Available online at www.sciencedirect.com
SCIENCE DIRECT®
Brain and Language xxx (2004) xxx–xxx
Brain and Language
www.elsevier.com/locate/b&l

Normative data for the Boston Naming Test in native Dutch-speaking Belgian children and the relation with intelligence☆

Gert Storms,[a,*] Jos Saerens,[b] and Peter P. De Deyn[b]

[a] *Department of Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium*
[b] *Laboratory of Neurochemistry and Behavior, Department of Biomedical Sciences, General Hospital Middelheim, ZINA, University of Antwerp, Belgium*

Accepted 24 March 2004

Abstract

This paper reports the results of a normative study of the 60-item version of the Boston Naming Test (BNT) in a group of 371 native Dutch-speaking Flemish children between the ages of 6 and 12 years. Analysis of test results revealed that BNT performance was significantly affected by age and gender. The gathered norms were shown to be significantly lower than published norms for comparable North-American children. Error analysis disclosed remarkable similarities with data from elderly subjects, with verbal semantic paraphasias and 'don't know' responses occurring most frequently. Finally, BNT scores were shown to correlate strongly with general intelligence as measured with the Raven Progressive Matrices. The relation between both measures can be of help in the diagnosis of identification naming deficits and impaired word-retrieval capacities.

Keywords: Boston Naming Test; Children Raven Progressive Matrices; Intelligence; Picture naming; Neurolinguistic error analysis

1. Introduction

The revised version of the Boston Naming Test (BNT) was introduced by Kaplan, Goodglass, and Weintraub (1983). The test consists of a visual picture naming task in which 60 outline drawings of objects and animals are presented. The items are presented in order of word frequency and difficulty.

The test has shown to be a highly sensitive tool to identify naming deficits and impaired word-retrieval capacities in adults and children. It was used in the study of a variety of cerebral pathologies in adults (e.g., Nicholas, Obler, Au, & Albert, 1996). For an overview of studies with adults and elderly where the BNT was used to study brain deficits, see Mariën, Mampaey, Vervaet, Saerens, and De Deyn (1998).

The Boston Naming Test has also been used extensively in neuropsychological studies of children with brain injuries due to intoxication (Yeates & Mortensen, 1994), tumors and acute lymphoblast leukemia (Hudson & Murdoch, 1992; Jordan, Murdoch, Hudson-Tennent, & Boon, 1996), closed head injury (Jordan, Cannon, & Mrdoch, 1992), attention deficit hyperactivity disorder and developmental language disorder (Semrud-Clikeman, Guy, Griffin, & Hynd, 2000; Weyandt & Willis, 1994), and Turner syndrome (Ross, Roeltgen, Feuillan, Kushner, & Cutler, 2000). The test has also been applied to explore the relations between oral and written language errors in studies with language disabled children (Rubin & Liberman, 1983; Wolf, 1984).

The vast majority of the published studies with the BNT in adults and in children were conducted with native-English speaking North-American adult participants (see, e.g., Mariën et al., 1998). Among the few exceptions are studies from Australia (Worrall, Yiu, Hickson, & Barnett, 1995), Switzerland (Thuillard-Colombo & Assal, 1992), Italy (Riva, Nichelli, & Devoti, 2000), Columbia (Rosselli, Ardila, Bateman, & Guzmán, 2001), and Belgium (Mariën et al., 1998).

In the studies cited above, several different versions of the BNT were used. While older studies usually used the

☆ The authors thank Annick Myszta, Liesbeth Van Oevelen, and Ellen Benaets for their help in gathering the data.
* Corresponding author. Fax: +32-16-325-916.
E-mail address: gert.storms@psy.kuleuven.ac.be (G. Storms).

0093-934X/$ - see front matter © 2004 Published by Elsevier Inc.
doi:10.1016/j.bandl.2004.03.005

YBRLN 3059 ARTICLE IN PRESS No. of pages: 8
DISK / 3/4/04 / Moses(CE) / Panneer (TE) DTD 4.3.1 / SPS

experimental 85-item version of the test or shortened (15-item) versions of the test, most recent publications on the BNT used the 60-item version (Kaplan et al., 1983).

Published norm studies, even with identical versions of the BNT, differ considerably for similar age groups due to cultural and language differences. Norm studies for North-American children from five different studies with moderately sized samples, conducted between 1983 and 1992, for instance, have been shown to result in significantly different averages (Yeates, 1994). Likewise, Mariën et al. (1998) showed significant differences in the average scores for comparable age groups in four different studies. Such findings illustrate the necessity to gather data that are culture and language specific, from large groups of participants.

Furthermore, it should be mentioned that 'norms' are not a static but a dynamic issue. They may change over time as has been repeatedly shown in intelligence tests such as the Raven Progressive Matrices and the Wechlser Adult Intelligence Scale (e.g., Flynn, 1984). Likewise, the results of visual naming tasks may be influenced by the fact that some of the represented objects have come in disuse or have changed in form. Furthermore, there may be the occurrence of a general rise of 'verbality' in a young population due to the availability of better examples of language use in television programs and young children's books.

1.1. The relation between the BNT and intelligence

Several studies have related scores on the BNT to other language-related measures, such as verbal fluency (Ardila & Rosselli, 1994; Halperin, Healey, Zeitchik, Ludman, & Weinstein, 1989; Riva et al., 2000), and to measures of reading disorders (Wolf, 1984). The BNT has also been related to memory and visuo-spatial abilities (Ardila & Rosselli, 1994; Halperin et al., 1989; Rosselli et al., 2001). To our knowledge, none of the published articles that used the BNT with normal children related the test scores to intelligence. However, in situations where a diagnosis concerning identification naming deficits and impaired word-retrieval capacities has to be made, the results of intelligence testing prior to the problems are often available. If the relation between intelligence and the BNT scores for a representative sample is known, then the BNT score that corresponds best with the intelligence score can be estimated and compared with the observed BNT score. Whenever the observed score is considerably below the predicted score, it is possible that the low observed score is not attributable to slow language development in the testee, but instead is indicative of impaired identification naming or word-retrieval.

Intelligence can be measured with a wide variety of tests and the choice of the test often depends on practical issues, such as collective versus individual administration, availability of recent norms and clinical time constraints. However, for the issue outlined above, there is an additional important concern. Even in cases where no pre-morbid intelligence measure is available, it may be interesting to test a patient's present intelligence level with a test that is not too much influenced by language abilities. Given these considerations, the Standard Progressive Matrices test (Raven, Court, & Raven, 1977) seems a good choice since the test is widely used (and therefore, appropriate norms are easily available), has been shown to be reliable (Raven et al., 1977) and its construct validity has been documented extensively. The test was constructed to test general intelligence (Spearman's *g*). Carroll (1993) showed that the test indeed measures general intelligence, and loads on fluid intelligence and inductive and spatial ability factors in large scale factor analyses.

The present paper describes a study that addressed several goals. First, we wanted to gather normative data for children of primary-school ages in Dutch-speaking Belgium. As argued above, due to linguistic and cultural differences, the normative data from North-American populations cannot be merely relied on in Belgium and the Netherlands. Therefore, a comparison of the Belgian and the North-American data from normally developing children was conducted. Second, a Dutch BNT error profile was deduced from the responses, which may instigate further research on aberrant naming behavior characteristics in children with traumatic brain injuries and other language development disorders. It is perfectly possible that a person obtains a test scores within the normal range, but that the exact error profile shows a language pathology (e.g., if virtually all errors are of a phonological nature). We are currently gathering data from a large group of children with several types of language development disorders. It is our aim to not only compare their BNT scores to the norms presented below, but to also compare the types of errors that both groups make. Third, the relation between the results of the BNT and non-verbal intelligence was investigated, which may be helpful in diagnosing whether or not deviant test scores are due to lower intellectual capability and therefore to slower vocabulary growth or to a language impairment.

2. Method

2.1. Participants

Three hundred and seventy-one native Dutch-speaking Belgian children from four different primary schools participated in our study. One hundred and seventy-four of the participants were male (47%) and 196 female (53%). Their age ranged from 6 to 12, with an average of

8.8 and $SD = 1.79$. Two of the schools were located in rural areas (Nijlen and Godsheide), one school was located in a city (Hasselt), and the fourth school was located in a metropolitan area (Merksem, Antwerp). The number of participants from the first to the sixth grade was 55, 52, 70, 49, 74, and 70, respectively. Only native speakers of Dutch participated in the study. The children were all screened for school readiness and for learning disabilities upon entrance in primary school. There were no indications whatsoever for abnormalities in any of those indices for any of the participants. Moreover, none of the children ever received special education services in the schools.

All participants completed the BNT. A subset of 135 participants (44 from the second grade, 41 from the fourth grade, and 47 from the sixth grade) also completed the Standard Progressive Matrices (Raven et al., 1977).

2.2. Administration of the BNT

All participants were tested individually with the BNT in a quite room in their school. Three last-year psychology students administered all test. For the purpose of the current normative study with children, the original procedure for the administration of the 60-item version of the BNT was altered. Standard BNT administration for non-aphasic adult subjects and older children, who may be expected to have no failures for the easier items of the test, starts with item 30 and, unless a failure is encountered before item 38, the test is continued. In case of an error before item 38, item 29 is presented and the test is continued backwards until eight consecutive items are named correctly without help. For the items preceding 'basal' level, full credit is allowed. Testing stops after six consecutive failures. In the current study, all participants were presented with the 60 items in the standard order and they were asked to name the object or animal that was pictured. Stimulus cues, supplying either phonemic or semantic information, were omitted and no time limits were imposed for responding. All responses obtained on the BNT items were recorded verbatim and orthographically or phonetically transcribed.

2.3. Administration of the SPM

The Standard Progressive Matrices (Raven, 1996) was administered in accordance with the instructions for group administration in a large classroom by a female research assistant and the regular school teacher of the children was present during testing. Standard instructions, as suggested in Raven et al. (1977), were given. The participants went through the test items at their own pace and there was no time limit imposed. The participants indicated their responses in the answer forms using a pencil. Completion of the test took approximately 1 h.

2.4. Target answers of the BNT

To determine the lexical appropriateness of responses, the same procedure as used by Mariën et al. (1998) was implemented. Thus, the English target words of the scoring booklet were translated using a standard English–Dutch dictionary (van Dale, 1984). In particular with respect to item 51, for which the English word 'latch' is proposed in the booklet, translation difficulties were encountered. English–Dutch dictionaries systematically translate the word 'latch' by 'klink' (door handle), an inappropriate, semantically related variant for the object pictured in item 51. We agreed on accepting only the semantically more closely related Dutch word 'grendel' for this item, which is translated in English as 'bolt' or 'bar.' Translations were further checked in a Dutch dictionary (van Dale, 1985) and in the official Dutch wordlist (Woordenlijst van de Nederlandse Taal, 1954). In this way, a list of Dutch synonyms matching the BNT representations was obtained. The list of accepted standard Dutch responses can be found in the Appendix of Mariën et al. (1998). Dialectic variants of the target words were not considered erroneous.

On the qualitative level, we applied the same system to classify errors as the one developed by Mariën et al. (1998). The error types considered in this classification system are: (1) phonemic or literal errors, which are alterations of the target word through addition, omission, substitution, or transposition of the phonemes (e.g., 'te*s*toscoop' for 'stetoscope' (stethoscope)); (2) verbal morphological errors, representing erroneous lexical items that highly resemble the target words with respect to their phonological structure (e.g., 'telescoop' for 'stetoscoop'); (3) verbal semantic errors, which are erroneous words that semantically relate to the target word (e.g., 'hoorapparaat' (hearing-aid) for 'stethoscope'); (4) verbal or unrelated errors, which are erroneous words that share no visual, phonological, or conceptual characteristics with the target item (e.g., 'rekker' (elastic band) for 'juk' (yoke)); (5) aborted words, that is, phonologically or morphologically incomplete responses (e.g., 'zee' (sea) for 'zeepaardje' (seahorse)); (6) non-words or neologisms, which do not belong to the lexicon (e.g., 'simek' for 'sfinx' (sphinx)); (7) portmanteau words or semantic neologisms, which consist of a neologistic combination of lexical elements that have a meaning on their own (e.g., 'deursluitingsklink' (doorlockinglatch) for 'klink' (latch)); (8) preservations or preservative errors, which are recurrent responses that lack any resemblance with the target item presented; (9) delayed responses which are chronologically wrong answers that relate to an item presented before and for which the right word was still lacking,

(10) ‘don’t know responses,’ which represent the absence of any response or the expression for not knowing the right target word; (11) non-specific utterances, which include empty words, interjections, and generalizations and which have no determining quality (e.g., ‘dinges’ (what-d’you-call-it)); (12) adequate circumlocutions, which are conceptually correct descriptions of the target item (e.g., ‘an instrument used by a doctor to listen to your heart’ for ‘stetoscoop’); (13) inadequate circumlocutions, which represent semantically inadequate or conceptually incomplete description (e.g., ‘a machine for hearing’ for ‘stetoscoop’); (14) foreign words, which correspond to the replacement of the target word by a correct alternative from another language (e.g., the French word ‘compas’ for ‘passer’ (compass)); and (15) mere visual misperceptions (e.g., ‘beker’ (drinking cup) for ‘masker’ (mask)) (Mariën et al., 1998).

3. Analysis and results

3.1. Statistical methods

Means and standard deviations were calculated. Means were compared between groups using analysis of variance. The degree of association between two variables was evaluated using Pearson product-moment correlation coefficient and linear regression analysis. For all tests a p value of less than .05 was considered to indicate a statistically significant difference.

3.2. BNT results

The split-half reliability of the BNT scores was estimated by calculating the total correct scores on the even items and on the uneven items and by applying the Spearman–Brown formula on the correlation between the half test scores (Lord & Novick, 1968). Estimates were .67, .85, .73, .76, .66, and .78, for grades 1 through 6, respectively. The reliability within the total group of 371 participants was .88.

The overall group mean on the BNT in the sample of 371 participants was 37.6 total correct with a range of 16–54 and $SD = 7.52$. The overall mean age was 8.8 years ($SD = 1.79$). The average score for the female participants was 37.31 ($SD = 7.13$) with scores ranging from 16 to 54, and for the male participants it was 37.91 ($SD = 7.84$) with scores ranging from 22 to 54.

An analysis of variance, with the BNT scores as dependent variable and the six grades of primary school as the independent variable yielded a significant grade effect, $F(5, 364) = 97.12$, $p < .001$. A further test showed a significant increasing linear trend in the data, $F(1, 364) = 483.66$, $p < .001$.

In an analyses in which both gender and grade functioned as independent variables, the grade variable yielded significance again, $F(5, 358) = 98.63$, $p < .001$, and also the gender effect was significant, $F(1, 358) = 8.08$, $p < .01$. The interaction effect of gender and grade was not significant ($F(5, 358) = 0.75$). Table 1 shows the mean scores, the standard deviations, the score range, and the median value within each age group between 6 and 12.

Recommended cut-off scores according to grade and gender for the BNT norms developed on our sample of native Dutch-speaking primary school children are presented in Table 1. The cut-off scores were calculated as two standard deviations below the mean of the BNT score for that group.

3.3. A comparison with North-American norms

Yeates (1994) and Mariën et al. (1998) presented data which show that BNT norms may differ significantly over populations, languages, and cultures. Yeates combined the norms of five published studies with North-American children between the ages of 5 and 13 and combined the norms of the different samples to

Table 1
Mean scores, standard deviations, score range, and median value for males and females for different age groups

Age	Gender	Mean	Standard deviation	Range	Median value	Cut-off score
6	Female ($N = 27$)	27.04	2.90	21–32	27	21.24
	Male ($N = 27$)	29.56	4.53	22–41	29	20.50
7	Female ($N = 20$)	30.55	6.38	16–41	31.5	17.81
	Male ($N = 22$)	33.00	4.80	26–41	33	23.40
8	Female ($N = 38$)	34.00	6.21	18–47	34	21.58
	Male ($N = 35$)	35.89	5.14	25–46	36	25.61
9	Female ($N = 18$)	40.39	3.88	32–49	40	32.63
	Male ($N = 26$)	40.12	5.29	30–53	39	29.54
10	Female ($N = 50$)	41.26	4.65	30–49	42	31.96
	Male ($N = 28$)	43.18	4.89	30–54	43.5	33.40
11	Female ($N = 30$)	43.37	6.01	30–54	44	31.35
	Male ($N = 31$)	43.52	4.91	31–52	44	33.70
12	Female ($N = 13$)	45.23	3.85	40–51	44	37.53
	Male ($N = 3$)	48.00	3.00	45–51	48	42.00

obtain averages (and standard deviations) for large samples per age category (with N up to 167). The resulting averages for the corresponding age groups were higher than the average values in our Dutch-speaking Belgian sample. For ages 6–12, Yeates reported means equal to 33.6, 36.9, 39.0, 41.7, 45.1, 46.8, and 48.6, respectively. The corresponding values (averaged over the two gender groups) in our sample were 28.3, 31.8, 34.9, 40.2, 41.9, 43.4, and 45.8. T tests showed that the North-American subjects scored significantly better in all seven age groups ($p < .01$).

3.4. BNT error analysis

The total number of correct responses was also counted for each of the 60 items (summed over the 371 participants). Construct validity of the test was evaluated by estimating the reliability of the percentages of correct responses of the different items. It was calculated using the split-half method, combined with the Spearman–Brown formula (Lord & Novick, 1968). The resulting reliability of these percentages was .998. Table 2 presents the percentage of correct responses per item.

The mean percentage of correct responses for all grades was 62.7% ($SD = 34.67$, range 16–100%). Twenty-four (out of 60) items yielded percentages of correct responses below 50%. The results were recalculated to evaluate a possible effect if, according to the original standard instructions (for adult participants), a starting point would have been used. Calculation of the scores according to the standard procedures (see above) resulted in a significantly higher average as compared to the average score when the participants were presented with all 60 items: respectively, 38.16 vs 37.59 ($t(370) = 13.70$, $p < .01$). The scores calculated according to both procedures correlate extremely high, though ($r = .995$).

On the qualitative level of analysis, 8291 erroneous responses (37.3% of all answers) were classified according to a 15-item neurolinguistic taxonomy (see also Mariën et al., 1998). Most responses could be classified unambiguously within one taxonomic category. To some responses, however, more than one single error type could be assigned. All ambiguities were resolved using the same standards as used in Mariën et al. (1998). Table 3 presents the incidence of the 15 error types, as well as the percentages, and the number of different items for which this type of error occurred. Two error types constitute the vast majority of errors in the corpus, namely 'verbal semantic paraphasias' and 'don't know responses.' Four error types account for 5–10% of the errors ('visual paraphasias,' 'adequate circumlocutions,' 'inadequate circumlocutions,' and 'portmanteau words'), and three other error types ('verbal paraphasias,' 'aborted words,' and 'phonemic paraphasias') accounted each for between 5 and 1% of the errors. The six remaining error types ('verbal morphological paraphasias,'

Table 2
Total number correct and percentage correct per item

Item	Target word	Total correct	%
1	Bed	369	99.5
2	Tree	370	100
3	Pencil	369	99.5
4	House	369	99.5
5	Whistle	367	99
6	Scissors	369	99.5
7	Comb	370	100
8	Flower	365	98.5
9	Saw	364	98.5
10	Toothbrush	366	99
11	Helicopter	349	94.5
12	Broom	334	90
13	Octopus	269	72.5
14	Mushroom	369	99.5
15	Hanger	343	92.5
16	Wheelchair	353	95.5
17	Camel	342	92.5
18	Mask	355	96
19	Pretzel	22	6
20	Bench	367	99
21	Racquet	342	92.5
22	Snail	370	100
23	Volcano	290	78.5
24	Seahorse	309	83.5
25	Dart	244	66
26	Canoe	354	96
27	Globe	348	94
28	Wreath	306	83
29	Beaver	246	67
30	Harmonica	156	42
31	Rhinoceros	312	84.5
32	Acorn	264	71.5
33	Igloo	303	82
34	Stilts	174	47
35	Dominoes	227	61.5
36	Cactus	303	82
37	Escalator	302	81.5
38	Harp	151	41
39	Hammock	259	70
40	Knocker	42	11.5
41	Pelican	154	41.5
42	Stetoscope	79	21.5
43	Pyramid	238	64.5
44	Muzzle	82	22
45	Unicorn	145	39
46	Funnel	111	30
47	Accordion	143	38.5
48	Noose	66	18
49	Asparagus	14	4
50	Compass	178	48
51	Latch	15	4
52	Tripot	17	4.5
53	Scroll	112	30.5
54	Tongs	24	6.5
55	Sphinx	87	23.5
56	Yoke	6	1.5
57	Trellis	78	21
58	Palette	23	6
59	Protractor	86	23
60	Abacus	168	45.5

Table 3
Different error types according to incidence, percentage, and spreading

Type of error	Incidence	%	Number of different items	Highest percentage (item number)
Verbal semantic paraphasias	2831	34.1	52	73 (51)
Don't know responses	2445	29.5	48	49.5 (56)
Visual paraphasias	663	8.0	31	73.5 (49)
Adequate circumlocutions	503	6.1	36	15.6 (42)
Inadequate circumlocutions	713	8.6	37	38.4 (58)
Portmanteau word	518	6.2	44	23.5 (58)
Verbal morphological paraphasias	50	0.6	9	7.6 (42)
Verbal paraphasias	202	2.4	33	5.9 (46)
Aborted words	138	1.7	24	6.2 (37)
Phonemic paraphasias	148	1.8	28	11.6 (57)
Foreign words	6	0.1	3	0.8 (45)
Neologisms	27	0.3	20	0.8 (58,54)
Empty words	30	0.4	21	1.1 (47)
Perseverations	11	0.1	11	0.3 (*)
Delayed responses	5	0.05	5	0.3 (**)

Note. (*) 5, 41, 42, 44, 45, 50, 52, 54, 57, 59, and 60; (**) 37, 41, 44, 56, and 57.

'foreign words,' 'neologisms,' 'empty words,' 'delayed responses,' and 'perseverations') accounted for less than 1% of the errors.

As further illustrated in Table 3, 'verbal semantic paraphasias' errors occur were most common overall (occurring on 52 different items), followed by 'don't know responses (on 48 items) and 'portmanteau words' (on 44 items).

Percentages of error types were also calculated per item. The item that generated the highest percentage of errors per category was selected. For instance, as shown in Table 3, item 51 ('grendel'), representing a latch, triggered the most mistakes (73%) within the class of verbal semantic errors, while item 56 ('juk'), showing a yoke, generated the most 'don't know' answers (56%) of all errors.

3.5. Predicting the BNT scores from the standard progressive matrices scores

The Standard Progressive Matrices (SPM; Raven, 1996) was administered to 135 of the participants in our study sample. The average raw SPM score (sum score across items) in our sample was 39.70 ($SD = 8.76$; range: 15–54). The average score in the second grade was 32.34 ($SD = 9.07$; range: 15–47). In the fourth grade, the average score was 40.64 ($SD = 6.01$; range: 23–52). Finally, the average SPM score in the sixth grade was 45.85 ($SD = 4.60$; range 32–54).

Within our sample, the correlations of the test scores of the SPM with the BNT and with the grade were .62 and .63, respectively (both $p < .0001$). The correlation between the SPM and the BNT scores after statistically controlling for the grade is still highly significant ($r = .3214$, $p < .001$). Furthermore, the CPM scores were also significantly correlated ($p < .05$) with the number of errors classified as verbal semantic paraphasias (−.43), visual paraphasias (−.43), inadequate circumlocutions (−.37), aborted words (−.32), and don't know responses (−.29).

A linear regression was fitted with the BNT scores as criterion variable and the CPM as predictor variable. The resulting R^2 value was .38 and the optimal regression equation was

$$\text{Predicted BNT} = 17.902 + 0.523 \times \text{Raw CPM score}.$$

The standard estimation error equaled 6.896. Given the sufficiently high relation between the CPM and the BNT scores, the above described regression equation can be of help in making an appropriate diagnosis in case of suspiciously low BNT scores. If the predicted BNT score is lower than 2.0 standard estimation error below the predicted BNT score based on the corresponding CPM score, careful investigation of the language abilities of the tested subject seems warranted.[1]

4. Discussion

Over the past two decades, several papers have been published which presented normative data for the BNT performances in children of primary school age (see, e.g., Lezak, 1995; Spreen & Strauss, 1991). However, it has been shown that significant differences in the average scores have been reported for comparable age groups in different studies, for instance, in the elderly (Mariën et al., 1998; Yeates, 1994). The latter findings clearly show the necessity to gather data that are culture

[1] Note that an empirically obtained scores that is lower than 2.0 standard estimation error below the predicted BNT score based on the corresponding CPM score is not at all a proof for the presence of a language disorder. We only argue that closer inspection is warranted in such cases.

and language specific, from sufficiently large groups of participants. Furthermore, the normative data should not be regarded as static. They may change over time because culture is in a constant evolution with the continuous introduction of new concepts and words while others may become in disuse.

Since there exist no Dutch BNT norms for children, the exact diagnostic validity of this frequently applied test remains essentially unknown for native Dutch language users in that age category. This study provides the first normative BNT data for 371 Dutch-speaking Belgian primary school children. Moreover, a comparison with published norms for similarly aged North-American children (Yeates, 1994) showed that the norms differed significantly, yielding further evidence for the necessity of the availability of language and culture specific normative data. The reason of the relatively lower performance of Belgian children is unclear, but it may reflect cultural differences in the exposure to the items represented on the BNT, which was developed by North-American authors.

The standard instructions, where the first item that is presented is Item 30, were shown to yield systematic overestimations as compared to a testing procedure where all 60 stimuli are presented. This finding argues for an administration of the test that starts with Item 1 instead, at least for a population of children between 6 and 12 years old.

The gathered data were shown to yield significant effects of age and gender on BNT performance. In line with previously published norms for children of other language communities (e.g., Ardila & Rosselli, 1994; Cohen, Town, & Buff, 1988; Guilford & Nawojczyk, 1988; Halperin et al., 1989; Kaplan et al., 1983; Kirk, 1992; Riva et al., 2000; Yeates, 1994), the age effects showed the expected (nearly linear) increase from ages 6 to 12. Previous norm studies have also reported significant sex differences, with boys performing better than girls. Our own data are again perfectly in line with these findings.

Unsurprisingly, the norms for 12-year-old children are considerably lower than the norms for elderly from the same cultural and language community (Mariën et al., 1998). However, the large similarities in the error profiles, based on a neurolinguistic classification system (see Mariën et al., for details) are less self-evident. Verbal semantic paraphasias were the most common error type in both the elderly and in children. Also, 'don't know responses' accounted for a large percentage of the errors in both groups.

General fluid intelligence scores, measured with the Raven Progressive Matrices, were correlated with the BNT scores. Since both measures showed a strong linear relation, the optimal linear regression to predict BNT scores from the RPM scores was presented. This prediction can be of use in the diagnosis of identification naming deficits and impaired word-retrieval capacities, in cases where premorbid intelligence scores are available. The prediction can most likely only be used in a Dutch-speaking Belgian population and cannot be generalized to other populations without replication. Moreover, even within this exact population, the regression equation most be used with caution, since its validity has not been verified yet in an independent sample.

The SPM scores were also shown to correlate significantly with several error types: verbal semantic paraphasias, visual paraphasias, inadequate circumlocutions, aborted words, and don't know responses. The correlation between CPM results and visual paraphasias is not unsurprising because both tasks carry a highly visual processing load. Inadequate circumlocutions, and don't know responses on the other hand represent the fact that the object is not known. These types of errors therefore may have a high conceptual load which may correlate with 'general intelligence.' The same may partially be true for verbal semantic errors.

We believe that the norms, the error analysis, and the relation between the BNT scores and intelligence may help psychologists in the field to diagnose children with language development problems. Research in progress, in which similar measurements from different sorts of language disabled children is compared with the above described norms, will further elaborate on this matter.

5. Uncited references

Goodglass and Kaplan (1982), Lecours and Lhermitte (1979).

References

Ardila, A., & Rosselli, M. (1994). Development of language, memory, and visuospatial abilities in 5- to 12-year-old children using a neuropsychological battery. *Developmental Neuropsychology, 10*, 97–120.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Cohen, M., Town, P., & Buff, A. (1988). Neurodevelopmental differences in confrontational naming in children. *Developmental Neuropsychology, 4*, 75–81.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932–1978. *Psychological Bulletin, 95*, 29–91.

Goodglass, H., & Kaplan, E. (1982). *The assessment of aphasia and related disorders*. Philadelphia: Lea and Febiger.

Guilford, A. M., & Nawojczyk, D. C. (1988). Standardization of the Boston Naming Test at the kindergarten and elementary school levels. *Language, Speech, and Hearing Services in Schools, 19*, 395–400.

Halperin, J. M., Healey, J. M., Zeitchik, E., Ludman, W. L., & Weinstein, L. (1989). Developmental aspects of linguistic and mnestic abilities in normal children. *Journal of Clinical and Experimental Neuropsychology, 11*, 518–528.

Hudson, L. J., & Murdoch, B. E. (1992). Language recovery following surgery and CNS prophylaxis for the treatment of childhood medulloblastoma: A prospective study of three cases. *Aphasiology, 6*, 17–28.

Jordan, F. M., Cannon, A., & Mrdoch, B. E. (1992). Language abilities of mildly closed head injuries (CHI) children 10 years post-injury. *Brain Injury, 6*, 39–44.

Jordan, F. M., Murdoch, B. E., Hudson-Tennent, L. J., & Boon, D. L. (1996). Naming performance in brain-injured children. *Aphasiology, 10*, 755–766.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The boston naming test*. Philadelphia: Lea and Febiger.

Kirk, U. (1992). Confrontation naming in normally developing children: Word-retrieval or word knowledge? *The Clinical Neuropsychologist, 6*, 156–170.

Lecours, A. R., & Lhermitte, F. (1979). *L'aphasie*. Paris: Flammarion.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading: MA: Addison-Wesley.

Lezak, M. D. (1995). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Mariën, P., Mampaey, E., Vervaet, A., Saerens, J., & De Deyn, P. P. (1998). Normative data for the Boston Naming Test in native Dutch-speaking Belgian elderly. *Brain and Language, 65*, 447–467.

Nicholas, M., Obler, L. K., Au, R., & Albert, M. (1996). On the nature of naming errors in aging and dementia: A study of semantic relatedness. *Brain and Language, 54*, 184–195.

Raven, J. C. (1996). *Standard Progressive Matrices: Sets A, B, C, D & E*. Oxford, UK: Oxford Psychologists Press.

Raven, J. C., Court, J. H., & Raven, J. (1977). *Manual for Raven's Progressive Matrices and vocabulary scales, section 3: Standard Progressive Matrices*. London: H.K. Lewis.

Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language, 71*, 267–284.

Ross, J. L., Roeltgen, D., Feuillan, P., Kushner, H., & Cutler, G. B. (2000). Use of estrogen in young girls with Turner syndrome. *Neurology, 54*, 164–170.

Rosselli, M., Ardila, A., Bateman, J. R., & Guzmán, M. (2001). Neuropsychological test scores, academic performance, and developmental disordersin Spanish-speaking children. *Developmental Neuropsychology, 20*, 355–373.

Rubin, H., & Liberman, I. Y. (1983). Exploring the oral and written language errors made by language disabled children. *Annals of Dyslexia, 33*, 111–120.

Semrud-Clikeman, M., Guy, K., Griffin, J. D., & Hynd, G. W. (2000). Rapid naming deficits in children and adolescents with reading disabilities and attention deficit hyperactivity disorder. *Brain and Language, 74*, 70–83.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Thuillard-Colombo, F., & Assal, G. (1992). Adaption française de dénomination de Boston: Versions abrégées. *Revue Européenne de Psychologie Appliquée, 42*, 67–73.

van Dale (1984). *Groot Woordenboek Engels-Nederlands*. Utrecht/Antwerpen: Van Dale: Lexicografie.

van Dale (1985). *Groot woordenboek der nederlandse taal*. Utrecht/Antwerpen: Van Dale: Lexicografie.

Weyandt, L. L., & Willis, W. G. (1994). Executive functions in school-aged children: Potential efficacy of tasks in discriminating clinical groups. *Developmental Neuropsychology, 10*, 27–38.

Wolf, M. (1984). Naming, reading, and the dyslexias: A longitudinal overview. *Annals of Dyslexia, 34*, 87–115.

Woordenlijst van de Nederlandse Taal. (1954). 's-Gravenhage.

Worrall, L. E., Yiu, E. M. L., Hickson, L. M. H., & Barnett, H. M. (1995). Normative data for the Boston Naming Test for Australian elderly. *Aphasiology, 9*, 541–551.

Yeates, K. O. (1994). Comparison of developmental norms for the Boston Naming Test. *The Clinical Neuropsychologist, 8*, 91–98.

Yeates, K. O., & Mortensen, M. E. (1994). Acute and chronic neuropsychological consequences of mercury vapor poisoning in two early adolescents. *Journal of Clinical and Experimental Neuropsychology, 16*, 209–222.