Fantova, and Van Loon, 2008; Schalock et al., 2005; Verdugo, Arias, and Gómez, 2006; Verdugo, Gómez, Arias, and Martin, 2006; Verdugo and Schalock, 2003). The main reason to choose this conceptual framework among other existing ones (such as those proposed by Cummins, 2005; Felce and Perry, 1995; Gardner and Carran, 2005; Perry and Felce, 2005) was that it was the most cited in publications about person-centered QOL in...
disabilities last years. Although this framework was initially developed in the field of intellectual disabilities, its use is being expanding not only to other kind of disabilities, but also to other target populations such as person’s with chemical dependency (De Maeyer, Vanderplasschen, and Broekaert, 2009) and elderly (Gómez, Verdugo, Arias, and Navas, 2008).

According to this framework, QOL is composed of eight domains and core indicators (listed in Table 1) that define operationally each QOL domain. The indicator measurement results in personal outcomes that can be used for both reporting purposes and guiding organization improvements (Keith and Bonham, 2005; Langberg and Smith, 2006; Schalock and Bonham, 2003; Schalock, Bonham et al., 2008; Schalock, Verdugo et al., 2008; Veerman and van Yperen, 2007). QOL is defined as a concept that a) is multidimensional; b) has etic (universal) and emic (culture-bound) properties (Jenaro et al., 2005; Schalock et al., 2005); c) is influenced by personal and environmental factors; and d) has objective and subjective components (Schalock and Verdugo, 2007). Depending on the purpose and the perspective of the instrument developed, indicator items will be used to assess either person’s perceived well-being on the item (‘self-report’) or the person’s life experiences and circumstances (‘direct observation’) (Schalock, Bonham, and Verdugo, 2008).

### Table 1. Domains and core indicators of quality of life.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Indicators</th>
<th>Item examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional well-being (EW)</td>
<td>Mental stability; satisfaction, self-concept; lack of stress/negative feelings</td>
<td>He/she shows symptoms of depression.</td>
</tr>
<tr>
<td>Interpersonal relations (IR)</td>
<td>Social relationships; family relationships; He/she complains about his/her relationships with friends.</td>
<td></td>
</tr>
<tr>
<td>Material well-being (MW)</td>
<td>Housing conditions; workplace conditions; Services conditions; employment; incomes/salary; possessions</td>
<td>His/her incomes are not enough to afford whims.</td>
</tr>
<tr>
<td>Personal development (PD)</td>
<td>Education; learning opportunities; work abilities; functional abilities (personal competency; adaptive behavior, etc.); activities of daily living.</td>
<td>He/she is involved in the development of his/her individual planning.</td>
</tr>
<tr>
<td>Physical wellbeing (PW)</td>
<td>Health care; sleep; health consequences (sorrow, medication, etc.); health; mobility; technical assistance</td>
<td>He/she has sleep problems.</td>
</tr>
<tr>
<td>Self-determination (SD)</td>
<td>Autonomy; goals and personal preferences; decisions; choices</td>
<td>Other people decide how to spend his/her money.</td>
</tr>
<tr>
<td>Social inclusion (SI)</td>
<td>Participation; integration; supports</td>
<td>His/her family supports him/her.</td>
</tr>
<tr>
<td>Rights (RI)</td>
<td>Knowledge of rights; defense of rights; exercise of rights; privacy; respect</td>
<td>He/she suffers exploitation, violence or abuse.</td>
</tr>
</tbody>
</table>
In this sense, we can speak about objective and subjective measures and measurement instruments, depending on their purpose, content, and respondent (Bonham, Basehart, and Marchand, 2005; Bonham et al., 2004; Brown, Keith, and Schalock, 2004; Cummins, 1997, 2005; Gardner and Carran, 2005; Keith and Bonham, 2005; Keith and Schalock, 2000; Perry and Felce, 1995, 2005; Schalock, Bonham, and Verdugo, 2008; Schalock et al., 2002; Schalock and Felce, 2004; Schalock, Gardner, and Bradley, 2007; Schalock and Verdugo, 2002, 2007; Verdugo, Arias, and Gómez, 2006; Verdugo et al., 2008b; Verdugo, Arias, Gómez, and van Loon, 2007; Verdugo, Gómez, and Arias, 2007; Verdugo, Gómez, Arias, and Schalock, 2009; Verdugo, Gomez, Schalock, and Arias, in press; Walsh, Erickson, Bradley, Moseley, and Schalock, 2006). If an evaluator desires to assess personal outcomes and develop person-centered programs, subjective Likert-type scales answered by the client or user of the service should be applied (Schalock and Felce, 2004). In distinction, when the goal is program evaluation, service quality improvement, or to assess organizational changes, it is recommended to use objective questionnaires based on the direct observation of personal experiences and circumstances.

The purpose of this two part study was to develop an instrument and evaluate its psychometric properties that serves the purposes discussed above. In terms of its first purpose, the GENCAT Scale was developed through a systematic and rigorous method, following the recommendations for instrumental studies (Carretero-Dios and Pérez, 2007). The GENCAT developmental method has served nationally and internationally as a model for developing other multidimensional QOL scales focusing on the context (Schalock, Keith, Verdugo, and Gómez, in press; Arias, Gómez, Schalock, and Verdugo, 2007; Van Loon, van Hove, Schalock, and Claes, 2008). In reference to the second purpose, psychometric properties of the measures provided by the Catalanian version of the GENCAT scale were analyzed. As this is an objective instrument, professionals assess the quality of life of social service users after observing them. So, it must be noted that the psychometric properties of the GENCAT measures belong to the ratings made by the professionals rather than the users assessed.

**STUDY 1: ITEM CONSTRUCTION AND ITEM THEORETICAL ASSESSMENT**

The development of the GENCAT Scale involved the suggested steps for developing multidimensional QOL questionnaires (Schalock et al., in press; Verdugo et al., in press; Verdugo, Schalock et al., 2007).

**Step 1: Selecting representative items**

Once the conceptual and measurement framework has been defined and the goal of the assessment has been specified, the next step in the development of an instrument entails selecting representative items. A concept mapping approach is typically used to select specific items (Schalock et al., in press). With this goal, an exhaustive review of previous QOL research and literature was carried out. As the final scale should have 8 items per domain at least \((n = 64)\), a pool of more than 300 items based on that review was developed; the items were organized by the 8 domains and 46 core indicators that were found in the literature. Next, items were organized by microsystem (individual and
family level), mesosystem (organizational level) and macrosystem (societal level). As the
goal of the instrument was being sensitive to organizational changes and person-
centered programs, items referred to macrosystem level were eliminated since they are
not dependent on organizational interventions. The most representative items of the
remaining classification categories were selected and new items were created when it
was necessary (there should be at least 2 items by indicator, 10 items by domain, and
2 items by system –micro and meso- in each domain). Once all categories were represented
as it has been indicated, the selected items \( n = 106 \) were reformulated as third-person
statements to be answered by an external observer with 4-frequency options \( \text{Never or}
\text{hardly ever, Often, Sometimes, and Always or almost always}. \)

**Step II: Panel of experts**

With the goal of studying the evidence based on the content of the test, the 106
items were evaluated by 73 judges that were experts in QOL. About 16.4\% \( n = 12 \) were
experts on QOL in elderly, 11\% \( n = 8 \) in mental health, 19.2\% \( n = 14 \) in physical
disabilities, 4.1\% \( n = 3 \) in auditory disabilities, 8.2\% \( n = 6 \) in visual disabilities, 24.7\% 
\( n = 18 \) in intellectual disabilities, 11\% \( n = 8 \) in drug dependences, and 5.5\% \( n = 4 \)
in AIDS/HIV. Their task was to evaluate the 106 items, sorted by domains, from 1 to 10
in three categories: Idoneity (degree in which the item is adequate to assess the
domain), importance (degree in which the item is important to assess person-centered
QOL) and observability (degree in which the action is subject of direct observation).
Besides, the experts were asked for making reformulations, suggestions or commentaries
about the proposed items, and for adding new items and indicators if they considered
as important. Obtained information was analyzed using qualitative and quantitative
methods.

Expert concordance was analyzed by Bangdiwala’s Weighted Agreement Coefficient
\( \text{BW}_n \) (Bangdiwala, 1987), a weighted kappa statistic which takes into account different
marginal frequencies in the data. As Cohen’s kappa, Bangdiwala’s coefficient ranges
from 0 (\text{No agreement}) to 1 (\text{Perfect agreement}). The agreement strength mean was
substantial (\( M = .61 \), as it is shown in Table 2 and taking into account the high number
of participating experts \( n = 73 \), when normally 6-10 participants participate in this kind
of studies). The mean rank comparison among groups (elderly, intellectual disabilities,
auditory disabilities, visual disabilities, mental health, drug dependences, and HIV/
AIDS) was calculated for idoneity, importance and observability using Kruskal-Wallis
test. The groups presented significant differences \( p < .05 \) in idoneity mean for 13 items,
in importance for 14, and in observability for 4 items. In total, 22 items showed significant
differences for one criterion (only one item showed significant differences for the three
criteria). In this way, a high agreement for idoneity, importance and observability was
found for 84 items (79.25\%). Thus, very high correlation coefficients \( p < .001 \) were
observed among idoneity and importance \( r > .80 \), and slightly lower among them and
observability \( r \approx .60 \) for all domains. The highest correlations were found for Self-
determination and the lowest ones for Social inclusion.
TABLE 2. Bangdiwala’s Weighted Agreement Coefficient ($B^W_{kn}$) means.

<table>
<thead>
<tr>
<th>Domains</th>
<th>Idoneity</th>
<th>Importance</th>
<th>Observability</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional well-being</td>
<td>.60</td>
<td>.58</td>
<td>.57</td>
<td>.59</td>
</tr>
<tr>
<td>Interpersonal relations</td>
<td>.57</td>
<td>.58</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>Material well-being</td>
<td>.54</td>
<td>.56</td>
<td>.57</td>
<td>.56</td>
</tr>
<tr>
<td>Personal development</td>
<td>.51</td>
<td>.53</td>
<td>.54</td>
<td>.53</td>
</tr>
<tr>
<td>Physical well-being</td>
<td>.68</td>
<td>.69</td>
<td>.67</td>
<td>.68</td>
</tr>
<tr>
<td>Self-determination</td>
<td>.69</td>
<td>.64</td>
<td>.63</td>
<td>.65</td>
</tr>
<tr>
<td>Social inclusión</td>
<td>.77</td>
<td>.76</td>
<td>.71</td>
<td>.75</td>
</tr>
<tr>
<td>Rights</td>
<td>.67</td>
<td>.69</td>
<td>.61</td>
<td>.65</td>
</tr>
<tr>
<td>$M$</td>
<td>.62</td>
<td>.62</td>
<td>.60</td>
<td>.61</td>
</tr>
</tbody>
</table>

The next step involved calculating item mean scores ($M$) and standard deviations ($SD$) for idoneity, importance and observability. The criterion was to eliminate those items presenting higher standard deviations and lower means. In order to safeguard the content validity, idoneity to assess the domain took priority over importance, and this one took it over observability. This led to the elimination of items presenting $M < 8$ and $SD \geq 2.5$ for idoneity and importance, and $M \leq 6$ and $SD \geq 3$ for observability. Only 20 items (18.86%) did not meet these requirements and so were eliminated ($EW = 5; IR = 3; MW = 5; PD = 4; PW = 2; RI = 1$). Among these 20 items, 15 did not meet the idoneity requirements, 16 did not meet the importance ones, and only 2 did not meet the one referred to observability. In conclusion, the analyses resulted in 86 items (81.13%) that were considered as valid by experts. Among these, the 55 most adequate items ($EW = 8; IR = 8; MW = 8; PD = 6; PW = 6; SD = 8; SI = 3$, and $RI = 8$) were selected not only taking into account their $M$ and $SD$, but also if they were representative of all core indicators and the two system levels in each domain (with the exception of Emotional wellbeing, whose items were only related to microsystem since it is something internal to individuals). Some of them were lightly reformulated following expert suggestions.

As it was determined that each domain should consist of 8 items at least, domains were completed with a selection of new items and indicators proposed by the experts. They proposed a total of 97 items ($EW = 12; IR = 11; MW = 15; PD = 13; PW = 11; SD = 12; SI = 12$, and $RI = 11$) and 17 indicators ($MW = 7; PD = 3; PW = 4; SI = 1; IR = 2$). After an exhaustive review of all them, the most adequate were selected according to the following criteria: a) number of experts proposing the item; b) theoretical evidence about their belonging to the domain; c) evidence about not similar items included in the scale; e) influence on person-centered QOL; and f) organizational opportunities to take part and achieve changes. The result of this review was the selection of 9 new items ($PD = 2; PW = 2; SI = 5$), and 4 indicators: limitations and abilities ($PD$), new technologies ($PD$), diet ($PW$), and accessibility ($SI$). This procedure ended with a total of 64 items (8 items per domain). More extensive information about the selected and eliminated items can be found in Verdugo et al. (2008b).
Step III: Focus groups

Despite of the exhaustive work described before, a step forward in developing the items to be included in field-test version of the scale was taken. As Bonham et al. (2004) and Schalock et al. (2007) suggested, all stakeholder groups should be involved in the selection of indicator items, and the collection and use of the resultant data. For that, the selected items and indicators were evaluated by several focus groups. The specific goals were: a) to confirm the idoneity, importance and observability of the selected items; b) to make sure that any relevant aspect had been forgotten, and c) to improve the item formulations.

A total of five groups were celebrated with a moderator \( n = 5 \) and 7-14 participants each \( n = 54 \). Users of social services, relatives, and stakeholders from Catalonia (Spain) were the participants. Each focus group was referred to one target population: elderly \( n = 10 \), mental health \( n = 12 \), drug addicts and HIV/AIDS \( n = 14 \), intellectual disabilities \( n = 11 \), and physical disabilities \( n = 7 \). The duration of each focus group was about 2 hours.

In order to systematize the procedure and the collected data, a discussion guide including the focus group questions was available for moderators. These questions were: a) «Do you think that the item/indicator have an influence on the person-centered QOL?»; b) «Do you think that this item/indicator is important to this population?»; c) «Do you think that the items to assess this indicator are observable for an external observer?»; and d) «Do you think any important item/indicator to assess this domain is not included?».

The results can be summarized as follows: a) all items were suitable for all populations, consequently all they were retained; b) some items were reformulated; c) clarifications were made to 13 items; and d) and 5 items were added based on the input from focus groups. Concerning to the first result, all items were considered as suitable, important and observable by the different focus groups. Only the observability was discussed for a few items but it was solved with its reformulation. For instance, the item ‘He/she is depressed’ was reformulated in order to be more observable like this ‘He/she shows symptoms of depression’. All the focus groups proposed the reformulation of most of the items or added one or more words to clarify the intent of the item. The goal of all reformulations and clarifications were to try to make the content of the items more specific and guarantee they were suitable for all people. An example of a reformulated item was: ‘He/she has a partner or he/she has the chance to have a partner if wanted’ was reformulated as ‘It is difficult to him/her to start a sentimental relationship’. An example of a clarification for the item was ‘He/she cares about his/her personal hygiene’, was the next: ‘It doesn’t matter if he does by himself/herself or he/she needs help to do it’. Regarding the last result, the 11 items were also suitable for all people. The focus group proposing the highest number of items was the one related to intellectual disabilities \( n = 6 \), followed by the one related to drug dependences and HIV/AIDS \( n = 5 \). The physical disabilities’ group only proposed one item (coinciding with one proposed by intellectual disabilities’ group) and the rest of the group considered that it was necessary to include any item. After an exhaustive review of the 11 items, 5 items \( IR = 2; RI = 2; SD = 1 \) and 1 indicator (‘sexuality’ for \( IR \)) were selected to be part of the field-test version of the scale. The selecting criteria were the same than for panel of experts.
STUDY 2: PSYCHOMETRIC PROPERTIES OF THE CATALONIAN VERSION OF THE SCALE

Study 1 resulted in a field–test version of GENCAT Scale that was composed of 69 items (55 came from the literature review, 9 from the panel of experts, and 5 from the focus groups). The purpose of study 2 was to evaluate the scale on a representative sample of social service users from Catalonia (Spain). The evaluation included reliability and validity. Reliability was evaluated by Cronbach’s alpha index and the standard error of measurement. Validity was evaluated by Confirmatory Factor Analysis. The goodness-of-fit of the eight-dimension model of QOL was compared with the goodness-of-fit of another conceptualization in which QOL is one-dimensional.

Method

Participants

With the purpose of guaranteeing the representativity, a probabilistic polietapic sample design was carried out. For elder people group, sample was selected by geographic areas (sampling error of 2.43 with 95% confidence and \( p = q \)) (see Table 3). If only residence settings were considered, the sampling error was increased to 2.66%, while a sampling error of 6.07% was found for day centres. For the remaining groups, given the found variety in number of centres by geographical areas, the sampling was carried out by target populations. In this case, a fixed and a proportional-to-size number of units were assigned to each group. Sampling errors ranged from 3.99 to 5.51 (see Table 4).

TABLE 3. Sampling design for elder people.

<table>
<thead>
<tr>
<th>Geographic areas</th>
<th>Residence settings</th>
<th>Day Centres</th>
<th>n</th>
<th>Sampling errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metropolitan region</td>
<td>570</td>
<td>89</td>
<td>659</td>
<td>3.75</td>
</tr>
<tr>
<td>Regions from Gerona</td>
<td>231</td>
<td>25</td>
<td>256</td>
<td>5.84</td>
</tr>
<tr>
<td>Field of Tarragona</td>
<td>144</td>
<td>22</td>
<td>166</td>
<td>7.28</td>
</tr>
<tr>
<td>Lands of the Ebre</td>
<td>89</td>
<td>21</td>
<td>110</td>
<td>7.90</td>
</tr>
<tr>
<td>West regions</td>
<td>194</td>
<td>44</td>
<td>238</td>
<td>5.96</td>
</tr>
<tr>
<td>Central regions</td>
<td>85</td>
<td>36</td>
<td>121</td>
<td>8.70</td>
</tr>
<tr>
<td>High Pirineus and Aran</td>
<td>41</td>
<td>28</td>
<td>69</td>
<td>11.32</td>
</tr>
<tr>
<td>Catalonia</td>
<td>1354</td>
<td>265</td>
<td>1619</td>
<td>2.43</td>
</tr>
</tbody>
</table>
TABLE 4. Sampling design for the remaining groups.

<table>
<thead>
<tr>
<th>Geographic areas</th>
<th>PD</th>
<th>ID</th>
<th>MH</th>
<th>DR</th>
<th>HA</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metropolitan region</td>
<td>181</td>
<td>263</td>
<td>188</td>
<td>30</td>
<td>51</td>
<td>713</td>
</tr>
<tr>
<td>Regions from Gerona</td>
<td>62</td>
<td>79</td>
<td>26</td>
<td>29</td>
<td>23</td>
<td>219</td>
</tr>
<tr>
<td>Field of Tarragona</td>
<td>33</td>
<td>56</td>
<td>18</td>
<td>-</td>
<td>7</td>
<td>114</td>
</tr>
<tr>
<td>Lands of the Ebre</td>
<td>4</td>
<td>24</td>
<td>16</td>
<td>-</td>
<td>-</td>
<td>44</td>
</tr>
<tr>
<td>West regions</td>
<td>40</td>
<td>67</td>
<td>15</td>
<td>16</td>
<td>-</td>
<td>138</td>
</tr>
<tr>
<td>Central regions</td>
<td>31</td>
<td>90</td>
<td>46</td>
<td>-</td>
<td>-</td>
<td>167</td>
</tr>
<tr>
<td>High Pirineus and Aran</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>Catalonia</td>
<td>355</td>
<td>586</td>
<td>313</td>
<td>75</td>
<td>81</td>
<td>1410</td>
</tr>
</tbody>
</table>

Sampling errors (%)  

Note. PD = physical disabilities; ID = intellectual disabilities; MH = mental health; DR = drug dependences; HA = HIV/AIDS. *No sampling error was calculated for drug dependences and HIV/AIDS since N was unknown.

A total of 608 professionals working in 239 centres of social services participated filling in the field-test version of the GENCAT Scale for 3,029 users from Catalonia. The mean number of evaluated persons per service was 12.67 (SD = 7.75) and the mean number of evaluated persons per professionals was five (M = 4.98). The requirements for professionals to participate in this study were: (a) to be working in some kind of social service for handicapped adults; and (b) to have been working directly with the client for at least 3 months. The only requirement to apply the scale to a social service user was that this was older than 16 years old.

Related to the main socio-demographic characteristics of professionals, most of them were female (85%), had been working with the client for more than 2 years (55.74%), were psychologists (23.01%) and social workers (18.41%), and had been working in social services more than 5 years (52.80%). Concerning to the social service users, 55.7% were female. Their ages ranged between 16 and 105 (M = 64.72; SD = 21.34). More than half of sample (57.57%; n = 1,711) was older than 60. Actually, the biggest group (n = 791) was composed of 81-90 years old people and only 17.39% (n = 515) were younger than 41. Concerning to people condition (see Table 5), the most representative group was the one composed of elder people living in residence settings (44.70%), followed by people with intellectual disabilities (19.35%), physical disabilities (11.72%), mental health (10.33%), and old people in day centres (8.75%). Percents of people with drug dependences and HIV/AIDS ranged from 2.48 to 2.67%.
Instruments

Since the GENCAT Scale application was going to be carried out in Catalonia, the field-test version needed to be translated to Catalan. Along the entire process of development, the International Quality of Life Assessment Project (IQOLA) and the International Test Commission (ITC) (Bartram, 2001; Hambleton, 2001; Muñiz and Hambleton, 1996; Van de Vyjver and Hambleton, 1996) guidelines on adapting tests were followed. It must be noted that this case is not exactly a test adaptation since the scale was developed from the beginning to be used with Catalan population. In fact, as we noted earlier, all items were validated by professionals, users and relatives from Catalonia. Nevertheless, a translation methodology meeting all the requirements for the best practice was applied: two translations, two back-translations, an expert committee and a pretest.

Translation. The first step was to produce two translations from Spanish to Catalonian, using two qualified and bilingual translators whose mother tongue was Catalan. Translations are of higher quality when undertaken by at least two independent translators since allows for the detection of errors and divergent interpretation of ambiguous items. One translator was aware of the objectives underlying the GENCAT Scale and the quality of life concepts, whereas the second was unaware of these objectives and concepts. The two translations were compared in order to find discrepancies and identify the more ambiguous words. Working from the initial version together with the first translator’s version (T-A) and the second translator’s version (T-B), the synthesis resulted in a common translation (T-AB).

Back-translation. Translating back from the final language into the source language has been shown to help improve the quality of the final version of an assessment instrument. From the T-AB version, two translators who were fluent in the idioms and colloquial forms of Catalonian translated this version into their mother tongue: Spanish. None of the translators was aware of the intent and content of the material. Both versions were synthesized (BT-AB). This version (BT-AB) was compared with the initial one in order to detect differences and reveal unexpected meanings or interpretations.

Expert committee. An expert committee was constituted in order to produce a final version of the GENCAT scale based on the various translations and back-translations.
Membership was multidisciplinary: a methodologist, two experts on quality of life, two experts on languages, the two translators and both back-translators. Each committee member reviewed the instructions, the items and the answer options. Committee member aimed for semantic, idiomatic, and conceptual equivalence. There were no difficulties in obtaining the consensus about the final version since all the items were developed in reference to Catalonian users and reformulated by persons from Catalonia. In addition, all items were very specific and clear. For these reasons, all versions were very similar and, once the equivalence was guaranteed, the synthesis of the two versions that were translated into Catalonian (without any modification) constituted the final Catalonian version.

Pre-test. The original and the final version were administered to two bilingual individuals. As discrepancies were not found between them, the translating process was concluded in order to administer the developed version and check its psychometric properties in a Catalonian sample.

This is a self-administered questionnaire in which professionals must answer objective and observable questions about user QOL. Administration time varies from 8 and 13 minutes. It is composed of eight scales -that correspond to the eight QOL-domains- and 69 items ($EW = 8; IR = 10; MW = 8; PD = 8; PW = 8; SD = 9; SI = 8; RI = 10$). All items are formulated as third person declarative statements and random organized by domains. Half items have positive ($n = 35$) and half have negative ($n = 34$) valence. The answer format is a frequency scale with 4 options (Never o hardly ever, Sometimes, Often, Always or almost always); however, in case that using this answer format turned out difficult, a 4-points Likert scale (Totally disagree, Disagree, Agree, and Totally agree) can be used instead. The instrument offers a direct score for each QOL domain that are converted to standard scores ($M = 10; SD = 3$) and to percentiles. It is also possible to illustrate all scores on a QOL profile that makes score interpretation easier. Finally, a QOL Index is also provided.

Procedure

Once the participant services were selected, a letter was sent in order to explain the researching goals and make the participation request. This letter was sent by post and by email. After that, the research team phoned every single service with the following goals: a) to confirm the letter reception; b) to confirm the post address since the scales were sent by a courier company; c) to give detailed information about the research project; and d) to confirm if they were determined to collaborate. When some service rejected to participate, it was replaced by other one with similar characteristics. Once their participation was confirmed, the specific number of scales they must complete plus 5 (to be sure of achieving a big enough sample) and an evaluator’s guide were sent. About 4,500 scales were sent. We must point out here that the Catalonian version of the GENCAT Scale was applied to all participants in this study. Phone and email contact was kept all along the process in case there were doubts or suggestions. Phone was also used to remind the deadlines when it was necessary.
Results

Reliability

Reliability was evaluated by Cronbach’s alpha index ($\alpha$) and the standard error of measurement ($SEM$) (see Table 6). The internal consistency indexes ($\alpha$) fluctuated between .47 and .88. The lowest coefficient was reported by the Physical Wellbeing domain; however, this result is congruent with findings in other similar studies (Gómez et al., 2008; Verdugo et al., in press). The internal consistency of the 69 items was quite high ($\alpha = .92$), whereas $SEM$ values were not too high (1.87 - 2.75).

<table>
<thead>
<tr>
<th>Domains</th>
<th>$\alpha$</th>
<th>$SEM$</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional well-being</td>
<td>.83</td>
<td>2.03</td>
<td>8</td>
</tr>
<tr>
<td>Interpersonal relations</td>
<td>.66</td>
<td>2.75</td>
<td>10</td>
</tr>
<tr>
<td>Material well-being</td>
<td>.57</td>
<td>1.87</td>
<td>8</td>
</tr>
<tr>
<td>Personal development</td>
<td>.74</td>
<td>2.39</td>
<td>8</td>
</tr>
<tr>
<td>Physical well-being</td>
<td>.47</td>
<td>2.05</td>
<td>8</td>
</tr>
<tr>
<td>Self-determination</td>
<td>.88</td>
<td>2.41</td>
<td>9</td>
</tr>
<tr>
<td>Social inclusion</td>
<td>.57</td>
<td>2.52</td>
<td>8</td>
</tr>
<tr>
<td>Rights</td>
<td>.69</td>
<td>2.18</td>
<td>10</td>
</tr>
<tr>
<td>GENCAT</td>
<td>.92</td>
<td>6.92</td>
<td>69</td>
</tr>
</tbody>
</table>

Construct validity

A Confirmatory Factor Analysis (CFA) was used to evaluate the goodness-of-fit for two measurement models. Model I contained the eight-factors proposed by Schalock and Verdugo (2002): Emotional wellbeing, Material wellbeing, Physical wellbeing, Social inclusion, Interpersonal relations, Rights, Self-determination, and Personal development. Model II was one-dimensional. Given the nature of the data, the CFA was performed implementing DWLS estimation method with the covariance and asymptotic covariance matrices. LISREL 8.8 (Scientific Software International, 2006) was the software used. As observed variables for each QOL factor, 4 parcels composed of the corresponding and random selected item means were used (Bandalos, 2002; Bandalos and Finney, 2001; Brown, 2006; Hall, Snell, and Singer Foust, 1999; Kishton and Widaman, 1994; Little, Cunningham, Shahar, and Widaman, 2002; West, Finch, and Curran, 1995). In this way, a total of 32 parcels (4 parcels per domain) were analyzed. Previously to CFA, one-dimensionality of each parcel was guaranteed.

To address the limitations of Chi-square test ($\chi^2$), several goodness-of-fit indexes (McDonald and Ho, 2002) were used to evaluate the goodness-of-fit of eight vs. one-factor models. Absolute fit indexes, such as the Goodness of Fit Index ($GFI$), directly assess how well a priori model reproduces the sample data; whereas relative fit indexes, such as Root Mean Square Error of Approximation ($RMSEA$), the Normed Fit Index ($NFI$), the Tucker-Lewis Index (TLI), and the Comparative Fix Index (CFI) measure the proportionate improvement in fit by comparing a target model with a more restricted nested baseline model (a null model in which all the observed variables are uncorrelated). Hu and Bentler (1999) identified cut off criteria for indexes. The $GFI$, the $NFI$, the $TLI$, $RMSEA$, $NFI$, $TLI$, and $CFI$.
and the $CFI$ should be above .95 for having a good fit to the data. For the $RMSEA$, a cut off value of less than .06 was suggested. As it is shown in Table 7, the goodness-of-fit indexes of the eight-factor model were acceptable (with the exception of $\chi^2/df$ that was slightly above 2). Anyway, the eight-factor solution had much better indexes than the one-dimension model, which showed a decrease of the fit indexes. The goodness-of-fit for the eight-factor model suggested that this model provided a better fit to the data than the other solution (see Figures 1 and 2).

**TABLE 7.** Goodness-of-fit indexes of the eight vs. one-factor models.

<table>
<thead>
<tr>
<th>Goodness-of-fit Indexes</th>
<th>Eight-factor model</th>
<th>One-factor model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
<td>436</td>
<td>464</td>
</tr>
<tr>
<td>S-B $\chi^2$</td>
<td>1251.16</td>
<td>5974.82</td>
</tr>
<tr>
<td>$\chi^2/df$</td>
<td>2.87</td>
<td>12.88</td>
</tr>
<tr>
<td>GFI</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td>NFI</td>
<td>.95</td>
<td>.76</td>
</tr>
<tr>
<td>TLI</td>
<td>.96</td>
<td>.75</td>
</tr>
<tr>
<td>CFI</td>
<td>.97</td>
<td>.77</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>$p$</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

**FIGURE 1.** Eight-factor measurement model of the GENCAT Scale.
Finally, regarding to nomological validity, correlation coefficients were calculated for all the latent variables (see Figure 1). The highest coefficient \( r = .93 \) was observed between Interpersonal relations and Social inclusion, followed by Self determination and Rights \( r = .82 \), Interpersonal relations and Emotional wellbeing \( r = .77 \), Social inclusion and Personal development \( r = .77 \), and Self determination and Personal development \( r = .75 \). The lowest coefficients were observed for Material wellbeing and Self determination \( r = .07 \), Self determination and Physical wellbeing \( r = .24 \), Personal development and Material wellbeing \( r = .25 \), and Rights and Emotional wellbeing \( r = .29 \). Since they are congruent with the outcomes of previous research (Bonham et al., 2004), the aforementioned correlations can be considered as an evidence of the scale’s nomological validity.

Discussion

The present study had two goals. Firstly, it aimed to develop an instrument to assess QOL of social service clients in an objective way and based on a multidimensional model (Schalock, Bonham, and Verdugo, 2008; Schalock and Verdugo, 2002). Based on it, it was expected that the GENCAT Scale measured QOL including Emotional wellbeing, Material wellbeing, Physical wellbeing, Personal development, Interpersonal relations, Self-determination, Rights, and Social inclusion. Secondly, it aimed to study the reliability and construct validity of the developed scale.

Regarding reliability, the GENCAT Scale shows a good internal consistency \( \alpha = .92 \), and a low standard error of measurement \( SEM = 6.92 \). Analyzing by domains, Self-
determination, Emotional well-being, and Personal development, with values higher than .70, are the most reliable; whereas Rights, Interpersonal relations, Material well-being, and Social inclusion are very close to .60. Finally, Physical wellbeing has the lowest internal consistency (α = .47). Since Physical wellbeing consists of eight items as other factors, its low level of internal consistency cannot be explained by the number of items. A more detailed analysis of the content of the items belonging to this domain found that there are three items with very low corrected homogeneity indexes and whose content seem to be slightly different to the other items within the domain. Although they undoubtedly contribute to Physical wellbeing, they also may assess other domains such as Material wellbeing or Rights. Given the strong consensus reached by the experts and the focus groups about their importance during the process of development of the scale, the items were retained. However, further studies should be done to determine its importance and reliability to assess this or other domain.

Concerning to construct validity, the aim is to demonstrate that indicators (items, or parcels in this study) load on only one factor (or domain) and are not influenced by any other systematic effect. To that end, two models were compared. The first one (Model I) approaches QOL as a one-domain construct and the second model (Model II) conceives QOL as composed of eight correlated domains. The Confirmatory Factor Analysis supported the eight-domain model is supported by the data, in comparison with the one-dimension solution. Also the scale showed an adequate nomological validity, although deeper research about this issue is recommended.

The sampling design guarantees the results’ representativeness and allows concluding that they can be generalized to users of social services in Catalonia. Besides, since each target-population involved in the development and validation of the scale shows a similar distribution all around Spain, the authors, other researchers and professionals of social services consider the GENCAT Scale as suitable to be applied to clients from other counties in Spain. In order to check it, it is being applied in other Spanish counties at present. Anyway, the study has also some limitations. For example, as the instrument was developed to assess QOL in an objective way, it is assumed that QOL scores should be invariant depending on the observer. Although this issue was not studied, we plan to do it after applying the scale in other areas.

The originality, importance and utility of this instrument reside on several aspects. Firstly, there is only one QOL questionnaire serving to the same goal, the FUMA T Scale (Gómez et al., 2008), but it was developed after this one and it is focused on elderly and disabilities. The GENCAT Scale, in contrast, can be applied to all kind of users of social and human services, including people with disabilities and elder people, but also people with mental health problems or AIDS/HIV. And secondly, it responds to all basic principles of conceptualization, measurement and application of QOL (Schalock, 2005, 2006; Schalock, Bonham, and Verdugo, 2008; Schalock and Verdugo, 2007; Schalock, Verdugo, et al., 2008; Verdugo, 2006; Verdugo, Schalock, Keith, and Stancliffe, 2005). In conclusion, the GENCAT Scale is in consonance with conceptual advances in the field and it constitutes a useful tool for organizations, entities and services interested in improving the assessed QOL of their service recipients, and for social service professionals that are concerned about improving their programs and practices.
References


Int J Clin Health Psychol, Vol. 10. Nº 1

VERDUGO et al. Objective instrument to assess quality of life


Received September 17, 2008
Accepted April 24, 2009