The Influence of Word Frequency on Accuracy and Latency in a Newly Developed Dutch Naming Test in Healthy Adult Participants

Name: Grietje Wijngaarden (Margriet)
Student number: 3949451
Status: Final version
Date: July 3, 2015
University of Utrecht
Master Clinical Health Sciences, Speech-Language Pathology Sciences UMC-Utrecht
Supervisor: Dr. Lizet van Ewijk
Teacher: Dr. Rob Zwitserlood
Internship institution: Hogeschool Utrecht
Journal: Brain and Language (reference style: APA, fewer than 10,000 words)
Number of words: 3771
Transparent report: STROBE
Number of words Dutch abstract: 278
Number of words English abstract: 297

Introduction

Approximately one third of the patients who experience a stroke suffer from aphasia(1)(2)(3)(4)(5). Aphasia affects some, or all areas of communication, including expression and understanding of speech, as well as the ability to read and write(4)(6). Difficulty in finding or producing words, also referred to as *anomia*, is a common problem for the majority of people with aphasia(7)(8)(9). Anomia affects daily conversation; it causes disruption of conversation and communication, thereby reducing the quality of life due to its significant impact(8)(10).

Speech production, which includes word retrieval from the mental lexicon (i.e. a store of word meanings[11]), is a complex process. According to Levelt's model, described by Jescheniak and Levelt, there are three major levels of processing involved in the production of spoken language(12)(13). The first level is the process of *conceptualization*: a pre-verbal message is created and specifies an idea or concept to be verbally expressed(13). The second level is the process of *formulation*: the outlining of the preverbal message into a linguistic form(13). Word meanings, stored in the semantic system, are activated in response to an idea or concept(11). The third level is the process of *articulation*, which involves three aspects: retrieval of a phonetic plan, initiation of articulation, and execution of articulation (i.e. the production of physical speech)(13). The processes of conceptualization and articulation are out of the scope of this study and will not be further discussed. This study will focus on the second level: the process of formulation. Therefore, this level will be further explained.

According to Jescheniak and Levelt, one of the functions of the level of formulation, also referred to as *lexicalization* or *lexical access*, is "the set of processes that govern the retrieval of lexical entries from the mental lexicon"(13)(14). Numerous studies have shown that lexical access follows two distinct steps(13)(14). The first step is the semantic step, which involves selection of a semantically appropriate item (e.g. cat)(15). During the second step, the phonological step, access to spoken word forms is provided by the phonological output lexicon (i.e. a store of spoken word forms[11]), resulting in retrieval of the word's phonological characteristics (e.g. kæt)(11)(14)(15).

Impairment of the phonological step due to stroke may result in impairment of word retrieval in people with aphasia. This particular impairment is often characterized by the presence of a frequency effect. Additionally, there is much evidence that, even in healthy speech, less frequent words tend to be more difficult to retrieve and produce than words that are used more frequently(13)(16)(17)(18). This so-called *word frequency effect* in speech production was discovered in 1965 by Oldfield and Wingfield(13)(14). Results from a picture-naming task showed that low frequent words took longer to name than high frequent words, and it was suggested that this effect was due to naming itself, instead of a discrepancy in speeds of object recognition(13). For people with aphasia the effect of frequency on naming ability is apparent in both speed and ease of retrieval. Retrieval of words with a low frequency of occurrence is more likely to be disrupted than those with a high frequency, and when retrieval is successful, it often takes much longer(11). A large body of research indicates that the frequency effect is strongly associated with the phonological step(13)(15), although there are some studies that have found frequency effects in production tasks involving the semantic step(11)(15). Nevertheless, most studies show that the effect of word frequency is stronger on the phonological step than on the semantic step(13)(14)(15).

As mentioned before, people with anomia have difficulty in finding words(7)(8)(9). In a clinical setting a naming test is used to determine whether, and to what extent, a patient has word finding difficulties. According to Harry and Crowe, the use of visual stimuli in a test requiring verbal responses is useful to determine naming impairment(9). Whitworth et al. suggest that picture-naming assessments provide a useful starting point in the assessment of spoken word production(11). They state that a test controlling for a variety of contrasts (e.g. frequency) "can provide some information about the possible level of any impairment"(11). For example: an impairment in the phonological output lexicon results in impaired word retrieval in spoken production, characterized by, inter alia, failures and delays in word retrieval, phonological errors and the presence of a frequency effect(11). A test of naming ability should be sensitive and valid in order to monitor naming deficits across time to assess the degree of recovery following stroke(9).

The Boston Naming Test (BNT)(19) is the most commonly used test to detect naming difficulties resulting from stroke(9)(16)(17)(20). The BNT consists of 60 line drawings, ordered according to word frequency from high frequent words to low frequent words(20). However, Harry and Crowe raise significant criticisms on the BNT and suggest it might not be the best choice in the assessment of picture naming(9). They argue that the BNT "has poor psychometric properties, is not adequately standardized, and has inadequate norms"(9). Moreover, the BNT does not live up to occurred advances in the neuropsychology of naming since 1983(9). Additionally, Brookshire and Nicholas studied the relationship between the order of items in the BNT and word frequency in daily life(17). They state that it is unlikely that most words of the BNT items are encountered in daily life conversations. Therefore they suggest that the BNT may provide information about the size of a person's vocabulary or the ability to retrieve words that occur infrequently in a naming test, but the test provides little information about word retrieval in daily life conversations(17). Furthermore, the BNT items are arranged by word frequency in American English. However, the arrangement of word frequency is different across languages(20). Besides, not all BNT items are appropriate for other cultures(21). These arguments clearly indicate that a new test of naming, arranged by word frequency consistent with the Dutch language, is needed.

This study is the first step in a process towards the development of a new naming test for patients with word finding difficulties due to aphasia. The aim of this study is to develop a picture-naming test arranged by word frequency according to the Dutch language. The study involved healthy Dutch adults, who were examined with the picture-naming test to determine if a frequency effect could be found in this particular test, based on latency (i.e. the period between the moment a stimulus is presented and the moment a response occurs). Furthermore, normative data will be collected from the participants' score on the test (i.e. accuracy). The study is focused on the following research question:

To what extent does word frequency influence accuracy and latency in a newly developed Dutch naming test in healthy, adult participants?

Method

Design

This quantitative study has an experimental design: a picture-naming test, arranged by word frequency according to the Dutch language, is developed. Word frequency is the manipulated variable in this experiment. The study lasted approximately six months, including two months of collecting data.

Pilot study

Prior to the actual study, a pilot study was conducted. A total of 20 participants were asked to name 120 images in order to select the definite set of 60 images. Participants' responses were labeled correct or incorrect. Additionally, participants were asked to give feedback on the pictures about familiarity, name agreement, and image agreement. Based on these results, 60 images were selected. The results of the pilot study can be obtained from the researcher.

Participants

A group of healthy adults was examined with the newly developed picture-naming test. A sample of subjects was drawn from the adult Dutch population. Participants were recruited by means of advertising in the circle of friends and acquaintances of the researcher. In order to be eligible to participate in this study, several criteria should have been met: a participant had to be an adult between 18-30 or 50-75 years old, Dutch as native language, and no history of neurological damage. Participants were excluded from the study when they had experienced neurological damage, a language disorder during childhood, a speech disorder (e.g. stuttering), a hearing problem not corrected-to-normal, or a vision problem not corrected-to-normal. A total of 59 healthy Dutch adults took part in the study, divided into two age groups: young adults (18-30 years old) and older adults (50-75 years old). Tabor Connor et al. state that age affects BNT score(22). Their longitudinal study demonstrated that lexical retrieval declines with age: performance on the BNT declined two percentage points per decade(22). By selecting two contrasting age groups, a potential effect of age on latency and accuracy in this particular picture-naming test was examined. Ashaie and Obler mention education as an additional factor influencing naming ability(21). Moreover, Tabor Connor et al. found that "the interaction of gender and education was a predictor of performance" in naming(22). Therefore this study controlled for education and gender in both groups as best as possible.

The setting of this study varied: participants were either tested in their home or work environment, or at the researcher's work space. The measurements took place under similar circumstances, regardless of the different settings.

Two participants were excluded, because they did not meet the inclusion criteria. One did not have Dutch as their mother tongue and the other recently suffered from a TIA. Another four participants were excluded due to technical problems with the equipment. Altogether, 53 participants completed the test. The group of young adults consisted of 23 participants, with a mean age of 26.87 (age range 18-30; SD 3.45). Average age for the 30 older adults was 60.03 years (age range 51-70; SD 5.29). Table 1 shows the participants' characteristics.

[INSERT TABLE 1 HERE]

Materials

The test consists of black and white line drawings from different categories (e.g. animals, body parts, tools, musical instruments, et cetera). These items were drawn from the Snodgrass and Vanderwart database(23). This set of images has normative data on name agreement, familiarity, and visual complexity(11)(23). Furthermore, the database is standardized in different cultures, languages, and among different populations(24). It is available for free download(23)(24). The test included 60 items: 20 high frequent words, 20 mid frequent words, and 20 low frequent words (appendix 1). The frequency of the words were extracted from the CELEX database(25). In 1993, frequencies were published by Baayen et al., based on written texts of 42 million words(25). Keuleers and Brysbaert argue that, since its publication, CELEX is the main source when it comes to lexical information for the Dutch language, and in the case of this study, word frequencies in particular(26).

Procedure

This study was conducted according to the principles of the Declaration of Helsinki(27). As confirmed by the Medical Ethical Screening Committee of the Faculty of Health of the University of Applied Sciences Utrecht (HU), no approval was needed by the Medical Research Involving Human Subjects Act of the Dutch Law Medical-scientific Research

(WMO). Written informed consent was obtained from all subjects prior to their participation. Each participant was tested once and individually. The session lasted approximately 10 to 15 minutes. All visual stimuli were displayed on a Windows 7 Intel CORE HP EliteBook laptop(28). The participants responded into a ZOOM H6 Handy Recorder with a ZOOM XYH-6 XY stereo microphone measuring speech-onset latencies(29). The microphone was placed on a König table tripod(30) in front of the participant. The whole picture appeared instantly on the screen. The timer started simultaneously with the picture onset. An external keyboard was used by the researcher to proceed to the next item. Image order was randomized for each participant eliminating possible effects, such as decreasing latency caused by appearance of two consecutive images from the same category (i.e. *zebra* following *giraffe*). Verbal responses of the participants were written down and recorded. Latency was measured manually, using the software application Audacity, an audio editor and recorder. In Audacity, the picture onset was displayed as a vertical bar. Latency time was measured from the beginning of the vertical bar (picture onset) to the start of the initial sound of the response.

All responses were given a code: 1 for correct (i.e. *bed* for *bed*); 2 for incorrect (i.e. *crab* for *lobster*); 3 for a filler (i.e. *a nose*; where *a* was considered a filler); and 4 for self correction (i.e. *goose, excuse me, swan*). At first, all responses were measured in seconds with three decimal places; subsequently the responses were converted to milliseconds, which was necessary for importing the data in SPSS from Excel. Responses with code 2 (incorrect response) were not included in the analysis of latency, because the target word (or a synonym) was not retrieved. In the case of a filler (code 3) or a self correction (code 4), the response was examined further: an item was excluded when it was too hard to distinguish where the filler, respectively incorrect answer, stopped and the initial sound, respectively self correction, started.

Data analysis

The primary study parameters were latency and accuracy. Latency is a continuous variable and was calculated in milliseconds. Accuracy is a categorical variable and was calculated with a predetermined points scale (code 1 to 4). Word frequency was the independent variable and is categorical, divided into three subcategories: high frequent words, mid frequent words, and low frequent words. Latency was determined for each participant. Latency was submitted to ANOVAs, with significance assigned at the 5% level, unless otherwise specified. Given the small number of participants, normality could not be assumed. Therefore the non-parametric Kruskal-Wallis test was performed to examine if the means of the three groups were equal. The Mann-Whitney test was performed three times to compare between groups (high-mid, high-low, and mid-low). In the case of missing data (e.g.

code 2, 3, or 4), the item was excluded from the analysis. To examine the influence of word frequency on accuracy, the Mann-Whitney test was used.

The secondary study parameter was age. The subjects were divided into two age groups. Latency of both age groups on the three levels of word frequency were compared using non-parametric ANOVAs, with significance assigned at the 5% level, unless otherwise specified. Spearman's rho was used to examine whether there was a correlation between latency and the frequency of the words as extracted from CELEX(25). Statistical analyses were performed using SPSS 20 for Windows(31).

Results

All 53 participants completed the naming test, adding up to 3180 responses; 1060 per frequency level. Two extreme values were found; a latency of 15833 ms on a mid frequent word (*anker*, anchor) and a latency of 15858 ms on a low frequent word (*spinnewiel*, spinning wheel). These two outliers were excluded from the analysis for statistical reasons. After exclusion of words with code 2, 3, or 4, and the two outliers, a total of 267 responses were listed as missing values: 49 of the high frequent words, 84 of the mid frequent words, and 134 of the low frequent words. The distribution of the latencies was highly skewed. This skewness was reduced by a logarithmic transformation for further analysis (lnReaction time).

To investigate the effect of frequency on latencies, Spearman's rho was used to correlate frequency with the lnReaction time. A significant negative correlation between lnReaction time and frequency (F(1.2913)=105,233, p<.001) was found, indicating that reaction time decreased as frequency of occurrence increased (graph 1).

[INSERT GRAPH 1 HERE]

The effect of frequency was further investigated by dividing the words into high, mid and low frequency words. Kruskal-Wallis revealed a significant difference in mean latency between the three levels of frequency, H(2)= 146,047, p<.001. Subsequently, the differences between the three frequency levels were examined. The mean reaction time of the high frequent words was 811.74 ms, 960.23 ms for the mid frequent words, and 1058.61 ms for the low frequent words. The Mann-Whitney test was carried out three times to compare high and mid frequency, high and low frequency, and mid and low frequency. The comparison of high and mid frequency revealed a significant difference in lnReaction time (U= 405826,500, z=-6.913, p<.001), with the high frequent words having an average rank of 907.62, while the mid frequent words had an average rank of 1085,69, indicating that the high frequent words were named faster than the mid frequent words. The comparison of high and low frequency revealed a significant difference in lnReaction time (U= 319908,500, z=-12.108, p<.001), with the high frequent words, with an average rank of 822.80, being named faster than the low frequent words, with an average rank of 1131.03. A significant difference in lnReaction time was also found in the comparison of mid and low frequency (*U*= 529183,500, *z*=-2.316, *p*=.021), with the mid frequent words having an average rank of 1029.73, while the low frequent words had an average rank of 1091.27, which indicates that mid frequent words were named faster than the low frequent words. Graph 2 shows three box plots of the lnReaction time per frequency.

[INSERT GRAPH 2 HERE]

Next, the hypothesis that latency increases with age was examined. Table 2 shows that the mean latency of age group 1 (young participants) is lower than age group 2 (older participants) on all three frequency levels. Three Mann-Whitney tests were used: age group 1 compared with age group 2 on high frequency; age group 1 compared with age group 2 on mid frequency; and age group 1 compared with age group 2 on low frequency. The Mann-Whitney tests revealed a significant difference between age group 1 and 2 in lnReaction time on high frequent words (*U*= 94909,500, *z*= -6.886, *p*<.001), mid frequent words (*U*= 93681,000, *z*= -5.480, *p*<.001), and low frequent words (*U*= 91301,500, *z*= -3.650, *p*<.001), with the group of young adults being faster than the group older adults on all three frequency levels (table 3).

[INSERT TABLE 2 HERE]

[INSERT TABLE 3 HERE]

In total, there were 2762 correct responses, 113 incorrect responses, 230 responses with a filler, and 75 self corrections. For the purpose of statistical analyses, all responses with a valid latency were labeled as code 1 and the remainder as code 2 (this included all incorrect responses and responses with code 3 or 4 which could not be measured), adding up to 2913 correct responses (since some of the answers that included fillers or self corrections did have a valid latency) and 267 incorrect responses. Three Mann-Whitney tests were performed to examine the influence of frequency on accuracy. A significant difference was revealed between the accuracy on high and mid frequent words (*U*= 543250,000, *z*= -3.134, *p*=.002), on high and low frequent words (*U*=516750,000, *z*= -6.572, *p*<.001), and on mid and low frequent words (*U*=535300,000, *z*= -3.574, *p*<.001)(table 4). This implies that there were more incorrect responses in the mid frequent words compared with the high frequent words and more incorrect responses in the low frequent words in comparison with the high and mid frequent words (table 5, graph 3).

[INSERT TABLE 4 HERE]

[INSERT TABLE 5 HERE]

[INSERT GRAPH 3 HERE]

Discussion

In this study, the influence of word frequency on latency in a newly developed Dutch picture-naming test was examined. The frequencies extracted from the CELEX database(25) were correlated with the lnReaction time on a naming task and showed a significant negative correlation, which indicates that reaction time decreased as frequency of occurrence increased. This finding resonates with previous studies that show that a frequency effect is present in healthy speech(13)(16)(17)(18). The differences between the lnReaction time in high, mid, and low frequent words were statistically significant; lnReaction time increased when word frequency decreased. The presence of a frequency effect in these three subsets of items, indicates that this test has the potential to become a clinically relevant naming test to assess word finding difficulties in patients. The influence of word frequency on accuracy was also tested in this study. In line with the findings on frequency, the number of incorrect responses increased as word frequency decreased.

It has been previously reported that naming performance is affected by age(20)(22). In this study, age appeared to have a strong effect on latency. On all three frequency levels, the young adults were faster than the older adults. These results show the importance of including participants of varying age ranges during the standardization process of diagnostic tests.

The picture-naming test of this study used visual stimuli in the assessment of speech production. Several authors have suggested that this is a useful starting point to detect naming impairment(9)(11). Up to now, the BNT is the most commonly used test to determine naming difficulties following stroke(9)(16)(17)(20). However, as discussed in the introduction, the BNT might no longer be the best choice. The naming task described in this study has been tailored to the Dutch situation and is based on the most recent psycholinguistic models of word retrieval(12)(13). Word frequency was ordered according to the Dutch language, whereas the BNT is arranged by word frequency in American English. This study is a first step towards a solid standardization with adequate norms by examining healthy adults, using standardized images (11)(23)(24).

A strength of this study is that the items of the test were drawn from the Snodgrass and Vanderwart database(23). This database is standardized in different cultures, languages, and among different populations(24). Moreover, the images have normative data on name agreement, familiarity, and visual complexity(11)(23). Furthermore, all 3180 responses were measured manually, enhancing the reliability of the measurements. Another strength is the comparison of the two age groups. However, a limitation is the small number of these groups; 23 and 30 participants per age group. Moreover, the groups were heterogeneous; the variables gender and education level were not equally distributed among the groups. Additionally, when subgroups would have been created, they would have been too small to draw conclusions from the analysis. Another limitation is that latency was measured by merely one researcher; inter-rater reliability was thus not assessed.

Conclusion and recommendations

This study is a first step towards a new picture-naming test, based on word frequencies of the Dutch language for people with anomia following stroke. In this test, a frequency effect was present among healthy Dutch adults. Latency increased significantly when word frequency decreased, with age having a strong effect on reaction time. Young adults named the images significantly faster than older adults on all three frequency levels. Additionally, when frequency decreased, the number of incorrect responses increased.

This new picture-naming test controls for frequency and will therefore be able to provide some information on where the impairment in the speech production of the person with aphasia is located. Eventually, this will help speech-language pathologists in their decision making on what treatment is best suited for their patients. For future research it is suggested that latency would be measured automatically, since measuring manually is very time consuming. Moreover, assessment of inter-rater reliability is preferred. Next, people with aphasia should be examined with this test and their results (latency and accuracy) should be compared with the results of the healthy participants of this study to complete the standardization of this test.

References

1. Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. J Intern Med. 2001;249:413–22.

2. Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: Type, severity and prognosis: The Copenhagen aphasia study. Cerebrovasc Dis. 2004;17:35–43.

3. Lazar RM, Speizer AE, Festa JR, Krakauer JW, Marshall RS. Variability in language recovery after first-time stroke. J Neurol Neurosurg Psychiatry. 2008;79:530–4.

4. Brady M, Kelly H, Godwin J, Enderby P. Speech and language therapy for aphasia following stroke. Cochrane Database Syst Rev. 2012;(5).

5. Geranmayeh F, Brownsett SLE, Wise RJS. Task-induced brain activity in aphasic stroke patients: what is driving recovery? Brain [Internet]. 2014;2632–48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24974382

6. Shah PP, Szaflarski JP, Allendorfer J, Hamilton RH. Induction of neuroplasticity and recovery in post-stroke aphasia by non-invasive brain stimulation. Front Hum Neurosci [Internet]. 2013;7(December):888. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3870921&tool=pmcentrez&rendertype=abstract

7. Nickels L. Therapy for naming disorders: Revisiting, revising, and reviewing. Aphasiology. 2002;16(October 2014):935–79.

8. Best W, Greenwood A, Grassly J, Herbert R, Hickin J, Howard D. Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. Cortex [Internet]. Elsevier Ltd; 2013;49(9):2345–57. Available from: http://dx.doi.org/10.1016/j.cortex.2013.01.005

9. Harry A, Crowe SF. Is the Boston naming test still fit for purpose? Clin Neuropsychol [Internet]. 2014;28(October):486–504. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24606169

10. Herbert R, Gregory E, Best W. Syntactic versus lexical therapy for anomia in acquired aphasia: differential effects on narrative and conversation. Int J Lang Commun Disord [Internet]. 2014;49:162–73. Available from: http://doi.wiley.com/10.1111/1460-6984.12054

11. Whitworth A, Webster J, Howard D. A cognitive neuropsychological approach to assessment and intervention in aphasia. A clinician's guide. East Sussex: Psychology Press; 2005.

12. Levelt WJM. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.

13. Jescheniak JD, Levelt WJM. Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. J Exp Psychol Learn Mem Cogn. 1994;20(4):824–43.

14. Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of frequency in word production? Insights from aphasic picture naming errors. Cogn Neuropsychol. 2008;25(4):463–92.

15. Nozari N, Kittredge AK, Dell GS, Schwartz MF. Naming and repetition in aphasia: steps, routes, and frequency effects. J Mem Lang. 2010;63(4):541–59.

16. Yochim BP, Rashid K, Raymond NC, Beaudreau S a. How frequently are words used on naming tests used in spoken conversation? Clin Neuropsychol [Internet]. 2013;27(March 2015):973–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23656300

17. Brookshire RH, Nicholas LE. Relationship of word frequency in printed materials and judgments of word frequency in daily life to Boston Naming Test performance of aphasic adults. Clin Aphasiology. 1995;23:107–19.

18. Hoffman P, Rogers TT, Ralph MAL. Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: insights using a novel method to quantify contextual variability in meaning. J Cogn Neurosci. 2011;23:2432–46.

19. Kaplan E, Goodglass H, Weintraub S. The Boston Naming Test. Philadelphia, PA: Lea & Fibiger; 1983.

20. Kavé G. Standardization and norms for a Hebrew naming test. Brain Lang. 2005;92:204–11.

21. Ashaie S, Obler L. Effect of age, education, and bilingualism on confrontation naming in older illiterate and low-educated populations. Behav Neurol. 2014;1–10.

22. Connor LT, Spiro A, Obler LK, Albert ML. Change in object naming ability during adulthood. J Gerontol B Psychol Sci Soc Sci. 2004;59(5):P203–9.

23. Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychol Hum Learn. 1980;6(2):174–215.

24. Bonin P, Guillemard-Tsaparina D, Méot A. Determinants of naming latencies, object comprehension times, and new norms for the Russian standardized set of the colorized version of the Snodgrass and Vanderwart pictures. Behav Res Methods [Internet]. 2013;45(3):731–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23239068

25. Baayen RH, Piepenbrock R, van Rijn H. The CELEX Lexical Database [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania; 1993.

26. Keuleers E, Brysbaert M, New B. SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. Behav Res Methods. 2010;42(3):643–50.

27. WMA. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects [Internet]. 2014. Available from: http://www.wma.net/en/30publications/10policies/b3/

28. Hewlett-Packard Development Company, L.P. Palo Alto, CA; 2015.

29. ZOOM Corporation. Tokyo, Japan; 2015.

30. König. ’s-Hertogenbosch; 2015.

31. IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.; 2011.

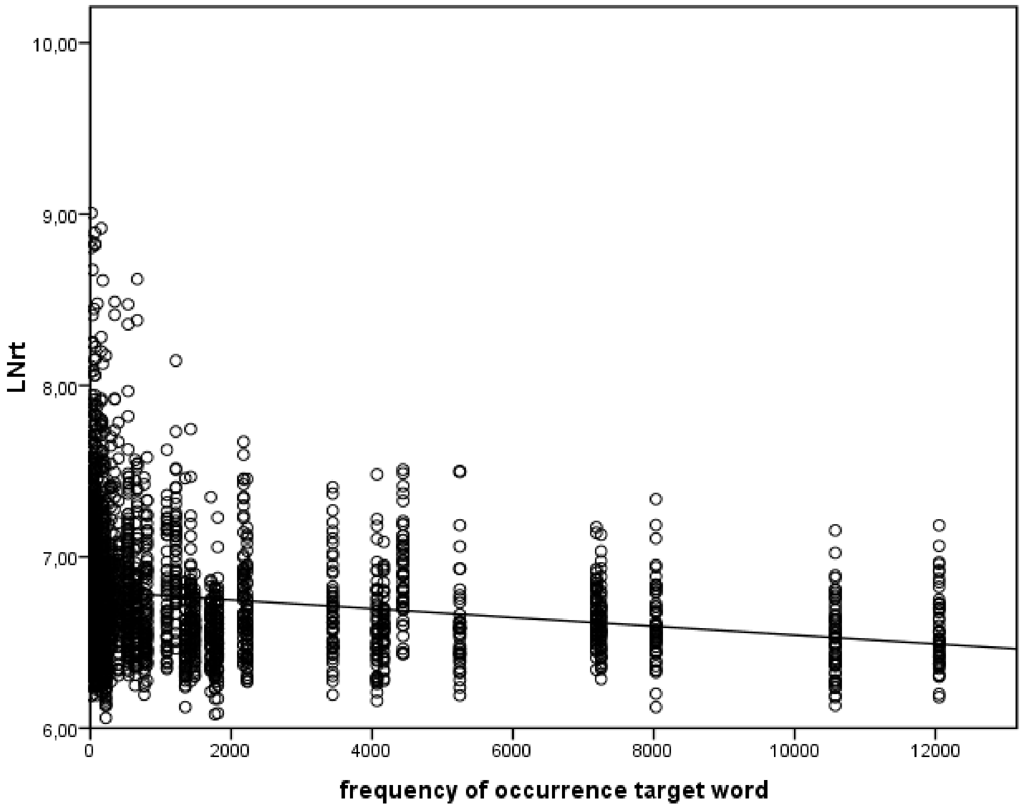
Tables and graphs

Table 1 Participants' characteristics

	young adults n = 23		**older adults n = 30**	
	mean	SD	mean	SD
age	26.87	3.45	60.03	5.29
gender	12 males / 11 females		13 males / 17 females	
county n				
1	3		11	
2	14		1	
3	2		18	
4	4		0	
education n				
1	0		0	
2	5		14	
3	18		16	

County: 1= Friesland; 2= Noord Holland; 3= Zuid Holland; 4= Utrecht
Education: 1= low (only elementary school or LBO); 2= mid (VMBO/MAVO/MULO/HAVO/MBO); 3= high (VWO/HBO/WO)

Graph 1 Correlation between lnReaction time and frequency of occurrence



Graph 2 Box plot of lnReaction time per frequency (1= high; 2= mid; 3= low)

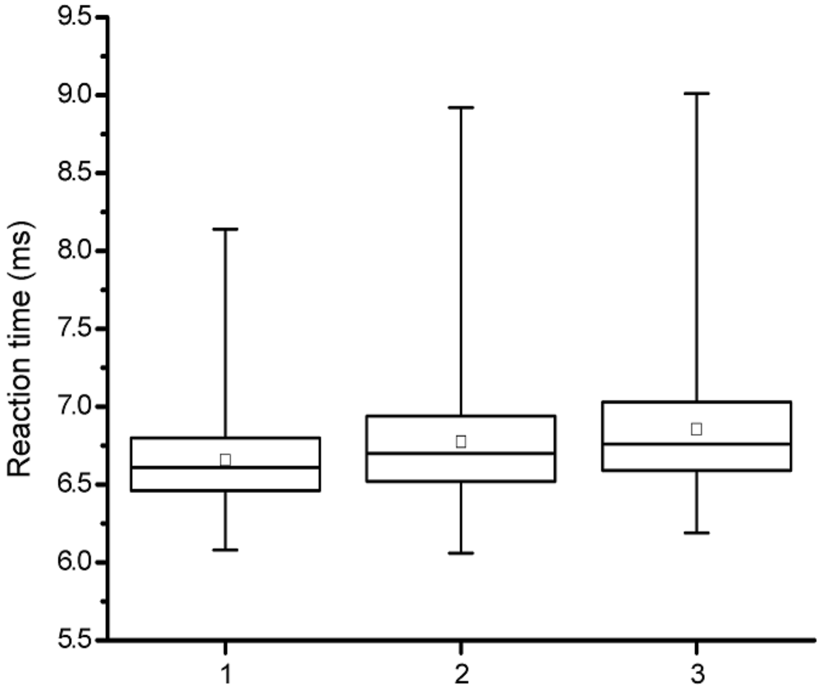


Table 2 Mean latency per age group on frequency

high frequent	
age group	*mean latency*
1	764.54
2	849.77
mid frequent	
age group	*mean latency*
1	872.88
2	1030.46
low frequent	
age group	*mean latency*
1	1005.28
2	1102.12

Table 3 Mann-Whitney U test frequency and age group

	Mann-Whitney U	**Z**	**Asymp. Sig. (2-tailed)**
high frequent			
age group 1 en 2	94909,500	-6.886	.000
mid frequent			
age group 1 en 2	93681,000	-5.480	.000
low frequent			
age group 1 en 2	91301,500	-3.650	.000

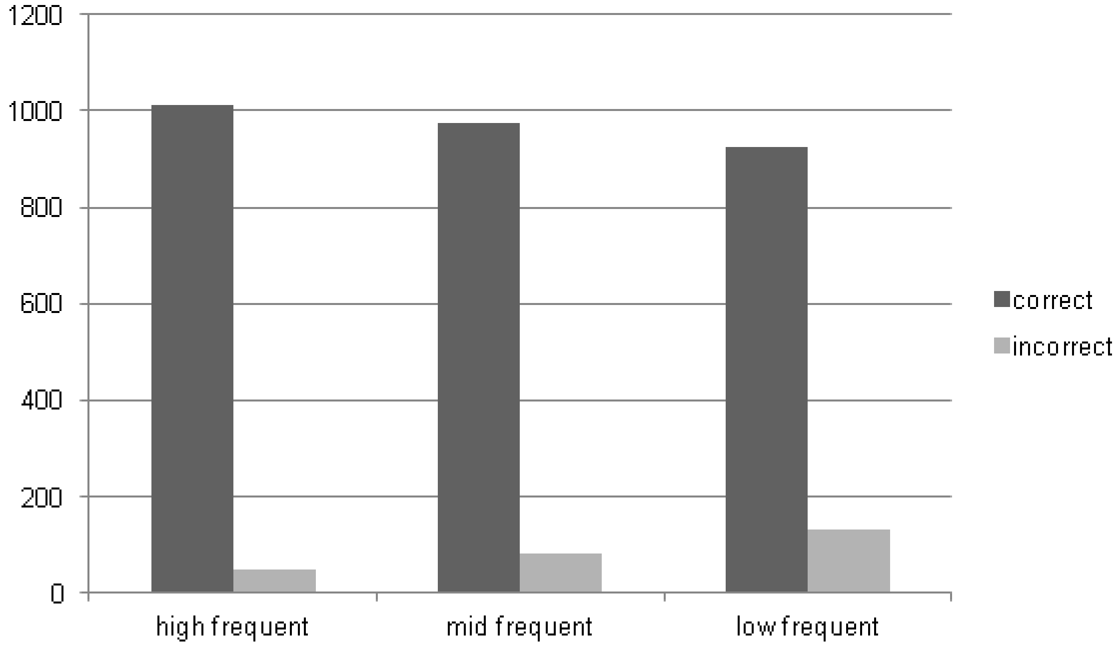
Table 4 Mann-Whitney U test accuracy and frequency (high, mid, low)

	Mann-Whitney U	Z	Asymp. Sig. (2-tailed)
accuracy high and mid frequent	543250,000	-3.134	.002
accuracy high and low frequent	516750,000	-6.572	.000
accuracy mid and low frequent	535300,000	-3.574	.000

Table 5 Accuracy per frequency level

	high frequent	mid frequent	low frequent	total
correct response	1011	976	926	2913
incorrect response	49	84	134	267
total	1060	1060	1060	3180

Graph 3 Accuracy per frequency level



Dutch Summary

Titel: De invloed van woord frequentie op de goedscore en bedenktijd in een nieuw ontwikkelde Nederlandse benoemtest bij volwassen, gezonde deelnemers.

Inleiding: Spraak productie, inclusief het ophalen van woorden, is een complex proces. Laag frequente woorden zijn moeilijker op te halen en te produceren dan hoog frequente woorden; het zogenaamde *woord frequentie effect*. Moeilijkheden met het vinden of produceren van woorden is een bekend probleem voor veel mensen met afasie na een beroerte. De Boston Naming Test (BNT) is de meest gebruikte test om woordvindingsproblemen vast te stellen. Vanwege matige psychometrische eigenschappen, een gebrek aan adequate standaardisatie en normering van de BNT, is een nieuwe benoemtest, gerangschikt volgens woord frequentie van de Nederlandse taal, nodig.

Doel en onderzoeksvraag: Het ontwikkelen van een benoemtest gerangschikt volgens Nederlandse woord frequentie, gericht op de volgende onderzoeksvraag: in hoeverre heeft woord frequentie invloed op de goedscore en reactietijd in een nieuw ontwikkelde Nederlandse benoemtest bij gezonde, volwassen deelnemers?

Methode: Negenenvijftig proefpersonen hebben deelgenomen aan het onderzoek, verdeeld in twee leeftijdsgroepen (18-30 en 50-75 jaar). De benoemtest bestaat uit 60 zwart wit lijntekeningen: 20 hoog frequente, 20 midden frequente en 20 laag frequente woorden. De goedscore en reactietijd werden gemeten.

Resultaten: Een frequentie effect werd gevonden onder gezonde, volwassen Nederlandse deelnemers: reactietijd nam significant toe wanneer woord frequentie afnam. De jongvolwassenen waren significant sneller dan de oudere volwassenen op de drie frequentie niveaus. Het aantal foute antwoorden nam toe naarmate de frequentie afnam.

Conclusie: Goedscore en reactietijd werden beïnvloed door woord frequentie in deze nieuw ontwikkelde Nederlandse benoemtest bij gezonde, volwassen deelnemers.

Aanbevelingen: Het verder standaardiseren en normeren van deze benoemtest door afasiepatiënten te testen.

Trefwoorden: afasie, woordvindingsproblemen, benoem test, woord frequentie, reactietijd

English abstract

Title: The influence of word frequency on accuracy and latency in a newly developed Dutch naming test in healthy adult participants.

Background: Speech production, including word retrieval, is a complex process. Low frequent words tend to be more difficult to retrieve and produce than high frequent words; the so-called *word frequency effect*. Difficulty in finding or producing words is a common problem for many people with aphasia following stroke. The Boston Naming Test is the most commonly used test to detect naming difficulties resulting from stroke. Due to poor psychometric properties, a lack of adequate standardization and norms, a new naming test, arranged by word frequency consistent with the Dutch language, is needed.

Aim and research question: To develop a picture-naming test arranged by word frequency according to the Dutch language, aiming at the following research question: To what extent does word frequency influence accuracy and latency in a newly developed Dutch naming test in healthy, adult participants?

Method: A total of 59 healthy Dutch adults took part in the study, divided into two age groups (18-30 and 50-75 years old). The test consists of 60 black and white line drawings: 20 high, 20 mid, and 20 low frequent words. Accuracy and latency were measured.

Results: A frequency effect was present among healthy Dutch adults; latency increased significantly when word frequency decreased. Young adults were significantly faster than older adults on all three frequency levels. The number of incorrect responses increased significantly as word frequency decreased.

Conclusion: In this newly developed Dutch naming test accuracy and latency were influenced by word frequency in healthy, adult participants.

Recommendations: It is suggested that people with aphasia should be examined with this test to complete the standardization of this test.

Key words: aphasia, anomia, naming test, word frequency, latency

Appendices

Appendix A Items of the picture-naming test

nr.	woord	nr.	
1	*bed* (bed)	31	*kruk* (stool)
2	*boek* (book)	32	*schaar* (scissors)
3	*tafel* (table)	33	*gitaar* (guitar)
4	*oog* (eye)	34	*spijker* (nail)
5	*kerk* (church)	35	*peer* (pear)
6	*glas* (glass)	36	*kanon* (cannon)
7	*arm* (arm)	37	*zwaan* (swan)
8	*neus* (nose)	38	*kreeft* (lobster)
9	*voet* (foot)	39	*zaag* (saw)
10	*bank* (couch)	40	*strik* (bow)
11	*boom* (tree)	41	*trompet* (trumpet)
12	*blad* (leaf)	42	*tomaat* (tomato)
13	*zon* (sun)	43	*banaan* (banana)
14	*fiets* (bike)	44	*schommelstoel* (rocking chair)
15	*oor* (ear)	45	*zeilboot* (sail boat)
16	*sleutel* (key)	46	*schroevendraaier* (screwdriver)
17	*jurk* (dress)	47	*waterput* (well)
18	*bril* (glasses)	48	*pompoen* (pumpkin)
19	*kom* (bowl)	49	*harp* (harp)
20	*pistool* (gun)	50	*liniaal* (ruler)
21	*bel* (bell)	51	*gloeilamp* (lightbulb)
22	*vlag* (flag)	52	*kangoeroe* (kangaroo)
23	*pen* (pen)	53	*zebra* (zebra)
24	*bloem* (flower)	54	*giraffe* (giraffe)
25	*wortel* (carrot)	55	*spinnewiel* (spinning wheel)
26	*knoop* (button)	56	*strijkplank* (ironing board)
27	*vrachtwagen* (truck)	57	*deegroller* (rolling pin)
28	*citroen* (lemon)	58	*sneeuwpop* (snowman)
29	*anker* (anchor)	59	*stinkdier* (skunk)
30	*wiel* (wheel)	60	*knijper* (clothespin)