



**HAL**  
open science

## Validation and psychometric properties of the "Reading the Mind in the Eyes Test"

Renaud F. Cohen, Anna Maria Berardi, Pascal Le Vaou, Jean-Pierre Kahn

### ► To cite this version:

Renaud F. Cohen, Anna Maria Berardi, Pascal Le Vaou, Jean-Pierre Kahn. Validation and psychometric properties of the "Reading the Mind in the Eyes Test". 2019. hal-02096607

**HAL Id: hal-02096607**

**<https://hal.science/hal-02096607>**

Preprint submitted on 11 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1       **Validation and psychometric properties of the “Reading the Mind in the Eyes Test”**

2

3       Renaud F. Cohen<sup>1, 2,3</sup> ; Anna Maria Berardi<sup>2</sup>; Pascal Le Vaou<sup>4</sup> ; Jean-Pierre Kahn<sup>1, 2,3</sup>

4

5       1 : Centre Psychothérapique de Nancy, 1, rue du Docteur Archambault, BP 11010, 54 521

6       LAXOU cedex. FRANCE.

7       2 : Université de Lorraine, FRANCE.

8       3 : FACE-BD network, fondation FondaMental

9       4: CHR Metz-Thionville

10

11       Renaud F. Cohen

12       Mail : [renaudcohen@hotmail.fr](mailto:renaudcohen@hotmail.fr)

13       Postal address :

14       1, rue du Docteur Archambault, BP 11010, 54 521 LAXOU cedex. FRANCE

15       Phone number: +33675870510

16       Professional phone number: +33383926844

17

18       Conflicts of interest: none

19       This research did not receive any specific grant from funding agencies in the public,  
20       commercial, or not-for-profit sectors.

21 **Abstract**

22           The French validation and psychometric investigation of the “Reading the Mind in the  
23 Eyes Test” (or “Eyes Test”; Baron-Cohen et al., 2001) was carried out on 661 French  
24 participants. Participants completed the Eyes Test, Facial Emotional Recognition test and  
25 Mill-Hill Vocabulary. This study is, to the best of our knowledge, the largest study assessing  
26 the psychometrics characteristics of the Eyes Test, allowing a Bayesian Item Response Model  
27 analysis and the study of its presumed unidimensionality. A subsample of 71 participants  
28 completed the Reading the Mind in the Eyes Test twice. The 3-PL model of Item Response  
29 Theory with Bayesian estimation was used to find the items with significant discriminant  
30 alpha coefficients. Parallel Analysis and Confirmatory Factor Analysis were used to assess the  
31 unidimensionality of the Eyes Test. Thirty one items were included in the final version of the  
32 French Eyes Test. Results suggest that the French Eyes Test with 31 items has good  
33 convergent and discriminant validities and that it fits a one-dimensional model ; moreover, the  
34 test is stable across time. However, the 3-PL Bayesian Item Response Theory Model fit  
35 suggests a high level of correct guessing. The Eyes Test measures a unique ability, namely  
36 affective theory of mind.

37 **Keywords:** Reading the Mind in the Eyes Test, Theory of mind, psychometric properties,  
38 Bayesian item response theory model, social cognition, French validation.

39           **1. INTRODUCTION**

40           Theory of Mind (ToM) is a reflection of social cognition and includes all the abilities  
41 involved in the relationships between humans and their social environment. An NIMH  
42 Workshop on Social Cognition in Schizophrenia defined social cognition as “the mental  
43 operations that underlie social interactions, including perceiving, interpreting, and generating  
44 responses to the intentions, dispositions, and behaviors of others” (Pinkham et al., 2014). The  
45 affective ToM corresponds to the ability to infer the roots of emotions identified in other  
46 people (Shamay-Tsoory & Aharon-Peretz, 2007). This ability is frequently assessed with the  
47 ‘Reading the Mind in the Eyes Test’ (Baron-Cohen et al., 2001) or validated translations of  
48 this test (Hallerbäck, Lugnegard, Hjärthag, & Gillberg, 2009; Sanvicente-Vieira et al., 2013;  
49 Vellante et al., 2013).

50           In the Eyes Test, participants are presented with a series of 36 photographs of the eye-  
51 region of different actors and actresses, and are asked to choose which of four words best  
52 describes what the person in the photograph is thinking or feeling (Baron-Cohen et al., 1997,  
53 2001). The target words and distractors are always complex mental states, like “reflective” or  
54 “irritated” (see the revised version; Baron-Cohen et al., 2001). Therefore, this test assesses  
55 affective ToM (Shamay-Tsoory et al., 2007), as it involves, for every item, the ascription of  
56 complex mental states (Baron-Cohen et al., 2001) and not of “basic emotions” (e.g. Ekman,  
57 1992a, 1992b).

58           The Eyes Test’s capacity to discriminate between subtle variations of ToM levels is  
59 exemplified by Domes et al. (2007), who showed that an intranasal administration of oxytocin  
60 improves the correct recognition of target words on the Eyes Test in healthy participants and  
61 by the evidence of a superior average performance in female compared to male participants  
62 on this test, a result that has been confirmed in a meta-analysis of the ‘Eyes Test’ (Kirkland,  
63 Peterson, Baker, Miller, & Pulos, 2013).

64           Therefore, it is one of the few tests that can assess high levels of performance in social  
65 cognition in healthy adults, where most other tests show strong ceiling effects. But, the  
66 hypotheses, that the test assesses a unique latent ability (i.e. attribution of complex emotion),  
67 deserves an empirical psychometric investigation. Therefore, the validity of the Eyes Test has  
68 been studied, but this does not imply a unique latent ability (unidimensionality of the test).

69           Our aim was also to validate the Eyes Test in French, because it is the most powerful  
70 test to assess ToM, and to demonstrate its psychometric properties, among which  
71 unidimensionality, validity and stability across time. To this purpose, a Bayesian 3-parameter  
72 logistic item response theory (IRT) model was fitted to compute all the coefficients associated  
73 with IRT, namely difficulty (beta), discriminant (alpha) and guessing (eta) parameters of the  
74 items (Lord, 1980). This article is the first to use Bayesian IRT analysis to study the  
75 psychometric properties of the Eyes Test (Fox, 2010). We were particularly interested in the  
76 evaluation of the alpha parameters (which have to be superior to 0) and the guessing  
77 parameters. Indeed, Johnston, Miles, & McKinlay (2008) found that a group of healthy  
78 subjects could infer the target word without the picture associated with it, suggesting high  
79 levels of correct guessing in the original test. The 3-PL model could be computed to obtain  
80 the “eta” parameters, which reflect the items’ associated levels of correct guessing. This  
81 parameter is particularly meaningful, indeed, it corresponds to the probability that a subject  
82 with very low levels of ability responds correctly to the item (DeMars, 2010).

83           One can expect the Eyes Test to correlate with the Facial Emotion Recognition Test  
84 (FERT), because even if the abilities assessed by the Eyes Test are wider than those assessed  
85 by FERT, in both tests facial emotional cues need to be interpreted. Therefore, a positive  
86 correlation between these two tasks of cognitive empathy was expected.

87           Demonstration of the unidimensionality of a test is a pre-requisite to the use of its total  
88 score. However, previous studies of the Eyes Test have generally neglected this psychometric

89 central element, apart from the studies carried out by Preti, Vellante, & Petretto (2017) and  
90 Vellante et al. (2013). The significance of this step in test validation has been pointed out by  
91 Ziegler & Hagemann (2015).

92 Vellante et al. (2013) used “Confirmatory Factor Analysis” (CFA) to test two models  
93 on the Italian Eyes Test. The first model was defined as unidimensional, i.e. all items loaded  
94 on a single latent dimension or factor. This model fitted well the data. The second model was  
95 defined by three dimensions and did not fit the data. The results supported the  
96 unidimensionality of the Italian Eyes Test. However, the method chosen did not take into  
97 account the binary nature of the variables. The second part of their analysis, reported in Preti  
98 et al. (2017), took into account the binary nature of the items with an IRT analysis, but did not  
99 analyze the guessing parameters.

100 The soundest methods to study unidimensionality are Confirmatory Factor Analysis  
101 (CFA) and Parallel Analysis (PA) (Cho, Li, & Bandalos, 2009; Ziegler & Hagemann, 2015).  
102 The CFA can help one to determine if one factor alone explains all responses to the test items  
103 (Brown, 2006). This method can be complemented with the optimal implementation of  
104 parallel analysis (Cho et al., 2009; Drasgow & Lissak, 1983; Lorenzo-Seva & Ferrando, 2013).

105 Therefore, the aim of this study was to assess the reliability (including the  
106 unidimensionality), and the convergent and discriminant validities of the French Eyes Test.  
107 We also wanted to study the fit of a 3-PL Bayesian IRT model, to select the items with non-  
108 zero alpha parameters and to demonstrate the importance of guessing in this test (by studying  
109 guessing parameters). These analyses have never been carried out on the French Eyes test, on  
110 its original English version, or in any of its validated translations.

111        **2. METHODS**

112            2.1.Participants

113    The sample was composed of 661 healthy French-speaking university students and their  
114    relatives, a random subsample of which took the Eyes Test twice ( $n = 71$ ). Participants were  
115    recruited on a voluntary basis. Five hundred twenty five women and 129 men take part to the  
116    study with nine subjects with unavailable gender information. The mean age was 23.1 years  
117    ( $\pm 7.25$ ). A psychologist (RFC) explained and completed the assessments. Participants were  
118    recruited from a University campus in France and later by the snow-ball sampling method.

119            All of the procedures were approved by the ethical committee of the University  
120    Medical Center of Nancy (CHU de Nancy, Université de Lorraine, France).

121    .A random subsample of 71 participants took part in the test twice at least one week apart  
122    (mean =  $13.6 \pm 9.9$  days).

123            2.2.Tests administered

124            Participants were administered a questionnaire to record their demographic  
125    characteristics and health status. A pre-validated version of the French Eyes Test with 35  
126    items (see below for its characteristics), was administered to all participants. The  
127    performances on all items were analyzed, including the practice item of the English version.

128            Simple emotion recognition was assessed with the *Facial Emotion Recognition Test*  
129    (FERT), a 35 item set of faces from the Ekman and Friesen (1975)'s set of pictures of facial  
130    affect; it includes five items for each basic emotion (fear, sadness, disgust, anger, surprise and  
131    happiness) and 5 items for neutral pictures. This classic task requires that subjects identify the  
132    emotional display of the face of an actor. The total score was used to obtain an index of global  
133    recognition of basic emotions. FERT was completed by 614 participants (participants with  
134    missing data were excluded).

135 Verbal intelligence was assessed with the Mill-Hill Vocabulary Test (Raven and  
136 Deltour, 1998) on a subsample of participants (n = 102 with complete data) to test for the  
137 separate effects of global verbal intelligence and of affective ToM. The information collected  
138 was anonymous.

### 139 2.3.Procedures

140 All participants were administered the items of the Eyes and Ekman's tests on a video  
141 screen. For the Eyes test, participants responded by pressing one of four buttons (1,2,4,5;  
142 according to the position of the target word on the screen) on a computer, and, for the  
143 Ekman's test, they selected the target word with the mouse.

144 Mill-Hill was administered in paper form to a subsample of 102 subjects. The size of  
145 this sample is sufficient to detect modest positive associations with either Eyes or Ekman's  
146 tests.

147 The Eyes Test is highly vulnerable to cultural factors (Hallerbäck et al., 2009;  
148 Sanvicente-Vieira et al., 2013; Yıldırım et al., 2011), a problem that elicited difficulties in  
149 previous validations in other languages. In their original version of the Eyes Test, Baron-  
150 Cohen et al. (2001) did not know what the persons in the pictures were thinking or feeling;  
151 this was determined by subjective judgments using groups of eight judges (Baron-Cohen et al.,  
152 1997, 2001). In the current study, the English Eyes Test was first translated into French and  
153 back-translated. Second, the original procedure of Baron-Cohen et al. (2001) with eight  
154 judges was used to select reliable items. Five judges out of eight had to agree and no more  
155 than two judges had to choose the same distractor as the target. After successive assessments  
156 of this type, a 35-items test without any practice items was obtained. An item from the child  
157 version of the test, with new words (target and distractors), was added. We carried out the  
158 statistical analyses described below on this 35-items pre-validated French version of the Eyes  
159 Test to select the items to be included in the final version.



160           2.4.Statistical procedures

161   The statistical analyses were carried out with the software R version 3.2.0 (R Development  
162   Core Team, 2014), FACTOR 10 (Lorenzo-Seva & Ferrando, 2013) and JAGS software (Lunn,  
163   Spiegelhalter, Thomas, & Best, 2009; Plummer, 2003), interfaced with R with the R2jags  
164   package.

165           2.4.1. Statistical procedures to obtain the final version of the French Eyes Test

166   To select the items for the final French version of the Eyes Test, a three-step procedure was  
167   followed. First, a 3-parameter logistic model (Lord, Novick, & Birnbaum, 1968) with  
168   Bayesian estimation was fitted as described in Curtis et al (2010). JAGS software interfaced  
169   with R (R2jags package) was used on the full test (35 items). Markov Chain Monte Carlo  
170   Methods estimation were used to obtain full posterior distributions of the parameters of the  
171   3-PL IRT model with the script from Curtis (2010). A small modification of the script was  
172   included, allowing for a small negative alpha in the prior random distribution (left truncation  
173   at  $\alpha = -0.5$  with a prior normal distribution of  $m = 1$  ;  $sd = 1$ ) This analysis allowed to compute  
174   the alpha parameters to observe if the credible interval at 95 % included 0 for some of the  
175   parameters, whereas the majority of alpha are constrained to be positive. This procedure  
176   would indicate the non-discriminant items to be suppressed from further analysis.

177   We used a burn-in period of 10 000 iterations, then we launched 100 000 iterations for 3  
178   chains (total iterations = 300 000), with a thinning interval of 10. Therefore, we used 30 000  
179   iterations for each parameter.

180   To confirm that the chains converge, multiple methods were used (Kruschke, 2015). The  
181   convergence of the 3 chains for each item (mixing of the 3 chains) was inspected visually.  
182   The potential scale reduction factor was used and should tend towards 1 at convergence (see  
183   Gelman & Rubin, 1992). Another diagnostic is the Geweke diagnostic, which takes two  
184   nonoverlapping parts (usually the first 0.1 and the last 0.5 proportions) of the Markov chain

185 and compares the means of both parts by using a difference among the means to verify  
186 whether the two parts of the chain are from the same distribution (Geweke, 1991). This last  
187 index could be used with a unique chain, but the 3 chains were successively tested and only  
188 the statistics, which were significant in at least two of the chains for the same parameter, were  
189 considered as significant.

190 Second, a Confirmatory Factor Analysis (CFA, with the *lavaan* package: Rosseel, 2012) to  
191 the remaining 31 items was fitted with the diagonal Weight Least Square method with Mean  
192 and Variance adjustment (WLSMV), which is specifically efficient with dichotomous items  
193 (Beauducel & Herzberg, 2009), to study their unidimensionality. This analysis was confirmed  
194 with the optimal implementation of parallel analysis as described in Lorenzo-Seva &  
195 Ferrando (2013) with FACTOR 10 software.

196 The fit of the one-factor CFA model on the Eyes Test-31 was assessed with the following  
197 criteria: RMSEA (Root Mean Square Error of Approximation)  $<.05$  (good fit of the model),  
198 CFI (comparative fit index)  $>.90$ , TLI (Tucker-Lewis index)  $>.90$  (Hu & Bentler, 1998,  
199 1999) and WRMR (weighted root-mean-square residual)  $<.90$  (Yu, 2002). If these four  
200 criteria were fulfilled, the model was considered to fit well the data. The standard criteria  
201 were used for parallel analysis based on tetrachoric correlations (see Lorenzo-Seva &  
202 Ferrando, 2013).

203 The most consensual indices of reliability were computed to study the homogeneity of the  
204 Eyes Test-31: McDonald's omega coefficients (McDonald, 1999), using the method suggested  
205 by Gadermann, Guhn, and Zumbo (2012), who proposed tetrachoric correlations rather than  
206 Pearson correlations to compute reliability indices for binary data. Shapiro-Wilk tests were  
207 used to determine if the scores followed a normal distribution.

208 Thirdly, a 3-parameter logistic model with Bayesian estimation (Fox, 2010) was fitted again  
209 on the 31 remaining items. This analysis (using the same computing characteristics as

210 described earlier) allowed us to compute the posterior distribution of guessing parameters and  
211 to study their characteristics (median values and distributions) and therefore their potential  
212 utility.

#### 213 2.4.2. Discriminant and convergent validities

214 To study convergent and discriminant validities, we computed two Spearman's correlations  
215 between the French Eyes-Test-31 and both FERT and Mill-Hill. A positive correlation  
216 between FERT and French Eyes-Test-31 would support convergent validity, because they both  
217 assess emotion recognition, whereas a non significant correlation between Mill-Hill and  
218 French Eyes-Test-31 would support discriminant validity (the relative independence of the  
219 Eyes Test from intellectual level). A positive partial correlation between French Eyes-Test-  
220 31 and FERT, adjusted on Mill-Hill score, would discard the hypothesis that the relation  
221 between the French Eyes-Test-31 and FERT is linked to non-specific factors like verbal or  
222 global intelligence.

#### 223 2.4.3. Test-retest stability

224 The intraclass correlation coefficient (ICC) "agreement" was computed to compare the test-  
225 retest stability of performance on a subsample of 71 participants (Weir, 2005). A high level of  
226 this coefficient ( $>.70$ ) indicates that the absolute score of the participants does not differ  
227 between the two occasions.

228

### 229 3. RESULTS

#### 230 3.1. Selection of items for the French Eyes Test-31

231 The French pre-validated version of 35 items was the point of departure for the selection of  
232 items to be included in the final French Eyes Test.

233 A 3-PL Bayesian IRT model was fitted on the 35 items and the credible intervals of  
234 the discriminant coefficients alpha were examined. The credible intervals included 0 for four  
235 items. Therefore, 31 items remained for the CFA (cf. table 1).

236 Insert Table 1 here

### 237 3.2. One factor Confirmatory Factor Analysis

238 The fit for the 31-items of the final French Eyes Test was good for the DWLS estimation  
239 (followed by Robust estimation) : CFI = .950 (.861), TLI = .946 (.851), RMSEA = .012 (.015)  
240 CI = [.000, .018] < .05, SRMR = 0.081, WRMR = .979  $\approx$  .90. These fit indices are acceptable  
241 according to generally accepted criteria (McDonald & Ho, 2002; Yu, 2002), but the robust  
242 estimation is less satisfactory than the classic estimation. The one factor model is supported.  
243 The modification indices suggest that including error covariance between residuals of items 4  
244 and 35 could increase the fit of the model. With this modification the fit is even better:  
245 CFI = .964 > .90 (.879), TLI = .962 > .90 (.870), RMSEA = .010 (.014) CI = [0.000, 0.017]  
246 < .05, SRMR = .080, WRMR = .965  $\approx$  .90.

### 247 3.3. Optimal implementation of Parallel Analysis

248 According to the chosen criteria, the parallel analysis confirmed the unidimensionality, with  
249 only one observed eigenvalue superior to the 95<sup>th</sup> quantiles of random eigenvalues (results  
250 from FACTOR 10 software).

### 251 3.4. Statistical characteristics of the Eyes Test–31

252 The final test includes 17 pictures of the eyes-region of men and 14 similar pictures of  
253 women. The reliability coefficients, based on the tetrachoric correlation matrix, are ordinal  
254 omega  $\omega$  = .79.

255 The Shapiro-Wilk test is significant ( $W = 0.954, p < .0001$ ), suggesting that the distribution of  
256 scores does not follow a normal distribution (see figure 1). The confidence interval of the total  
257 score at 90 percent (from 5 to 95 quantile) is  $CI = [18 - 29]$ .

258 Insert Figure 1 here

259  
260 The total average of the means of each item on the 661 participants is .776 and the  
261 average of the standard deviations for each item is .389. The range of item-total corrected  
262 correlations in our sample was of 0.092 to 0.406 with a median of  $r_{pbis} = 0.202$ .

263 The properties of the items are presented in table 2, where one can see that the target words  
264 that obtained the lowest percent of correct responses (44.8 % for item 26 and 47.5 % for item  
265 8) have a significant discriminant coefficients in the 3-PL model (their credible interval did  
266 not include 0). Therefore two items obtained a mean success score inferior to .50 and their  
267 distractors obtained more than 25 percent of total responses (Baron-Cohen et al.'s criteria,  
268 2001); but they present good psychometrical characteristics in the 3-PL IRT model  
269 (discriminant coefficients superior to 0) and the CFA model.

270 Insert Table 2 here

### 271 3.5.Three-Parameters Logistic Bayesian Item Response Model Estimation of the Eyes 272 Test-31

273 The 3 chains converge after the initial burn-in period of 10 000 iterations, as suggested by  
274 visual inspection showing mixing of the chains, with a potential scale reduction factor tending  
275 to 1 and Geweke diagnostics being non significant for each parameter in more than one chain.  
276 All alpha parameters' credible intervals are superior to zero (quantiles 2.5% of alpha  
277 parameters  $> 0$ ), confirming that all items contribute to the measure of the unique latent  
278 dimension.

279 Insert Table 3 here

280 The guessing coefficients (eta) are particularly interesting (see table 3). The mean of the 31  
281 eta coefficients is 0.504 [min = 0.154; max = 0.820] (estimated as the median of the values  
282 obtained at each iteration), which is substantially higher than 0.25. We could have hypothesed  
283 the .25 value from the fact that it is a multiple choice decision with 4 possible choices.  
284 Only two items obtained a computed median value inferior to .25 (item 8 and item 19).  
285 Twelve items out of 31 had 95% credible intervals which did not include .25 (strictly superior  
286 to .25). These results show a remarkably high level of guessing.  
287 The high levels of guessing and the evidence for the good convergence of the model  
288 estimation indicate the usefulness of computing a 3-PL item response model for the French  
289 Eyes-Test-31.

### 290 3.6. Convergent and discriminant validities of the Eyes Test-31

291 The Spearman correlation between the Mill-Hill and the French Eyes-Test-31 is not significant  
292 in the subsample with complete data on all three tests (N = 102;  $r = .11$ ,  $p > .10$ ). The  
293 correlation between FERT and Mill-Hill is statistically significant ( $r = .26$ ;  $p < .01$ ).  
294 Moreover, even after partialling out the Mill-Hill score, the partial correlation between the  
295 French Eyes-Test-31 and FERT is still significant ( $r = .22$ ;  $p < .05$ ). Therefore, as expected,  
296 the French Eyes-Test-31 is associated with the FERT even when the global effect of  
297 intellectual level is statistically controlled. Finally, the results suggest a stronger correlation  
298 between FERT and Mill-Hill than between the French Eyes-Test-31 and Mill-Hill.

### 299 3.7. Test-retest reliability

300 The ICC for agreement was .749, CI = [.62, .84] suggesting that the scores are stable across  
301 two evaluations.

302       **4. DISCUSSION**

303       The psychometric properties of the French Eyes Test-31 are similar to those of the original  
304       English version (Baron-Cohen et al., 2001). The results showed that a unique latent dimension  
305       could explain the pattern of performance on this test, namely the affective ToM. The results  
306       showed good reliability and validity of the Eyes Test-31 (test-retest reliability, convergent and  
307       discriminant validities).

308               4.1. Psychometric properties of the French Eyes Test-31

309               Given the criticisms with respect to Cronbach's alpha, a more consensual coefficient  
310       was computed, namely the "ordinal" version of the *omega coefficient* (Gadermann et al.,  
311       2012; McDonald, 1999)  $\omega = .79$ , which is satisfactory according to the rule of index  $\geq .70$ .

312               The item-total correlations for the French Eyes Test-31 present values in the small  
313       range for effect size (Cohen, 1988). The mean corrected item-total polychoric correlations are  
314       low, as are the mean inter-item tetrachoric correlations. These results are related to the  
315       method chosen to select reliable items, a process that was named the *consensus method* in the  
316       original article by Baron-Cohen et al. (2001). By definition, the percentage of correct  
317       responses to an item cannot be inferior to 50%, inducing necessarily a small variance.  
318       Therefore, selected items with less than fifty percent of correct responses but with good  
319       psychometric properties is justified, with the aim to increase the difficulty of the test. Indeed,  
320       the variance for a binary variable is mathematically linked to its mean and therefore the range  
321       of item-total Pearson correlations for binary items is frequently restricted if the mean of the  
322       items is not .50 (McDonald, 1999). The consensus method requires that the variance of each  
323       item is small, because all the items should have a mean of correct response superior to .50.  
324       This bias (low variance of items) is clearly apparent in the change of the size of omega when  
325       it is computed as an ordinal coefficient: .79 versus .61 for the "classic" omega coefficient.

326 This surprising change could be explained by the dichotomous nature of the items and their  
327 low variance. Indeed, the tetrachoric correlations appear to be a good measure for the low  
328 covariance and allow the omega coefficient to be probably closer to its true value, contrary to  
329 the Pearson correlations. A proof of the interpretation of the high guessing as caused by the  
330 design of the items and the small variance induced, is that one of the two items with the  
331 smallest guessing parameters is one with an associated high level of choice of the distractors  
332 (31.7%); according to Baron-Cohen et al. (2001), this item should be removed. However, the  
333 results suggest that this item should not be excluded.

334 Confirmatory Factor Analysis and parallel analyses, two among the strongest methods  
335 to confirm dimensionality, support the unidimensionality of the French Eyes-Test-31. Our  
336 data support the hypothesis that a unique latent variable explains the results on the test and  
337 this unique ability is probably affective ToM. This hypothesis is further confirmed by the  
338 positive relationship between the French Eyes Test-31 and the FERT. Item 4 and 35 present a  
339 covariance that cannot be explained uniquely by the general factor, probably because of the  
340 use of the same target word (“charmeur”).

341 Responses to the French Eyes-Test-31 involve various abilities, like the understanding  
342 of the word, visual face perception, the ability to detect the target word, elimination of the  
343 implausible words or distractors (Johnston et al., 2008), and probably the *g factor*. Peterson  
344 and Miller (2012) found that 25 % of the variance in the Eyes Test was explained by an index  
345 of the *Verbal Intelligence Quotient* (VIQ). In a meta-analysis, Baker et al. (2014) showed that  
346 this result is strong and involves both verbal and performance intelligence quotients. However,  
347 their correlation between intelligence and the score at the Eyes Test was not high with  $r = .24$ ,  
348 which is a relatively small effect size.

349 The processes involved in the success on the Eyes Test are complicated by the high  
350 level of guessing revealed in this study. Indeed, for the first time, the high level of guessing in



351 the Eyes-Test has been studied. Our results confirm those of Johnston et al. (2008), who  
352 experimentally studied this effect. Therefore, we confirm the importance of guessing in this  
353 task. The high level of both guessing and correct target recognition affects probably  
354 negatively the sensitivity of the test. The Bayesian item response model analysis showed its  
355 potential to inform us about the characteristics of the test. The consensus method for the  
356 choice of items probably induces a high level of successful guessing.

#### 357 4.2. Convergent and discriminant validities

358 A result of our study is that FERT was significantly positively correlated with verbal  
359 intelligence, as well as with the French Eyes Test-31 ( $p < .05$ ). This result is surprising,  
360 because the words used in the French Eyes Test-31 seem far more complex than the words  
361 used for basic emotions, which are universal; therefore one could presume that verbal  
362 intelligence would be involved in the French Eyes Test-31, by the simple need to understand  
363 the meaning of complex words. However, this does not appear to be the case. This result  
364 suggests that it is not only verbal comprehension that explains the association between the  
365 Eyes Test and intelligence in previous studies, but rather a more general level of intelligence,  
366 like the *g factor* (e.g. Sternberg & Grigorenko, 2002). The present study supports the  
367 hypothesis of an implication of a *g-like factor*, explaining its involvement even in the simpler  
368 facial emotion recognition test.

#### 369 4.3. Test-retest reliability

370 The ICC coefficient confirms the test-retest reliability, with a value superior to .70, which is a  
371 sufficiently high level of test-retest reliability. Therefore, the reliability of the Eyes Test-31 is  
372 supported by the test-retest procedure.

373 **CONCLUSION**

374 The French Eyes-Test-31 is a reliable and valid test, with discriminant items and good  
375 psychometric properties. Its reliability coefficients (ordinal  $\omega$ ) are within the range of  
376 acceptable values (i.e.  $\geq .70$ ); its unidimensionality is supported by the results. The present  
377 results support the use of “ordinal” reliability coefficients. A unique latent ability seems to be  
378 involved in the success on this test, the affective ToM. Evidence is provided for the  
379 convergent and discriminant validities of the French Eyes Test-31, as well as its test-retest  
380 reliability. Its optimal properties are reflected in the quasi absence of "ceiling effects" (only  
381 three of the normal participants out of 661 attained the maximum score). This conclusion is  
382 supported by the computed significant discriminant coefficient (3-PL model) of all the  
383 selected items, which all attest that they participate to the test’s ability to discriminate  
384 between high and low functioning individuals. However, the test could be improved, given  
385 the high level of correct guessing associated with this test. In conclusion, the French Eyes  
386 Test-31 is ‘an advanced test of theory of mind’. It can be used in the general population to  
387 search, for example, for the endophenotypes of neurological or psychiatric syndromes in the  
388 relatives of patients, in contrast to many other tests of ToM, which show strong ceiling effects.

389 **REFERENCES**

- 390 Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis  
391 of the relationship between intelligence and “Reading the Mind in the Eyes”. *Intelligence, 44*,  
392 78–92.
- 393 Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another Advanced  
394 Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or  
395 Asperger Syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813–822.  
396 <https://doi.org/10.1111/j.1469-7610.1997.tb01599.x>
- 397 Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “Reading the  
398 Mind in the Eyes” Test Revised Version: A Study with Normal Adults, and Adults with  
399 Asperger Syndrome or High-functioning Autism. *Journal of Child Psychology and Psychiatry,*  
400 *42*(2), 241–251. <https://doi.org/10.1111/1469-7610.00715>
- 401 Beauducel, A., & Herzberg, P. Y. (2009). On the Performance of Maximum Likelihood  
402 Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA, (October  
403 2014), 37–41. <https://doi.org/10.1207/s15328007sem1302>
- 404 Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford  
405 Press.
- 406 Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with  
407 polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748–759.
- 408 Curtis, S. M. (2010). BUGS code for item response theory. *Journal of Statistical Software,*  
409 *36*(1), 1–34.
- 410 DeMars, C. (2010). *Item response theory*. Oxford University Press.
- 411 Drasgow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining  
412 the latent dimensionality of dichotomously scored item responses. *Journal of Applied*  
413 *psychology, 68*(3), 363.

414 Ekman, P. (1992a). An argument for basic emotions. *Cognition & Emotion*, 6(3-4), 169–200.

415 Ekman, P. (1992b). Are there basic emotions? *Psychological Review*.

416 Ekman, P., & Friesen, W. V. (1975). *Pictures of facial affect*. Consulting Psychologists Press.

417 Fox, J. (2010). *Bayesian item response theory*. New York: Springer.

418 Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for  
419 Likert-type and ordinal item response data: A conceptual, empirical, and practical guide.  
420 *Practical Assessment, Research & Evaluation*, 17(3), 1–13.

421 Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple  
422 sequences. *Statistical science*, 457–472.

423 Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation*  
424 *of posterior moments* (Vol. 196). Federal Reserve Bank of Minneapolis, Research Department  
425 Minneapolis, MN, USA.

426 Hallerbäck, M. U., Lugnegard, T., Hjärthag, F., & Gillberg, C. (2009). The Reading the Mind  
427 in the Eyes Test: test–retest reliability of a Swedish version. *Cognitive neuropsychiatry*, 14(2),  
428 127–143.

429 Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to  
430 underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453.  
431 <https://doi.org/10.1037/1082-989X.3.4.424>

432 Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure  
433 analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1  
434 –55. <https://doi.org/10.1080/10705519909540118>

435 Johnston, L., Miles, L., & McKinlay, A. (2008). A critical review of the eyes test as a  
436 measure of social-cognitive impairment. *Australian Journal of Psychology*, 60(3), 135–141.  
437 <https://doi.org/10.1080/00049530701449521>

438 Kirkland, R. A., Peterson, E., Baker, C. A., Miller, S., & Pulos, S. (2013). Meta-analysis  
439 Reveals Adult Female Superiority in "Reading the Mind in the Eyes Test". *North American*  
440 *Journal of Psychology, 15*(1).

441 Kruschke, J. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*.  
442 Elsevier Science. Consulté à l'adresse <http://books.google.co.uk/books?id=FzvLAWAAQBAJ>

443 Lord, F. M. (1980). *Applications of item response theory to practical testing problems*.  
444 Routledge.

445 Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*.

446 Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2 A Comprehensive Program for  
447 Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied*  
448 *Psychological Measurement, 37*(6), 497–498.

449 Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution,  
450 critique and future directions. *Statistics in Medicine, 28*(25), 3049-3067.  
451 <https://doi.org/10.1002/sim.3680>

452 McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press.

453 McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural  
454 equation analyses. *Psychological Methods, 7*(1), 64-82. [https://doi.org/10.1037/1082-](https://doi.org/10.1037/1082-989X.7.1.64)  
455 [989X.7.1.64](https://doi.org/10.1037/1082-989X.7.1.64)

456 Peterson, E., & Miller, S. F. (2012). The Eyes Test as a Measure of Individual Differences:  
457 How much of the Variance Reflects Verbal IQ? *Frontiers in Psychology, 3*.  
458 <https://doi.org/10.3389/fpsyg.2012.00220>

459 Pinkham, A. E., Penn, D. L., Green, M. F., Buck, B., Healey, K., & Harvey, P. D. (2014). The  
460 Social Cognition Psychometric Evaluation Study: Results of the Expert Survey and RAND  
461 Panel. *Schizophrenia Bulletin, 40*(4), 813-823. <https://doi.org/10.1093/schbul/sbt081>

462 Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using  
463 Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical*  
464 *computing* (Vol. 124, p. 125). Technische Universit at Wien.

465 Preti, A., Vellante, M., & Petretto, D. R. (2017). The psychometric properties of the “Reading  
466 the Mind in the Eyes” Test: an item response theory (IRT) analysis. *Cognitive*  
467 *Neuropsychiatry*, 22(3), 233-253. <https://doi.org/10.1080/13546805.2017.1300091>

468 R Development Core Team. (2014). *R: A Language and Environment for Statistical*  
469 *Computing*. Vienna, Austria: R Foundation for Statistical Computing. Consulté à l’adresse  
470 <http://www.R-project.org/>

471 Raven, J. C., & Deltour, J. (1998). *Echelle de vocabulaire Mill Hill*. Ed. et applications  
472 psychologiques.

473 Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of*  
474 *Statistical Software*, 48(2). Consulté à l’adresse  
475 [http://www.lce.esalq.usp.br/arquivos/aulas/2013/encontro\\_ppg/Lucia/paper.pdf](http://www.lce.esalq.usp.br/arquivos/aulas/2013/encontro_ppg/Lucia/paper.pdf)

476 Sanvicente-Vieira, B., Kluwe-Schiavon, B., Wearick-Silva, L. E., Piccoli, G. L., Scherer, L.,  
477 Tonelli, H. A., & Grassi-Oliveira, R. (2013). Revised Reading the Mind in the Eyes Test  
478 (RMET)-Brazilian version. *Revista Brasileira de Psiquiatria*, (AHEAD), 000–000.

479 Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for  
480 cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, 45(13), 3054-3067.  
481 <https://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2007.05.021>

482 Shamay-Tsoory, S. G., Shur, S., Barcai-Goodman, L., Medlovich, S., Harari, H., & Levkovitz,  
483 Y. (2007). Dissociation of cognitive from affective components of theory of mind in  
484 schizophrenia. *Psychiatry research*, 149(1), 11–23.

485 Sternberg, R. J., & Grigorenko, E. L. (2002). *The general factor of intelligence: How general*  
486 *is it?* Psychology Press.

487 Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti,  
488 A. (2013). The “Reading the Mind in the Eyes” test: systematic review of psychometric  
489 properties and a validation study in Italy. *Cognitive neuropsychiatry*, 18(4), 326–354.

490 Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient  
491 and the SEM. *The Journal of Strength & Conditioning Research*, 19(1), 231–240.

492 Yıldırım, E. A., Kaşar, M., Güdük, M., Ateş, E., Küçükparlak, I., & Ozalmete, E. O. (2011).  
493 Investigation of the reliability of the « reading the mind in the eyes test » in a Turkish  
494 population. *Turkish journal of psychiatry*, 22(3), 177-186.

495 Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models*  
496 *with binary and continuous outcomes*. University of California Los Angeles.

497 Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items: Pitfalls and  
498 loopholes. *European Journal of Psychological Assessment*, 31(4), 231-237.

499 <https://doi.org/10.1027/1015-5759/a000309>

500

501 **Table 1: 3-PL Bayesian analysis on 35 items, medians and 95 % credible interval of alpha coefficients**

English denomination	French number	2.50%	50%	97.50%
practice	Item 1	0.163	0.601	1.775
Item 1	Item 2	1.033	1.689	2.888
Item 2	Item 3	0.282	0.762	2.126
Item 3	Item 4	0.771	1.264	2.33
Item 4	Item 5	0.187	0.845	2.639
Item 5	Item 6	0.268	0.68	1.992
Item 6	Item 7	0.612	1.042	1.971
Item 7	Item 8	0.143	0.453	2.174
Item 8	Item 9	0.234	0.548	1.787
Item 9	Item 10	0.558	0.948	1.704
Item 11	<b>Item 11</b>	<b>-0.181</b>	<b>0.234</b>	<b>1.957</b>
Item 12	Item 12	0.326	0.656	1.317
Item 14	Item 13	0.343	0.862	2.276
Item 15	Item 14	0.589	0.949	1.648
Item 16	Item 15	0.396	0.911	2.351
Item 18	Item 16	0.205	0.78	1.621
Item 20	Item 17	0.516	1.299	2.909
Item 21	Item 18	1.185	1.98	3.269
Item 22	Item 19	0.4	1.025	2.452
Item 23	<b>Item 20</b>	<b>-0.129</b>	<b>0.16</b>	<b>1.515</b>
Item 24	Item 21	0.082	0.34	1.168
Item 25	Item 22	0.332	0.627	1.268
Item 26	Item 23	0.689	1.063	1.748
Item 27	<b>Item 24</b>	<b>-0.29</b>	<b>0.096</b>	<b>2.249</b>
Item 28	Item 25	0.233	0.724	2.274
Item 29	Item 26	0.405	1.069	2.595
Item 30	Item 27	0.729	1.155	2.082
Item 31	Item 28	0.121	0.49	1.942
Item 32	Item 29	0.134	0.425	1.317
Item 34	Item 30	0.026	0.471	2.254
Item 36	Item 31	0.073	0.568	1.582
Item 13	Item 32	0.425	0.937	2.104
Item 33	Item 33	0.415	0.824	1.869
Item 36	<b>Item 34</b>	<b>-0.051</b>	<b>0.211</b>	<b>2.246</b>
child version#	Item 35	0.823	1.296	2.197

502  
503 In bold, the 4 items with credible intervals including 0 and excluded from the final version of  
504 the French Eyes-Test-31; with the correspondence with English item numbers  
505

506



507 **Table 2. Percent of responses for targets and distractors in the Validation of the French Eyes Test-31**

508

English denomination	French number	Target	Foil 1	Foil 2	Foil 3
practice	1	81.9	7.5	6.2	4.4
Item 1	2	93.1	2.9	2.7	1.3
Item 2	3	84.0	8.7	3.7	3.7
Item 3	4	77.7	17.7	3.3	1.3
Item 4	5	82.9	8.7	5.8	2.7
Item 5	6	76.9	12.1	10.4	0.6
Item 6	7	67.9	13.3	11.9	6.9
<b>Item 7</b>	<b>8</b>	<b>47.5</b>	<b>26.3</b>	<b>23.7</b>	<b>2.5</b>
<b>Item 8</b>	9	59.0	<b>31.7</b>	6.9	2.3
Item 9	10	87.1	7.1	5.2	0.6
Item 12	12	77.3	17.1	4.0	1.5
Item 14	13	88.5	8.7	1.5	1.3
Item 15	14	81.3	9.8	4.8	4.0
Item 16	15	80.6	15.0	2.5	1.9
Item 18	16	91.3	4.0	3.1	1.5
Item 20	17	92.3	4.0	3.7	0.0
Item 21	18	97.3	1.7	1.0	0.0
Item 22	19	91.0	4.4	3.8	0.8
Item 24	21	56.7	16.2	15.2	11.9
Item 25	22	63.5	18.1	10.2	8.3
Item 26	23	78.3	12.5	6.0	3.3
Item 28	25	81.9	12.1	3.7	2.3
<b>Item 29</b>	<b>26</b>	<b>44.8</b>	<b>33.1</b>	<b>16.5</b>	<b>5.6</b>
Item 30	27	80.4	7.3	6.7	5.6
Item 31	28	66.7	16.2	10.2	6.9
Item 32	29	66.5	23.7	7.5	2.3
Item 34	30	78.8	12.1	5.0	4.0
Item 36	31	88.7	8.5	2.9	0.0
Item 13	32	89.2	6.2	4.0	0.6
Item 33	33	70.8	22.1	4.0	3.1
child version#	35	76.5	10.0	8.5	5.0

509

510 \* In boldface, items with less than 50 % of correct responses or with higher than 25 % of  
 511 distractor choice

512 # item adapted from the child version of the English Eyes Test

513

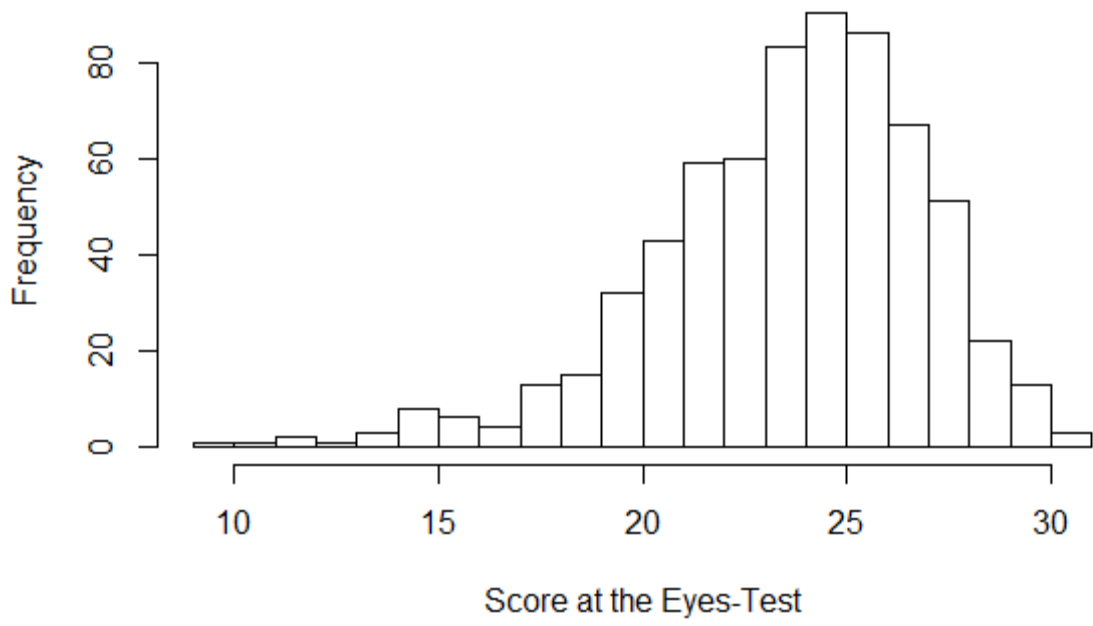
514 **Table 3. Results of the Bayesian Item Response Model of the French Eyes Test-31, quantiles of the**  
515 **distribution of guessing parameters**

Item	2.50%	25%	50%	75%	97.50%
Item 1	<b>0.373</b>	0.591	0.657	0.713	0.802
Item 2	0.107	0.48	0.634	0.733	0.843
Item 3	<b>0.325</b>	0.577	0.657	0.72	0.803
Item 4	0.042	0.261	0.395	0.498	0.634
Item 5	<b>0.495</b>	0.666	0.727	0.777	0.83
Item 6	0.212	0.459	0.55	0.628	0.726
Item 7	0.035	0.214	0.328	0.421	0.548
Item 8	0.005	0.063	0.154	0.315	0.449
Item 9	0.026	0.17	0.278	0.392	0.546
Item 10	0.074	0.377	0.526	0.635	0.765
Item 12	0.103	0.366	0.478	0.567	0.69
Item 13	<b>0.349</b>	0.617	0.698	0.756	0.825
Item 14	0.042	0.271	0.415	0.532	0.677
Item 15	<b>0.278</b>	0.54	0.631	0.698	0.775
Item 16	<b>0.454</b>	0.721	0.792	0.838	0.892
Item 17	<b>0.564</b>	0.765	0.82	0.857	0.901
Item 18	0.106	0.514	0.687	0.794	0.899
Item 19	<b>0.482</b>	0.73	0.795	0.839	0.889
Item 21	0.011	0.087	0.157	0.253	0.473
Item 22	0.031	0.189	0.299	0.4	0.553
Item 23	0.017	0.15	0.268	0.383	0.547
Item 25	<b>0.32</b>	0.534	0.612	0.68	0.757
Item 26	0.029	0.199	0.287	0.341	0.408
Item 27	0.047	0.286	0.429	0.542	0.685
Item 28	0.101	0.292	0.386	0.493	0.626
Item 29	0.032	0.173	0.264	0.362	0.54
Item 30	<b>0.344</b>	0.51	0.576	0.662	0.752
Item 31	<b>0.519</b>	0.706	0.756	0.795	0.861
Item 32	<b>0.324</b>	0.634	0.725	0.784	0.854
Item 33	0.048	0.257	0.378	0.475	0.595
Item 35	0.018	0.151	0.263	0.368	0.522

516

517

518 Figure 1. Distribution of total scores on the Validation of the French Eyes-Test 31 (N = 661)



519

520

521 Appendix 1: Words used in the Validation of the French Eyes Test-31

English denomination	French number	Choice 1	Choice 2	Choice 3	Choice 4
practice	1	jaloux	<b>paniqué</b>	arrogant	haineux
Item 1	2	<b>malicieux</b>	consolateur	irrité	ennuyé
Item 2	3	terrifié	<b>bouleversé</b>	arrogant	agacé
Item 3	4	plaisantant	troublé	<b>charmeur</b>	convaincu
Item 4	5	plaisantant	<b>insistant</b>	amusé	détendu
Item 5	6	agacé	sarcastique	<b>soucieux</b>	amical
Item 6	7	consterné	<b>rêveur</b>	impatient	inquiet
Item 7	8	se justifiant	amical	<b>mal à l'aise</b>	déprimé
Item 8	9	<b>découragé</b>	soulagé	timide	impatient
Item 9	10	agacé	hostile	horrifié	<b>préoccupé</b>
Item 12	12	indifférent	embarrassé	<b>sceptique</b>	déprimé
Item 14	13	irrité	déçu	abattu	<b>accusateur</b>
Item 15	14	<b>contemplatif</b>	confus	encourageant	amusé
Item 16	15	irrité	<b>pensif</b>	encourageant	compatissant
Item 18	16	<b>décidé</b>	amusé	consterné	ennuyé
Item 20	17	impatient	<b>amical</b>	coupable	horrifié
Item 21	18	embarrassé	<b>charmeur*</b>	confus	paniqué
Item 22	19	<b>préoccupé</b>	reconnaisant	insistant	implorant
Item 24	21	<b>méditatif</b>	irrité	impatient	hostile
Item 25	22	paniqué	incrédule	découragé	<b>intéressé</b>
Item 26	23	alarmé	timide	<b>hostile</b>	anxieux
Item 28	25	<b>intéressé</b>	plaisantant	affectueux	satisfait
Item 29	26	impatient	consterné	irrité	<b>méditatif</b>
Item 30	27	reconnaisant	<b>flirtant</b>	hostile	déçu
Item 31	28	honteux	<b>confiant</b>	plaisantant	abattu
Item 32	29	<b>grave</b>	honteux	perplexe	alarmé
Item 34	30	consterné	déconcerté	<b>méfiant</b>	terrifié
Item 36	31	honteux	nerveux	<b>soupçonneux</b>	indécis
Item 13	32	décidé	<b>espérant</b>	menaçant	arrogant*
Item 33	33	satisfait	coupable	rêveur	<b>préoccupé</b>
child version#	35	haineux*	nerveux*	plaisantant*	<b>charmeur*</b>

522

523

524 \*: word meaning changed from the original English version (Baron-Cohen et al., 2001) for  
525 psychometric reasons.

526 #: item from the English child version of the Eyes Test with new words (targets and  
527 distractors)

528 Words in boldface are target words.