To cite this article: Wim Van der Elst, Helen Reed & Jelle Jolles (2013) The Logical Grammatical Structures Test: Psychometric Properties and Normative Data in Dutch-Speaking Children and Adolescents, The Clinical Neuropsychologist, 27:3, 396-409, DOI: 10.1080/13854046.2012.745610

To link to this article: http://dx.doi.org/10.1080/13854046.2012.745610

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
The Logical Grammatical Structures Test: Psychometric Properties and Normative Data in Dutch-Speaking Children and Adolescents

Wim Van der Elst, Helen Reed, and Jelle Jolles
LEARN! Research Institute and Faculty of Psychology & Education, VU Universiteit Amsterdam, Amsterdam, The Netherlands

Logical grammatical structures comprehension is the ability to understand the relations between objects, actions, and qualities in spoken and written sentences. The Logical Grammatical Structures Test (LGST) was especially devised to assess these abilities in children and adolescents. In the present study the LGST was administered to 405 healthy Dutch children and adolescents. The aims of the present study were (i) to evaluate the psychometric properties of the LGST (using an Item Response Theory framework), and (ii) to establish demographically corrected normative data. The results showed that there was a strong curvilinear relationship between age and LGST performance, i.e., the relative improvement in ability level was much more pronounced for younger children (aged below 14 years) than for older children (aged above 14 years). Level of parental education was positively associated with the LGST performance. Normative data that took the relevant demographic variables into account were established, and it was shown that the LGST had sound psychometric properties.

Keywords: Language development; Educational differences; Item Response Theory; Normative data.

INTRODUCTION

Language development has been extensively studied in children and adolescents. For example, it has been established that young children show the first signs of word comprehension when they are 8–10 months old, produce their first words when they are 11–13 months old, use word combinations by the age of about 2 years, and produce multiclause sentences by the age of about 4–5 years (Bates, Thal, Finlay, & Clancy, 2003; Hoff, 2009). The basics of language development are thus accomplished by the age of about 4–5 years, but this does not mean that language development stops at this point (Hoff, 2009). For example, children continue to improve their definitional skills and their knowledge of morphologically complex words (Benelli, Belacchi, Gini, & Lucangeli, 2006; Marinellie & Johnson, 2003; Nippold & Sun, 2008), and language development in adolescence is further characterized by important syntactic attainments (e.g., the ability to use the passive voice) and by gradual increases in the use of subordinate clauses and clausal density (Berman, 2004). For a detailed overview of the development of language abilities in childhood and adolescence, see Nippold (2007) and Hoff (2009).
Another important aspect of language development in childhood and adolescence relates to the understanding of so-called “logical grammatical structures”; i.e., the ability to understand the relations between objects, actors, actions, and qualities in spoken and written sentences (Luria, 1980). Understanding a logical grammatical structure such as “Nick was struck by Pete: Who was the victim?” is a complex process that not only requires the understanding of the individual words and actions in the sentence, but also depends on various non-language related higher-order cognitive abilities. These cognitive abilities include the inhibition of prepotent responses, cognitive flexibility, and the simultaneous analysis and synthesis of verbal material (Christensen, 1993; Colman, Koerts, Stowe, Leenders, & Bastiaanse, 2011; Golden, Purisch, & Hammel, 1991; Lee, Grossman, Morris, Stern, & Hurtig, 2003; Luria, 1966; Wassenberg, 2007).

Various standardized tests have been developed to assess the productive and the receptive language abilities of children and adolescents (such as the Clinical Evaluation of Language Fundamentals and the Peabody Picture Vocabulary Test; Dunn & Dunn, 2007; Semel, Wiig, & Secord, 2003), but none of these instruments specifically assesses logical grammatical structures comprehension. Yet it would be advantageous to have a brief, psychometrically sound instrument available which allows for assessing these abilities. Such an instrument would be useful in diagnostic settings, because a poor logical grammatical structure comprehension ability is associated with different neuropsychological and cognitive problems (e.g., Traumatic Brain Injury, learning disabilities, and poor school performance; Cook, Murdoch, Cahill, & Whelan, 2004; Facon, Magis, & Courbois, 2012; Lerner, Lowenthal, & Egan, 2002; Wassenberg, 2007). Such an assessment tool would also be useful in research settings, e.g., to evaluate the impact of growth spurts in the brain maturation of a child on improvements in his or her logical grammatical structure comprehension abilities.

The Logical Grammatical Structures Test (LGST) was especially devised to assess the logical grammatical structure comprehension abilities of children and adolescents. The LGST is based on the Receptive speech scale of the Luria-Nebraska Neuropsychological Battery (Christensen, 1993; Golden et al., 1991; Luria, 1980). The items of this scale were originally developed to assess logical grammatical structures comprehension in brain-damaged adults. Of the 24 items of the original scale, 13 were administered. Items that were too easy for normally developing school-aged children and adolescents (e.g., pointing to objects) were not included in the LGST, and some of the original items were modified to make them more appropriate for administration in children and adolescents (Wassenberg, 2007). For example, the original item “If I had lunch after I had cleaned up the house, what did I do first?” was modified into “I had breakfast after I had walked the dog. What did I do first?”.

In the present study the LGST was administered to a large sample of cognitively intact Dutch-speaking children and adolescents ($N = 405$, age range 5.57–15.87 years). The main aims of the present study were (i) to evaluate the psychometric properties of the LGST, and (ii) to establish normative data for the LGST. Normative data are useful tools in both research and in clinical settings, because they allow for the quantification of the relative level of logical grammatical structures comprehension of a tested child (i.e., what is the level of logical
grammatical structures comprehension of this child as compared to his or her
demographically matched peers?; Capitani, 1997; Mitrushina, Boone, Razani, &
D’Elia, 2005; Van der Elst, 2006).

METHOD

Participants

The data were derived from the COOS, a large-scale study into “normal”
cognitive development (Cognitief Ontwikkelings Onderzoek bij Schoolgaande
kinderen, in English: Cognitive developmental study in school-aged children). In
the COOS, \( N = 29 \) regular primary and secondary schools in Maastricht (a city in
the South of the Netherlands) were contacted for study participation. Schools that
agreed to participate received information packages for the parents of their pupils.
These information packages contained a letter about the purpose of the study, a
request for their child to participate, and a stamped return envelope. Of the
approximately \( N = 3000 \) information packages that were provided to the parents,
\( N = 1086 \) packages were returned. A total of \( N = 892 \) of the parents (82.14\%) gave
consent for their child to participate in the research. In this group the LGST was
administered to a random sample of \( N = 408 \) children. Children who (1) had
repeated or skipped a grade, (2) used medication known to affect cognitive
performance (such as Ritalin), or (3) suffered from severe medical conditions known
to affect cognitive performance (such as epilepsy) were excluded from the data
analyses (\( N = 3 \)). This participant sampling procedure and these exclusion criteria
were aimed at selecting “typically developing” children (i.e., children who were
attending a school for regular education, who were in the appropriate grade, and
whose cognitive abilities were not affected by medical conditions or medication).

Basic demographic data for the \( N = 405 \) eligible children are provided in
Table 1. All children were native Dutch speakers. The children were aged between
5.57 and 15.87 years. Age was used as a continuous variable in the analyses (see
below), but it was categorized into bands of about 2 years in Table 1 (for descriptive
purposes). The educational level of the children’s parents (or caregivers) was
measured with a commonly used Dutch educational 8-point rating scale that ranges
from primary school to university degree (De Bie, 1987). Mean Level of Parental

<table>
<thead>
<tr>
<th>Age</th>
<th>Female : Male</th>
<th>MLPE Low</th>
<th>MLPE High</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \leq 8 ) year</td>
<td>6.51 (0.61)</td>
<td>40 : 35</td>
<td>46</td>
<td>29</td>
</tr>
<tr>
<td>8–10 year</td>
<td>8.66 (0.58)</td>
<td>27 : 38</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>10–12 year</td>
<td>10.68 (0.55)</td>
<td>48 : 30</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>12–14 year</td>
<td>12.97 (0.54)</td>
<td>59 : 67</td>
<td>62</td>
<td>64</td>
</tr>
<tr>
<td>( \geq 14 ) year</td>
<td>14.51 (0.34)</td>
<td>28 : 33</td>
<td>40</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>10.87 (2.83)</td>
<td>202 : 203</td>
<td>223</td>
<td>182</td>
</tr>
</tbody>
</table>

MLPE = Mean level of parental education.
Education (MLPE) was dichotomized into Low and High groups (after a median split) for parents who had a maximum educational level that was \(5\) and \(\geq 5\) on the 8-point scale, respectively (with 5 = preparatory vocational education). These two levels of education corresponded with a mean (SD) of about 9.88 (2.59) and 14.68 (3.30) years of full-time education, respectively. The Ethics Committee of the Faculty of Psychology and Neuroscience of Maastricht University (the Netherlands) approved the study protocol.

**Procedure and instruments**

The Dutch version of the LGST was individually administered to the children at their own school. The LGST items (see Table 2 below, left column) were read aloud to the child by the experimenter. Items were repeated if requested. Each item was scored as 1 (correct response) or 0 (incorrect response). All children were allowed to answer the items without a time limit being set. The LGST took less than 10 minutes to administer. The LGST was part of a larger cognitive battery that took about 90 minutes to be completed. The cognitive battery focused on language abilities and on several other cognitive domains, such as memory and sense of time. All tests were administered in the same order for each child.

### Table 2. Psychometric properties of the items of the Logical Grammatical Structures Test as established under the 2PL IRT framework

<table>
<thead>
<tr>
<th>Item</th>
<th>Item Difficulty ((b_i))</th>
<th>Item Discrimination ((a_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Is my father’s brother the same person as my brother’s father?</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2: Draw a cross beneath a circle</td>
<td>−2.24</td>
<td>1.42</td>
</tr>
<tr>
<td>3: Draw a circle to the right of a cross</td>
<td>−1.45</td>
<td>1.83</td>
</tr>
<tr>
<td>4: Draw a circle to the right of a cross but to the left of a triangle</td>
<td>−0.17</td>
<td>1.68</td>
</tr>
<tr>
<td>5: Which is correct: “Spring comes before summer” or “Summer comes before spring”?</td>
<td>−2.35</td>
<td>0.92</td>
</tr>
<tr>
<td>6: Who is shorter if John is taller than Pete?</td>
<td>−2.06</td>
<td>1.27</td>
</tr>
<tr>
<td>7: Which is correct: “A fly is bigger than an elephant” or “An elephant is bigger than a fly”?</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>8: Which girl is lightest if Olga is lighter than Sonia but darker than Kate?</td>
<td>−0.71</td>
<td>1.41</td>
</tr>
<tr>
<td>9: And which girl is darkest?</td>
<td>−0.33</td>
<td>1.28</td>
</tr>
<tr>
<td>10: Nick was struck by Pete: Who was the victim?</td>
<td>−1.21</td>
<td>1.38</td>
</tr>
<tr>
<td>11: I had breakfast after I had walked the dog. What did I do first?</td>
<td>−1.37</td>
<td>1.07</td>
</tr>
<tr>
<td>12: The girl who worked in the zoo came to Margret’s school to give a presentation. Who gave the presentation?</td>
<td>−0.95</td>
<td>1.18</td>
</tr>
<tr>
<td>13: What did Margret do?</td>
<td>−0.34</td>
<td>1.79</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>−1.20 (0.78)</td>
<td>1.38 (0.29)</td>
</tr>
</tbody>
</table>

– = The item was excluded because of insufficient variability (item 7) or because it did not fit the model (item 1).
Statistical analyses

Confirmatory Factor Analysis (CFA) was conducted to evaluate the dimensionality of the LGST. An a priori model was specified in which one factor (i.e., logical grammatical structures comprehension) underlies the item responses for the LGST. The Diagonally Weighted Least Squares method for polychoric correlation matrices was used instead of the standard Maximum Likelihood estimation method (because the items were dichotomous; Schumacker & Beyerlein, 2000). The fit of the models was evaluated with the Root Mean Square Error of Approximation (RMSEA; < .08 acceptable, < .05 excellent; Browne & Cudeck, 1992), the Comparative Fit Index (CFI; > .90 acceptable, > .95 excellent; Bentler, 1990; Bentler & Bonett, 1980), and the Normed Fit Index (NFI; > .90 acceptable; Bentler & Bonnet, 1980).

Next, a 2-Parameter Logistic (2PL) Item Response Theory (IRT) model was fitted to the data. The 2PL model assumes that the probability that person \(s\) solves item \(i\) can be modelled as

\[ P(X_s = 1|\theta_s, \beta_i, \alpha_i) = \frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]} \]

The item difficulty parameter (\(\beta_i\)) refers to the ability level that is needed for an individual to answer LGST item \(i\) correctly with a 50% probability. The item discrimination (or slope) parameter (\(\alpha_i\)) indicates how quickly the probability of a correct response changes as a function of ability level. Item fit was evaluated with Bock’s \(\chi^2\) index (Bock, 1972). A Monte Carlo procedure with 500 replications was used to approximate the distribution of the item-fit statistics under the null hypothesis (i.e., under the assumption that the observed data were generated under the 2PL model). Items that did not fit the model (i.e., items with \(p\)-values of the Bock \(\chi^2\) index < .01) were removed. Expected A Posteriori (EAP) scoring with 21 quadrature nodes was used to estimate the ability levels of the children. The EAP is a Bayesian estimator derived from the mean of the posterior distribution. The standard errors of the EAP trait level estimates were derived by computing the posterior standard deviations (Bock & Mislevy, 1982).

The logical grammatical structure comprehension ability estimates of the children that were obtained as based on the IRT analyses (i.e., the \(\theta_s\) estimates) were subsequently regressed on Age, Age\(^2\), Gender, MLPE, and all two-way interactions between these predictors. Age was centred (age = calendar age – 11) before quadratic terms and interactions were calculated (to avoid multicollinearity; Kutner, Nachtsheim, Neter, & Li, 2005). Both Age and Age\(^2\) were added as predictors in the model, to allow for a curvilinear relationship between age and the logical grammatical structures comprehension ability. Sex was coded as boy = 1 and girl = 0. MLPE was coded as high = 1 and low = 0. Non-significant predictors (\(p > .01\)) were excluded from the full model, but no predictor was removed as long as it was also included in a higher order term in the model (Aiken & West, 1991). The assumptions of regression analysis were tested for each model. Homoscedasticity was evaluated by grouping the participants into two groups (children who had low and high predicted scores, based on a median split) and applying the Levene test. Normality of the residuals was investigated by conducting Kolmogorov-Smirnov tests on the standardized residuals. The occurrence of multicollinearity was checked by calculating Variance Inflation Factors (VIFs), which should not exceed 10.
Normative conversions of the $y$ scores into demographically corrected percentile values were established by means of a four-step procedure (Van Breukelen & Vlaeyen, 2005; Van der Elst, 2006; Van der Elst, Hurks, Wassenberg, Meij, & Jolles, 2011). First, the expected ability scores were computed by means of the final regression model ($y = B_0 + B_1X_1 + \cdots + B_nX_n$, with $B_0$ = the intercept, $B_n$ = the regression weight(s), and $X_n$ = the predictor values). Second, the residuals were calculated ($e_i$ = observed ability score – expected ability score). Third, the residuals were standardized ($Z_i = e_i / SD_{\text{residual}}$, with $SD_{\text{residual}}$ = the standard deviation of the residuals in the normative sample). Fourth, the standardized residuals were transformed into percentiles by means of the standard normal distribution of the residuals.

All analyses were conducted with PRELIS/LISREL 8.8 for Windows and the R 2.11.1 software package (R Development Core Team, 2010) for Linux. The ltm library (Rizopoulos, 2006) was used to conduct the IRT analyses. An $\alpha$ level of .01 was used in all analyses.

RESULTS

Psychometric properties of the LGST

Only 2 out of the 405 children (i.e., less than .005% of the sample) answered item 7 incorrectly. This uninformative item was thus removed from the analyses. Next a one-factor CFA model was fitted to the remaining items. The Modification Indices suggested that an error covariance path should be added between items 12 and 13. This error covariance path was not unexpected, because both items refer to the same initial logical grammatical structure (i.e., “The girl who worked in the zoo came to Margret’s school to give a presentation.”; Item 12: “Who gave the presentation?”; Item 13: “What did Margret do?”). After making this modification the model had an excellent fit with the data ($RMSEA = .04$, $NFI = .98$, $CFI = .99$), and all items loaded significantly on the underlying factor. Note that the LGST consists of items that require two different response modalities, i.e., motor responses (items 2, 3, and 4) and verbal responses (the other items). The question arises whether a two-factor model (in which response modality is reflected) provides a better fit to the data than the one-factor model. Thus a second CFA was conducted in which items 2, 3, and 4 were assumed to load on a first latent variable, and items 5, 6, 8, 9, 10, 11, 12, and 13 were assumed to load on a second latent variable. Both models fitted the data equally well—two-factor model: $\chi^2 = 75.41$ ($df = 42$), one-factor model: $\chi^2 = 74.01$ ($df = 43$); $\chi^2$ difference test $= 1.40$ ($df = 1$), $p = .24$—but the one-factor model is preferred over the two-factor model because it is more parsimonious.

Next the IRT analyses were conducted. The first item of the LGST scale did not fit the 2PL model (Bock $\chi^2$ values $= 35.59$, $p$-value $< .01$). After removal of this item, all remaining items had an excellent fit with the model (all $p$-values of the Bock $\chi^2$ values $> .15$). The item difficulty ($\beta_i$) and item discrimination ($\alpha_i$) parameters of the 11 remaining items are shown in Table 2. The mean ($SD$) item difficulty and
item discrimination parameters equalled $-1.20 (0.78)$ and $1.38 (0.29)$, respectively. Item 5 and item 4 had the lowest and highest difficulty levels, respectively (i.e., $\beta_5 = -2.35$ and $\beta_4 = -0.17$). Item 5 and Item 3 had the lowest and highest discrimination levels, respectively (i.e., $\alpha_5 = 0.92$ and $\alpha_3 = 1.83$). The item characteristic curves of these items (which depict the probabilities of a correct item response as a function of $y_s$) are shown in Figure 1.

The LGST information curve is shown in Figure 2. This curve depicts test information as a function of $\theta_s$—where Information($\theta_s$) = $1/\text{SEM}(\theta_s)$. As shown, the measurement precision of the LGST peaked at $\theta_s$ levels that were in the range between about $-2$ and $0$ (i.e., information was $>3.33$, which corresponds to reliability levels $>.70$). The internal consistency of the items of the LGST equalled $.73$ (as estimated with Cronbach’s $\alpha$ coefficient).

### The effects of age, gender, and mean level of parental education on the LGST scores

Table 3 presents the final regression model for the $\theta_s$ score. Age, age$^2$, and MLPE had significant effects on the $\theta_s$ score. As shown in Figure 3, there was a strong non-linear effect of age on logical grammatical structures comprehension ability. That is, the relative improvement in test performance was much more...
pronounced for younger children than for older children. Moreover, children who had parents with a high MLPE had significantly higher \( \theta_s \) scores as compared to children who had parents with a low MLPE. There were no gender differences in the \( \theta_s \) scores. The regression model explained 42.9\% of the variance in the \( \theta_s \) score. The \( SD(\text{residual}) \) value for the final model equalled 0.64.

All assumptions of regression analysis were fulfilled. There was no serious influence of outliers (the maximum Cook’s distance equalled 0.06) or

<table>
<thead>
<tr>
<th>Variable</th>
<th>( B )</th>
<th>Std. Error ( B )</th>
<th>Standardized ( B )</th>
<th>( T )</th>
<th>Adjusted ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(constant)</td>
<td>0.067</td>
<td>0.058</td>
<td>1.162</td>
<td></td>
<td>.429</td>
</tr>
<tr>
<td>Age</td>
<td>0.162</td>
<td>0.013</td>
<td>0.542</td>
<td>12.927*</td>
<td></td>
</tr>
<tr>
<td>( Age^2 )</td>
<td>-0.020</td>
<td>0.005</td>
<td>-0.174</td>
<td>-4.143*</td>
<td></td>
</tr>
<tr>
<td>MLPE</td>
<td>0.239</td>
<td>0.064</td>
<td>0.141</td>
<td>3.740*</td>
<td>.429</td>
</tr>
</tbody>
</table>

Coding of the predictors: Age = calendar age (years) – 11; \( Age^2 = (\text{calendar age (years) – 11})^2 \); MLPE: Low = 0, High = 1. MLPE = Mean level of parental education.

*\( p < .01 \).
multicollinearity (maximum VIF = 1.25) observed for the model. The Levene test suggested that there was no heteroscedasticity (Levene statistic = 0.147, \( p = .70 \)), and the standardized residuals of the models were normally distributed (Kolmogorov–Smirnov \( Z = 1.12, p = .17 \)).

**Normal range of performance**

Normative LGST data are established by means of the four-step procedure described above. For example, suppose that a 12-year-old child who has parents with a low MLPE had the first 8 items of the LGST correct and the last 3 items incorrect. This item response pattern corresponds with a \( \theta \) score equal to \( -0.12 \) (see next section). In the first step of the normative procedure the expected \( \theta \) score of the child is computed, i.e., \( 0.209 (=0.067 + (0.162 \times (12 - 11)) + (-0.020 \times (12 - 11)^2)) \). Second, the residual is calculated, i.e., \( -0.329 (= -0.12 - 0.209) \). Third, the residual is standardized, i.e., \( -0.51 (= -0.329/0.64) \). Fourth, the standardized residual is converted into a percentile value using the standard normal distribution. A standardized residual equal to \( -0.51 \) corresponds with a percentile value of 30, indicating that 30\% of the 12-year-old children who have parents with a low MLPE
had lower levels of logical grammatical structures comprehension as compared to this child.

Note that, in some situations, a clinician or researcher may find it more useful to use age-corrected normative data (instead of age- and MLPE-corrected normative data). To allow for this possibility we established a regression model in which MLPE was not included as a covariate in the final model (i.e., expected \( \theta_s = 0.186 + \text{age} \times 0.161 + \text{age}^2 \times -0.021; SD(\text{residual}) = 0.65; R^2 = .41 \)). Based on this regression equation and the \( SD(\text{residual}) \) value, the four-step normative procedure can be applied to convert a tested child’s raw LGST score into an age-corrected percentile value. For example, the expected \( \theta_s \) score of the child from the previous example equals 0.326 \((= 0.186 + (0.161 \times (12 - 11)) + (-0.021 \times (12 - 11)^2))\), the residual is \(-0.446 \((=-0.12 - 0.326)\), and the standardized residual equals \(-0.69 \((=-0.446/0.65)\), which corresponds to a percentile value of 25.

**Automatic scoring program**

Converting the item response pattern of a tested child into a \( \theta_s \) score is based on EAP scoring, which requires a large number of computations and is difficult to do manually. We therefore devised a computer-based algorithm that makes the required computations automatically. The user of this algorithm simply types in the age of the child, the MLPE of his or her parents, and the LGST item responses. Next the algorithm automatically computes the \( \theta_s \) level of the child, and converts this measure into an age- and MLPE-corrected percentile or an age-corrected percentile value (depending on the choice of the user). The algorithm runs under Microsoft Excel (or a compatible software package), and can be downloaded from http://home.deds.nl/~wimvde/.

**DISCUSSION**

The results of the present study suggested that age had a highly significant non-linear effect on logical grammatical structures comprehension ability, i.e., the relative improvement in LGST performance was much more pronounced for younger children than for older children (see Figure 3). For example, the average logical grammatical structures comprehension ability of a 7.5-year-old child was about 35% higher as compared to the ability level of a 6.5-year-old child, while the average logical grammatical structures comprehension ability of a 15.5-year-old child was identical to the ability level of a 14.5-year-old child. In addition to age and \( \text{age}^2 \), MLPE profoundly affected logical grammatical structures comprehension. For example, the average 6.5-year-old child who has parents with a high MLPE had a level of logical grammatical structures comprehension that was approximately equal to the ability level of an average 7.25-year-old child who has parents with a low MLPE. Similar observations have been made in younger children regarding other language abilities. For example, 2-year-old children who have parents with a high educational level were found to have a substantially higher rate of vocabulary acquisition than their counterparts who have parents with a low educational level (Hoff, 2003). The effect of parental education on language development in children
and adolescents has been attributed to differences in the quantity, sentence complexity, and lexical richness of parents’ speech to their children (Hoff, 2003). In addition to environmental factors, significant heritability for language has been established (Hayiou-Thomas et al., 2006; Hoekstra, Bartels, & Boomsma, 2009; Spinath, Price, Dale, & Plomin, 2004). Recent studies have indicated that there is an increase in the heritability for language between early and middle childhood, i.e., environmental factors account for most of the variance in early language ability while genetic factors are more important in middle childhood and early adolescence (Hayiou-Thomas, Dale, & Plomin, 2012). We did not use a genetically sensitive design in the present study, so environmental and genetic influences on the development of logical grammatical structure comprehension could not be distinguished.

Note also that the effect of MLPE on logical grammatical structures comprehension ability was not moderated by age (i.e., there was no significant MLPE × Age interaction term in the regression model, see Table 3). This suggests that the logical grammatical structure comprehension abilities of children who have parents with a low MLPE never “catch up” with their counterparts who have parents with a high MLPE (at least not in the age range that we considered).

The psychometric analyses showed that a one-factor model had an excellent fit with the data (supporting the factorial validity of the LGST), and the 2PL model adequately fitted the LGST items (with the exception of one item). Classical Test Theory (CTT) is the dominant psychometric approach in neuropsychology, but we used an IRT-based approach in the present study. This has several advantages (for details, see Embretson & Reise, 2000). For example, IRT models allow for estimating item-invariant and person-invariant parameters. This means that the parameters that characterize the examinees (i.e., their ability levels) are independent from the specific items that were administered, and the parameters that characterize the items (i.e., their difficulty and discrimination levels) are independent of the ability distribution in the normative sample (Hambleton, Swaminathan, & Rogers, 1991). This property is important in the context of establishing unbiased normative data. Another advantage of IRT models is that they facilitate an item-based interpretation of the examinee’s ability level. For example, knowing the ability level of a tested child immediately allows for an interpretation of the probability by which the tested child will answer a particular LGST item correctly (see Figure 1). Moreover, IRT models do not assume an equal measurement precision across the entire ability range (i.e., they allow for different measurement errors as a function of ability level). The analyses showed that the measurement precision of the LGST was especially high (>3.33, corresponding with reliability levels >.70) in the range between having poor to average ability levels (i.e., between −2 and 0; see Figure 2). Outside this range the LGST was less discriminative. The standard errors of measurement thus become larger for more extreme (i.e., very high or very low) ability levels, which can be intuitively easily understood. For example, children who have low (e.g., 0s = −2) and very low ability levels (e.g., 0s = −2.5) will have a very similar item response patterns (i.e., most items will be answered incorrectly), and it will consequently be difficult to accurately distinguish the trait levels of both children.
The LGST and its normative data provide useful tools in applied and research settings when logical grammatical structure comprehension abilities need to be assessed, but there are some limitations that warrant further discussion. First, IRT models have several advantages as compared to CTT models (see above), but they also have some disadvantages. For example, the application of IRT-based tests in applied and research settings is often hampered because converting the item responses into ability estimates is computationally involved (in contrast to the situation for CTT-based tests, where the ability estimates are simply computed as the sum of the item scores). To increase the use of IRT-based tests in applied settings, it is important that user-friendly scoring procedures are provided (as we did with the automatic scoring program).

Second, LGST performance no longer improved from the age of about 14.5 years (see Figure 3). This finding may suggest that adult levels of logical grammatical structures comprehension are achieved at the age of about 14.5 years, but an alternative explanation is that ceiling effects are responsible for this finding. The latter possibility seems less probable, because the mean (and Standard Error of the Mean) number of LGST items that were correctly answered by the children in the normative sample who were aged at least 14.5 years equaled 9.33 (0.33) (while the LGST contains 11 items). Nonetheless, future studies should administer the LGST in a sample in which a broader age range is considered, so that this issue can be evaluated in more detail.

Third, all children who participated in the present study were native Dutch speakers. It remains to be determined whether the established normative data can also be applied to children who have a different native language. Future studies should evaluate this issue.

REFERENCES


Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29–51.


