Abstract

Semantic memory tests assess long-term memory for facts, objects, and concepts as well as words and their meaning. Since it holds culturally shared information, the development of normative data adjusted to the cultural and linguistic reality of the target population is of particular importance. The present study aimed to establish normative data for the Pyramids and Palm Trees Test, a commonly used test of semantic memory, in the French-Quebec population. The normative sample consisted of 214 healthy French-speaking adults and elderly persons from various regions of the province of Quebec. The effects of participants’ age, gender, and education level on test performance were assessed. Results indicated that participants’ level of education and age, but not sex, were found to be significantly associated with performance on this test. Normative data are presented as means and standard deviations. Overall, the present norms are consistent with those of previous studies with Spanish samples.

Keywords: Neuropsychological tests; Adult; Normative data; Percentiles; Semantic memory; French Canadian

Introduction

Semantic memory is a long-term memory system for facts, objects, and concepts as well as words and their meaning (Tulving, 1972). In contrast to other types of memory, semantic memory is culturally shared and is not temporally or spatially specific. It enables us to interpret and act on sensory perceptions and experiences. Therefore, semantic memory occupies a central role in cognitive processes, such as language production and comprehension, object and environmental sound recognition, and object use, among others. Because of this central role, damage to this type of memory system may lead to various cognitive deficits (e.g., Buxbaum, Schwartz, & Carew, 1997; Patterson, Nestor, & Rogers, 2007; Warrington, 1975). In clinical and research settings, semantic memory is usually assessed with tasks requiring access to semantic representations with various levels in depth of processing. These tasks include, among others, picture naming, defining spoken or written words, word–picture and picture–picture matching, generation of exemplars on category fluency tests, and semantic similarity judgment.

The Pyramids and Palm Trees Test (PPTT; Howard & Patterson, 1992) is a commonly used test of associative semantic knowledge comprising 52 items. In this test, participants are shown a stimulus (e.g., a pyramid) and asked to match it to either of two images: A target (e.g., a palm tree) or a distracter (e.g., a fir tree). Subjects are required to use explicit semantic information of encyclopedic (e.g., pyramid: Palm tree vs. fir tree) or functional (e.g., wood: Saw vs. hammer) nature to select
the correct match. According to theoretical models of object recognition and naming (e.g., Humphreys & Riddoch, 2006), the PPTT requires that subjects (a) recognize each of the items presented and (b) retrieve, activate, and compare the corresponding semantic information in order to identify semantic associates. The test is available in both word and picture modalities to assess verbal and non-verbal semantic memory, respectively. The three-picture modality is, however, the most commonly used by clinicians and researchers. It was initially normalized in England with a group of 60 healthy English-speaking subjects. Participants’ performance was very consistent and strong. None of the subjects made more than three errors and the authors concluded that a subject scoring 90% or better on the PPTT does not have clinically significant impairment in this task; consequently, researchers and clinicians generally consider that a score lower than 47/52 on this test indicates significant semantic memory deficit. Unfortunately, although the use of such cut-off scores may be useful, this technique fails to take into account variables likely to influence semantic memory performance, such as age and formal education level. Indeed, it is generally admitted that these variables may influence cognitive performance. In the case of semantic memory, age is typically shown to be inversely associated with performance, with younger adults performing better than older adults (e.g., Brickman et al., 2000), although this finding has been disputed by certain researchers (Perlmuter & Tun, 1987). In contrast, education level generally appears to be positively correlated with test performance, with more highly educated individuals obtaining better results on tests of semantic memory than those with basic education (Da Silva, Petersson, Faisca, Ingvar, & Reis, 2004). Owing to the influence that age and education level may exert on semantic memory performance, it seems pertinent to develop normative data which take at least these variables into consideration.

In addition, tests of semantic memory often include items that are culture-specific. The PPTT includes several such items. For instance, the association of a windmill to a tulip is typical to Holland (item 16). Such items may be problematic for individuals not familiar with certain cultures, making it essential that stimuli presented in tests of semantic memory be adapted to the cultural and linguistic reality of the target population.

Recently, normative data were made available for the PPTT in Spanish-speaking populations (Gudayol-Ferre et al., 2008; Rami et al., 2008). In both studies, education level was found to be the only significant predictor of performance on the PPTT, although in one study (Gudayol-Ferre et al., 2008) this effect was modulated by age.

To our knowledge, normative data for the PPTT exist only in English (Howard & Patterson, 1992) and in Spanish (Gudayol-Ferre et al., 2008; Rami et al., 2008). Normative data for this test are not yet available for French Quebecers, despite the fact that it is used among numerous French-language clinical and research groups in Quebec. The aim of the present study was to establish normative data for the PPTT, adjusted to the linguistic and cultural reality of the French-Quebec population.

Materials and Methods

Participants

Two hundred and fourteen healthy French-speaking adults, whose mother tongue and usual language was French, were recruited via public advertisements and among the relatives of patients recruited for research projects. The province of Quebec is officially divided into 17 administrative regions. In order to control for geographic variation, subjects were recruited from 15 of these urban and rural regions (no participants from Abitibi-Témiscamingue or from Nord-du-Québec). The majority of subjects were recruited from the regions Capitale-Nationale (37.38%) and Estrie (28.50%). A self-reported medical and psychiatric history was obtained from each participant. Any person with a history of neurological disease, psychiatric illness, head injury, or stroke was excluded. All participants were screened using standard neuropsychological tests (e.g., Montreal Cognitive Assessment, Nasreddine et al., 2005; Mini-Mental State Examination, Folstein, Folstein & McHugh, 1975) to rule out the possibility of cognitive impairment. All participants scored within normal ranges. In some cases, participants received financial compensation to cover travel expenses. The sample was composed of 77 men (36%) and 137 women (64%). In addition to sex, we considered age and education level as demographic variables that may have influenced test performance. There were no significant differences between men and women in age (men: M = 53.8; SD = 16.21; women: M = 54.97; SD = 18.76; t = 0.44, p = .66) or years of education (men: M = 12.90; SD = 3.24; women: M = 12.31; SD = 3.36; t = −1.25, p = .21). Our sample had an overall overrepresentation of highly educated individuals compared with actual Quebec demographics (Institut de la statistique du Québec, 2003; Table 1).

Materials and Procedure

The three-picture modality of the PPTT was administered to all subjects. Each page in the test booklet depicted the stimulus centered at the top of the page. Presented directly below the stimulus were two additional images, the target and a distracter.
Subjects were asked to decide which of the additional items could best be matched to the stimulus. There was no time limit to complete the test. The maximum possible score was 52.

Statistical Analyses

To identify the best combination of sociodemographic variables influencing the performance on the PPTT, a standard regression analysis was performed with age, sex, and education as predictors. Years of age and years of education were entered as continuous variables in the analysis. Women and men were, respectively, coded as 0 and 1. Since age and education were significantly correlated ($r = -0.33; p < 0.001$), we also included an age x education interaction as a predictor. To reduce multicollinearity, the interaction was calculated with “centered” variables, each value being replaced with its difference from the mean. Visual and statistical analyses of the residuals were conducted to verify the underlying assumptions of the regression model. As recommended by Tabachnick and Fidell (2007), we examined the shape of the distribution rather than using formal normality inference tests, which are not adequate for large samples. Since the distribution of PPTT scores was slightly skewed (Fig. 1), a reflect and square-root transformation ($\sqrt{(53-score)}$) was applied for the regression (Tabachnick & Fidell, 2007). Then, we verified that the residuals from the regression analyses had skewness and kurtosis values below 1 in absolute value.

Qualitative inspection of the data revealed that some items of the PPTT tended to have particularly high error rates. Items with error rates above 25% were identified. To assess the relationship between these items and test performance, a point-biserial correlation was performed between each of these items and residual total PPTT score. All statistical analyses were performed using the SPSS v.13.0 package for Windows.

Results

The mean PPTT score was 49.30 ($SD = 1.98$) and the median was 50.00. The transformed mean PPTT was 1.86 ($SD = 0.50$). Results of the regression analysis revealed a modest ($R^2 = 0.05$) but significant association between participants’

![Fig. 1. Frequency distribution of PPTT scores for all participants.](image-url)
characteristics and PPTT score, \( F(4,209) = 2.57, p < .05 \). According to the standardized \( \beta \) (Table 2), education had the strongest influence on the PPTT score, although neither sex, age, education, or the interaction age \( \times \) education were individually statistically significant.

Since age \((r = .13; p < .05)\) and education \((r = -.19; p < .05)\), but not sex \((r = -.10; p = .08)\), were significantly correlated with the transformed PPTT score, normative data were stratified according to age and education level. Five age \((19–39, 40–49, 50–59, 60–69, \text{and} 70 \text{ or more years})\) and two educational \((0–12 \text{ and} 13 \text{ or more years})\) clusters were formed. Data for each cluster are presented in Table 3 and include statistical cut-off scores set at 5% (equivalent to a 40–49, 50–59, 60–69, and 70 or more years) and two educational (0–12 and 13 or more years) clusters were formed. Data for each cluster are presented in Table 3 and include statistical cut-off scores set at 5% (equivalent to a Z-score of \( -1.65 \)), which is the generally accepted statistical cut-off score to determine pathological performance in comparison to healthy adults.

PPTT items 12 (pyramid—palm tree), 16 (windmill—tulip), and 40 (acorn—pig) had particularly high error rates (26.2%, 34.6%, and 54.2%, respectively) compared with the mean error rate of the 49 other items (3.19%). The point-biserial analysis revealed that only item 16 was significantly, albeit weakly, correlated with overall test performance (\( r = .16, p = .02 \)), whereas items 12 \((r = .05, p = .47)\) and 40 \((r = .07, p = .33)\) were not. Because the point-biserial correlation for item 16 was small, based on Cohen’s (1988) criteria, all three items were excluded from the data set and the regression analysis was repeated. The distribution of PPTT scores was again transformed for the analysis \((\sqrt{50}\text{-score})\). When the regression analysis was performed without these three items, the model was significant, \( F(4,209) = 2.97, p < .05 \). Once more, no predictor was individually significant and the same pattern of correlations was observed between predictors and transformed PPTT scores (Table 2). Therefore, Table 4 presents means and standard deviations for PPTT scores excluding items 12, 16, and 40. The scores are

### Table 2. Regression coefficients for the standard regression analyses

<table>
<thead>
<tr>
<th></th>
<th>All test items</th>
<th>Excluding items 12, 16, and 40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard coefficient (( \beta ))</td>
<td>( t )</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.08</td>
<td>-1.14</td>
</tr>
<tr>
<td>Age</td>
<td>0.07</td>
<td>1.03</td>
</tr>
<tr>
<td>Education level</td>
<td>-0.15</td>
<td>-1.90</td>
</tr>
<tr>
<td>Education ( \times ) age</td>
<td>-0.03</td>
<td>-0.34</td>
</tr>
</tbody>
</table>

Note: Residual mean square, all test items: 0.24; excluding items 12, 16, and 40: 0.22.

### Table 3. Norms for PPTT stratified for age (19–39, 40–49, 50–59, 60–69, and 70+ years) and years of education (0–12 and 13+)

<table>
<thead>
<tr>
<th>Age (Years)</th>
<th>0–12 years of education</th>
<th>13+ years of education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19–39 ((n = 19))</td>
<td>40–49 ((n = 12))</td>
</tr>
<tr>
<td></td>
<td>19–39 ((n = 36))</td>
<td>40–49 ((n = 22))</td>
</tr>
<tr>
<td>( M )</td>
<td>49.89</td>
<td>49.42</td>
</tr>
<tr>
<td>( SD )</td>
<td>1.10</td>
<td>1.83</td>
</tr>
<tr>
<td>5% cut-off score</td>
<td>48</td>
<td>46</td>
</tr>
</tbody>
</table>

Notes: \( M \) = mean; \( SD \) = standard deviation. The 5% cut-off score corresponds to a Z-score of \( -1.65 \).

### Table 4. Norms for PPTT stratified for age (19–39, 40–49, 50–59, 60–69, and 70+ years) and years of education (0–12 and 13+), excluding test items 12, 16, and 40

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>0–12 years of education</th>
<th>13+ years of education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19–39 ((n = 19))</td>
<td>40–49 ((n = 12))</td>
</tr>
<tr>
<td></td>
<td>19–39 ((n = 36))</td>
<td>40–49 ((n = 22))</td>
</tr>
<tr>
<td>( M )</td>
<td>48.00</td>
<td>47.17</td>
</tr>
<tr>
<td>( SD )</td>
<td>0.75</td>
<td>1.47</td>
</tr>
<tr>
<td>5% cut-off score</td>
<td>46</td>
<td>44</td>
</tr>
</tbody>
</table>

Notes: \( M \) = mean; \( SD \) = standard deviation. The 5% cut-off score corresponds to a Z-score of \( -1.65 \).
clustered again by age and education level and include cut-off scores set at 5% (Z-score of −1.65) to determine pathological performance.

Discussion

The aim of the present study was to establish normative data for the PPTT for French Quebecers. In sum, participants’ level of education and age, but not sex, were found to be significantly associated with performance on this test. Moreover, age was found to be moderately correlated with the level of education.

The results reported here replicate those of previous studies reporting normative data for the PPTT in Spanish (Gudayol-Ferre et al., 2008; Rami et al., 2008), in which education level was found to be a significant moderator of PPTT score. Our findings are particularly consistent with those of Gudayol-Ferre and colleagues (2008), who reported that although only education contributed significantly to their statistical model, this variable correlated highly with age.

These results are somewhat supported by current scientific literature, which shows that younger individuals (Brickman et al., 2000) as well as those with high levels of formal education (Da Silva et al., 2004) perform better on tests of semantic memory. This is not surprising, considering that semantic memory contains knowledge that is generally acquired in school, such as historical, political and geographical facts, as well as word meanings (Tulving, 1972).

An original aspect of the present study was to consider the error rates of individual test items. Items 12, 16, and 40 were found to have higher error rates than all other test items. An item-total analysis for each of these items revealed a significant but weak correlation between item 16 and overall performance and non-significant correlations for items 12 and 40. This suggests that these three items are likely not representative measures of semantic memory performance within this sample, compared with the 49 other items on the PPTT. It is interesting to note that these three items require semantic knowledge that is highly culturally specific (for item 12, the association of a pyramid with a palm tree is typical to Egypt; for item 16, the association of a windmill to a tulip is typical to Holland; for item 40 the association of an acorn to a pig is typical to France and Europe), thus emphasizing the importance of establishing normative data adapted to the cultural and linguistic reality of the target population. In light of these data, it will be important for Quebec clinicians to exercise caution in interpreting results for these three items on the PPTT. Furthermore, it seems appropriate to consider the possibility of replacing these items by other, more culturally representative items in a PPTT adapted to French-Quebec individuals. The development of new test items was outside the scope of the present study, but future research should certainly address this consideration.

Another original aspect of this work was to include individual cut-off scores for each demographic cluster, set at 5%. These scores may be used to identify a pathological semantic memory deficit. Unlike the arbitrary cut-off (90%) originally proposed by Howard and Patterson (1992), the scores presented here take into account subjects’ age and level of education and are therefore more likely to represent participants’ actual performance.

One limitation to the present study was the overrepresentation of women, who represented 64% of the total sample. This may have reduced the sensitivity of statistical tests in detecting a significant effect of sex. Nonetheless, sex differences have generally not been reported in tests of semantic memory, with the exception of differences in category fluency (Capitani, Laiacoma, & Barbarotto, 1999; Laws, 1999; McKenna & Parry, 1994). A second limitation is the relatively small number of participants in each age–education cluster. However, the results reported here remain consistent with those reported in the scientific literature, making it probable that they are representative of the French-Quebec population. A final possible limitation to the present study is the use of an incidental sampling method. Ideally, a random sampling method would have been preferable because it would have allowed each Quebecker an equal chance at participating in this study, maximizing the representativeness of the sample. Our sampling method resulted in an overrepresentation of highly educated individuals; nonetheless, considering the absence of existing norms, we believe that our sampling method is, at the very least, a practical starting point in the establishment of PPTT norms for this population.

It is important to note that the PPTT may tap into several aspects of cognition in addition to semantic memory, such as visual and reasoning abilities. As such, although the PPTT is generally recognized by researchers and clinicians as a valid measure of semantic memory, performance on this test has been found to be only moderately associated with performance on traditional tests of semantic memory, such as fluency (Klein & Buchanan, 2009). It is possible, therefore, that impairments on this test may reflect deficits in other cognitive domains rather than in semantic memory alone; this possibility should be considered by researchers and clinicians alike when interpreting PPTT scores.

To our knowledge, the present study is the first to establish normative data for the PPTT in the French-Quebec population. This test is frequently used in clinics as well as in research. However, with one exception (Klein & Buchanan, 2009), no studies have investigated the psychometric properties of the PPTT. In a recent study conducted with American college students, Klein and Buchanan (2009) showed that although the PPTT has good discriminant validity, the instrument displays relatively poor test–retest reliability as well as poor convergent validity. Further research should explore the reliability and the validity of the
measure more widely and in different populations. Moreover, as already pointed out by Gudayol-Ferre and colleagues (2008), the ceiling effect observed in the performance of healthy Quebecers on the PPTT directly addresses the question of the instrument’s sensitivity to detect mild semantic memory impairments. As such, further studies are needed.

**Funding**

This work was funded by the Réseau Québécois de recherche sur le vieillissement. In addition, BLC is supported by a Canadian Institutes of Health Research Master’s Award. CH is supported by a Chercheur-boursier Award from the Fonds de la recherche en santé du Québec. NB was supported by the Canadian Institutes of Health Research at the time of data collection. OP is supported by a postdoctoral trainee award from the Fonds de recherche en santé du Québec and Alzheimer Society of Canada Partnership Program.

**Conflict of Interest**

None declared.

**References**


