A Dutch, Computerized, and Group Administrable Adaptation of the Operation Span Test

Wim De Neys*, Géry d'Ydewalle, Walter Schaeken, & Gerrit Vos

K.U. Leuven, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium

* Corresponding author. Fax: +32 16 326099; Tel: +32 16 326143 ;
E-mail: Wim.Deneys@psy.kuleuven.ac.be

Running head: Operation Span

De Neys, W., d'Ydewalle, G., Schaeken, W., & Vos, G. (2002). A Dutch, computerized, and group administrable adaptation of the operation span test. *Psychologica Belgica, 42,* 177-190.

One of the most popular tests to measure Working Memory (WM) capacity is the operation span task (OSPAN) by Turner and Engle (1989). We present a Dutch, computerized, and group administrable adaptation (GOSPAN) of this test. The GOSPAN requires no active intervention of the experimenter and allows testing large groups at the same time. Participants received sets of operation-word strings (e.g., 'IS 4/2 – 1 = 5 ? BALL') on the computer screen. Participants first read the operation silently and pressed a key to indicate whether the answer was correct or not. The number of correct responses and mean response latencies were recorded. After the participant typed down the response, the corresponding word (e.g., 'BALL') from the operation-word string was presented shortly (800 ms). We tested 424 first year psychology students with the GOSPAN. Forty-six participants were individually retested with the standard OSPAN task. The alpha coefficient for the GOSPAN was .74 and the correlation with the standard OSPAN reached .50 (.70 when corrected for attenuation). The study provides researchers with a time saving, reliable, and valid adaptation of the OSPAN task.

A Dutch, Computerized, and Group Administrable Adaptation of the Operation Span Test

Individual differences in working memory (WM) capacity have received a great deal of attention from cognitive scientists. WM capacity has been related to performance on a wide range of higher-order cognitive tasks, including reading and language comprehension, vocabulary learning, note taking, writing, reasoning, bridge playing, and even dealing with life event stress (e.g., Engle, 2001; Klein & Boals, 2001; Kyllonen & Christal, 1990)

A number of tests have been proposed to assess peoples WM capacities (e.g., Reading Span, Daneman & Carpenter, 1980; Counting Span, Case, Kurland, & Goldberg, 1982; Sentence Span, Waters & Caplan, 1996). Over the last few years, the operation span task (OSPAN, see La Pointe & Engle, 1990; Turner & Engle, 1989) has become one of the most popular WM tasks (e.g., Barrouillet & Fayol, 1998; Hitch, Towse, & Hutton, 2001; Klein & Fiss, 1999; Miyake et al., 2000).

In the OSPAN participants read aloud a series of operation-word strings (e.g., 'IS (4 : 2) – 1 = 5 ? BALL'). They first read the operation, respond as to whether or not the equation is correct and then read the word aloud. After a set of two to six operation-word strings, participants have to recall the list of presented words. The WM capacity score is the total number of correctly recalled words.

The OSPAN has excellent validity and reliability characteristics (Engle, Tuholski, Laughlin, & Conway, 1999; Klein & Fiss, 1999). Test scores correlate well with standard measures of higher cognitive functioning (e.g., the Scholastic Aptitude Test and Raven Progressive Matrices scores, see Engle et al., 1999). Klein and Fiss also reported a high internal and test-retest reliability.

In the present study we introduce a Dutch and computerized, group administrable adaptation of the OSPAN. Since the only language specific component in the OSPAN are the words presented with the operations, constructing a Dutch version was rather straightforward. We simply replaced the set of high frequency, English words in the operation-word strings by a set of high frequency Dutch words.

More important are the adaptations that allowed a group assessment. In the standard OSPAN task participants are tested individually. After a participant has read and answered an operation-word string, the experimenter provides the next string. A major advantage of the procedure is that it allows the participant to read and calculate the operation at his or her own pace. A fixed presentation time for all participants (as in Singer, Andrusiak, Reisdorf, & Black, 1992) runs the risk of not allowing some participants to finish reading, while others will have additional time to rehearse the words. Such rehearsal is detrimental for the reliability of a WM measure (Turner & Engle, 1989; Waters & Caplan, 1996).

A major disadvantage of the guided, individual assessment, however, is that it is very time and attention demanding for the experimenter. This individual assessment disadvantage is especially clear in the working memory experiments of Engle and colleagues (e.g., Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2000). These studies typically involve specific participants from the top and bottom quartile of the OSPAN distribution in a large (over 400 participants), pretested sample. Since the OSPAN takes about 20 minutes this means that the individual assessment demands about a month of OSPAN testing from a single researcher.

A reliable OSPAN adaptation that allows testing multiple participants at the same time would therefore mean a major, time-saving improvement. In the present study we introduce such a computerized, group administrable adaptation (GOSPAN). It requires no active intervention of the experimenter and allows testing large groups at the same time.

The crucial adaptation is that we first presented the operation on screen (e.g., 'IS (4 : 2) – 1 = 5 ?'). Participants had to read the operation silently and pressed a key to indicate whether the answer was correct or not. The number of correct responses and mean response latencies were recorded. Deviating reaction times allowed to identify participants that were actively rehearsing the word sets during the operation solving. After the participant had typed down the response, the corresponding word (e.g., 'BALL') from the operation-word string was shortly presented for 800 ms.

We tested the first year psychology students from the University of Leuven with the GOSPAN. In order to check the validity of the GOSPAN we also retested 46 participants individually with the standard OSPAN task.

Method

Participants

The GOSPAN test was presented to 424 first-year psychology students from the University of Leuven (Belgium) for partial fulfilment of a course requirement. Forty-six of these students were also invited for a session with the standard OSPAN task. The 46 students received 5 euro for their participation in the standard OSPAN session.

Material

We selected 132 high frequency (see Uit den Bogaert, 1975; Vingerhoets, 1993), one and two syllable, Dutch words. The one syllable words were used for the OSPAN task and the two syllable words for the GOSPAN task[1]. Both tasks used the same set of 66 operations (taken from Engle et al., 1999).

Each operation was paired with a one (OSPAN) or two (GOSPAN) syllable word. We randomly constructed sets of two to six operation-word strings. Three series of each set size (2-6) were performed in the actual tests. Thus, a total of 15 (3 x 5) series were presented. Three additional series, each consisting of two operation-word strings, were provided as practice for the participants. The order in which the sets appeared was random except for the first and second presented sets which were of size three and two. A participant could thus not know the number of words to be recalled until prompted.

Set size varied in the same random order for every participant. A different set order was chosen for the OSPAN and GOSPAN.

Both in the OSPAN and GOSPAN participants were presented 60 operation-word strings. A subjects span score was the sum of correctly recalled words for sets that were perfectly recalled in the correct order. Thus, span score could range from zero to 60. For example, if a participant was presented the following set

IS (9 : 3) + 2 = 5 ? JOB

IS (5 x 1) – 4 = 2 ? BALL

IS (3 x 4) – 5 = 8 ? MAN

she/he was given credit for three words if she/he recalled JOB, BALL, MAN. The participant would not receive credit if the recall was incomplete (e.g., ‘JOB, BALL’) or in the incorrect serial order (e.g., ‘BALL, JOB, MAN’).

OSPAN. Participants saw individual operation-word strings (e.g., IS (9 : 3) + 2 = 5 ? JOB) centered on the monitor of a computer. The experimenter instructed the participant to begin reading the operation-word pair aloud as soon as it appeared. If the participant paused before reading aloud, the experimenter explained again that pausing was not allowed. After reading the equation aloud, the participant verified aloud whether the provided answer was correct and then read the word aloud. Then, the experimenter pressed a key, and the following operation-word pair was presented. The sequence continued until three question marks cued the recall of the words presented in the set. Participants were requested to write the words on an answer sheet in the same serial order in which they had been presented. If participants made math errors frequently, the experimenter repeated that they could take as much time as they needed to answer the operation and that it was crucial that the answer was correct.

GOSPAN. Participant saw first the operation part of an operation-word string centered on the computer monitor (e.g., IS (9 : 3) + 2 = 5 ?). Under the operation the text '1 - CORRECT 2 - FALSE' was presented. Participants were instructed to begin reading the operation silently as soon as it appeared and to press the '1' or '2' key as soon as they had verified the answer. Instructions stressed that no additional pausing was allowed. If the response was not typed down within 6 s from presentation, a text line (in red, capital letters) appeared on screen that reminded participants to type down their response. After participants had typed down their response, the operation disappeared and the corresponding word was presented for 800 ms. Pilot work showed that all the operations could be solved within the 6 s interval. Likewise, the pilot study indicated that the 800 ms interval was sufficient to focus on and read the 'popped-up' word. The sequence continued until three question marks cued the recall of the words presented in the set. Participants wrote the words on an answer sheet in the same serial order in which they had been presented. After a participant finished writing down the words, he/she pressed a key to start presentation of the next set. After a sixth and ninth math error, a text that stressed the importance of solving the operations correctly appeared on screen (after recall and before the start of the next set).

Procedure

OSPAN. Twenty-three participants took the OSPAN before the GOSPAN, while the remaining 23 took both tests in reversed order. Five to 121 days intervened between the two test sessions. All participants were tested individually.

GOSPAN. Participants were tested in groups of 38 to 48 at the same time in a large computer room. Every participant took place behind a computer. Two experimenters answered possible questions during the instruction phase and checked whether participants, as instructed, only wrote down words when prompted by the three question marks.

Results and discussion

Four participants were discarded from the GOSPAN sample prior to any analysis because they were non-native Dutch speakers. As Engle et al. (1999), we also discarded participants that made more than 15% math errors. In the GOSPAN task this was the case for only one participant. This resulted in a total of 419 participants for the GOSPAN task. No participants were discarded in the OSPAN task.

GOSPAN. As reported above, all the 419 participants correctly solved the vast majority (85%) of the operation problems. This guarantees that the processing requirements of the WM task were met (Waters & Caplan, 1996): The storage of the words and recall performance could not be boosted by simply not spending resources to the operation processing.

A further control comes from the operation response latencies. We assume that most participants will comply with the instructions and start reading an operation as soon as it appears and give their response as quickly as possible. Therefore, if a participant is systematically pausing and rehearsing the presented words before giving his/her answer to the operation, this should result in increased response latencies.

The mean operation response latency in the GOSPAN was 4296 ms (SD = 855 ms). This is well within the range of the 6000 ms cut-off value that was used in the task to remind participants to respond to the operation. We decided to discard participants whose mean operation response latencies deviated by more than 2.5 standard deviations of the mean of the sample (= 6434 ms). This was the case for 13 participants (about 3% of the sample).

The mean GOSPAN score for the 406 remaining participants was 31.33 (SD = 10.17, top and bottom quartiles at 24 and 38). In order to further check the possibility of a general rehearsal bias in the GOSPAN we calculated the correlation between participants' mean operation response latencies and GOSPAN score. If high GOSPAN scores would simply result from rehearsal and thus spending more time at the operation, then we should see a positive correlation. However, results showed that there was no relation between latency and GOSPAN score (Pearsons product moment correlation, $\underline{r} = -.09$, n = 406, $\underline{p} > .05$).

Finally, we looked at the internal reliability of the GOSPAN test. The GOSPAN consists of three different presentations at each set size (e.g., from two to six items for recall). As Engle et al. (1999), we combined the first presentation of all the sets of different lengths into a single score, the second presentation into a single score, and the third presentation into a single score. We thus obtained three subscores that were used to compute Cronbach's alpha as a measure of reliability. The resulting alpha coefficient reached .74. This is comparable to the alpha of .69 that Engle et al. reported for the standard OSPAN and indicates that our GOSPAN measure is reliable.

GOSPAN and OSPAN. In order to test the validity of the GOSPAN task, the WM capacity of 46 participants was assessed with both the GOSPAN and standard OSPAN tasks. If the GOSPAN is a valid WM task, the scores on both tasks should be related. Results indicated that this is indeed the case. GOSPAN and OSPAN scores showed a correlation of .50 ($n = 46$, $p < .0001$). This is well within the range of the correlation between the OSPAN and other standard WM tests[2]. The correlation corrected for attenuation reached .70.

One should note that the raw correlation of .50 between OSPAN and GOSPAN is somewhat lower than the test-retest reliability that Klein and Fiss (1999) reported for the standard OSPAN (raw correlations ranged from .67 to 81). An important factor that has to be taken into account here is the time interval between the different testing sessions. Waters and Caplan (1996) suggested that longer inter-test intervals might reduce the correlations.

The different WM tasks in Engle et al. (1999) were always administered over the course of approximately seven days. Klein and Fiss (1999) used inter-test intervals of 21 to 49 days. Difficulties in participant recruitment resulted in considerably longer intervals (from five to 121 days) in our study. Although we did not keep track of the precise testing dates of every individual participant we could trace for which participants the inter-test interval did not exceed a two week term. When the analysis was restricted to these participants that were retested within 14 days the GOSPAN and OSPAN score correlation indeed increased, $r = .63$, $n = 27$, $p < .001$. Nevertheless, note that despite the larger test-retest interval variability for the complete set of participants the correlation was still as good as those between the standard WM tasks in Engle's study.

Interestingly, participants' GOSPAN scores were higher than their OSPAN scores [Mean GOSPAN = 27.96 vs. Mean OSPAN = 15.43, t-test for independent samples, $n = 46$, $p < .02$]. This was also reflected in a post-experimental question where most participants reported that the GOSPAN was easier. Participants indicated that reading the operations aloud in the OSPAN distracted them from the actual calculations. A higher load of the processing component of the WM task could indeed decrease the recall performance (see Waters & Caplan, 1996).

Recall in the OSPAN might also be harder because the reading aloud interferes with a rehearsal process that would otherwise facilitate recall (e.g., Beaman & Jones, 1998). Note that the reported GOSPAN rehearsal controls were aimed at identifying systematic 'inter-individual' rehearsal differences: We wanted to avoid a bias in GOSPAN scores due to the fact that some people would be deliberately rehearsing the to-be-remembered words while others were not. This does not exclude that due to the absence of an interfering reading aloud process all participants could benefit from a, possibly more automatic, rehearsal in the GOSPAN. As long as all participants would benefit equally from the rehearsal (i.e., the relative ranking of the participants on the GOSPAN and OSPAN tasks is maintained) this would not be problematic.

There was some support for the 'equal benefit' hypothesis in the data. The hypothesis implies that the increase in GOSPAN scores is similar for all participants. We tried to check this by classifying the 46 retested participants in a high and low span group based on their OSPAN score. Participants with an OSPAN score of 13 or less (n = 25) were classified as low spans and participants with OSPAN scores of 15 or more (n = 21) were classified as high spans. We ran a 2 (span group) x 2 (WM task) ANOVA on the number of correctly recalled words with span group as between-subjects factor and WM task (OSPAN or GOSPAN) as within-subjects factor. If everyone benefits equally well from the easier retrieval in the GOSPAN, the increase in the number of correctly recalled words from OSPAN to GOSPAN should not be affected by Span Group. Figure 1 shows the results.

Insert Figure 1 here

Not surprisingly there was a main effect of Span group and WM task: High spans scored better than low spans, $\underline{F}(1, 44) = 38.8$, MSE = 76.25, $\underline{p} < .0001$ and GOSPAN scores were higher than OSPAN scores, $\underline{F}(1, 44) = 76.13$, MSE = 45.38, $\underline{p} < .0001$. More important, the increase in recall scores did not significantly differ for the high and low span, Span group x WM task interaction, $\underline{F}(1, 44) = 2.05$, MSE = 45.38, $\underline{p} > .15$. This gives us some indication that people from different OSPAN levels do indeed benefit equally well from the easier recall in the GOSPAN.

An interesting extension of the current study would be to look at the correlation of the OSPAN and GOSPAN scores with a higher-order cognitive task. Since the ability to predict higher-order cognitive task performance is an important touchstone of a working memory test such a study would allow to play both tasks off against one another (i.e., test which part of the variation in the higher-order task performance both tasks account for). In the present study the OSPAN was used as criterion against which the quality of the GOSPAN was measured. However, it should be stressed that the computerized nature of the GOSPAN has also a clear advantage over the OSPAN: The GOSPAN ‘s stimulus presentation is standardized while some of the variability in the OSPAN can be attributed to the experimenter. For example, after a participant has read a word aloud, the timing of the presentation of the next operation-word string in the OSPAN depends on how fast the experimenter presses the ‘Enter-key’ on the keyboard. It also depends on the experimenter ‘s personal judgement whether or not a participant is starting to make too much reasoning errors or whether the participant is taking additional time for rehearsal. Furthermore, when the experimenter does decide to admonish the participant, the personal, face-to-face nature of this intervention can be quite intrusive for some participants.

Although we cannot directly compare the OSPAN and GOSPAN tasks, it is important to note that recent studies did successfully link GOSPAN performance to performance in such higher-order cognitive tasks as conditional reasoning (De Neys, Schaeken, & d'Ydewalle, in press, 2002) and semantic memory retrieval (De Neys et al., 2002.; Verschueren, De Neys, Schaeken, & d'Ydewalle, 2002). Together with the present results these findings further support the use of the GOSPAN task as a measure of WM capacity.

Conclusion

In this study we presented a Dutch, computerized and group administrable adaptation of the OSPAN task. Our GOSPAN task requires no active intervention of the experimenter and allows testing large groups at the same time. The task showed good reliability characteristics and GOSPAN scores correlated well with the standard OSPAN task. This provides researchers with a time-saving, reliable, and valid WM capacity measure.

We finally remark that since the only language specific component of the GOSPAN task are the words in the operation-word pairs, the task can be easily adopted for other language groups. This should allow a wide range of researchers to benefit from the proposed group administrable adaptations of the GOSPAN.

Acknowledgements

This research was supported by the Fund for Scientific Research Flanders. We would like to thank Philip Beaman and Randy Engle for valuable suggestions and the OSPAN material. The Dutch OSPAN and GOSPAN tasks can be obtained from the first author.

References

Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short–term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Baayen, R. H., Piepenbrock, R., & van Rijn, A. (1993). *The CELEX lexical database [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Barrouillet, P., & Fayol, M. (1998). From algorithmic computing to direct retrieval: Evidence from number and alphabetic arithmetic in children and adults. *Memory & Cognition, 26*, 355-368.

Beaman, C. P., & Jones, D. M. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51A*, 615-636.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (in press). Working memory and everyday conditional reasoning: A trend analysis. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.*

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). *Everyday conditional reasoning and working memory capacity: Retrieval and inhibition of background knowledge*. Paper presented at the joint meeting of the Experimental Psychology Society and the Belgian Psychology Society, Leuven, Belgium.

Engle, R. W. (2001). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*, 309-331.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General, 130*, 184-198.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled attention view of Working-Memory Capacity. *Journal of Experimental Psychology: General, 130*, 169-183.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336-358.

Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, & Computers, 31*, 429-432.

Klein, K., & Boals, A. (2001). The relationship of life event stress and working memory capacity. *Applied Cognitive Psychology, 15*, 565-579.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning is (little more) than working-memory capacity. *Intelligence, 14*, 389-433.

La Pointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118-1133.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. (1992). Individual differences in bridging inference processes. *Memory & Cognition, 20,* 539-548.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language, 28*, 127-154.

Uit den Bogaert, P. C. (1975). *Woordfrequenties in geschreven en gesproken Nederlands.* [Word frequencies in written and spoken Dutch]. Utrecht, Netherlands: Oosthoek, Scheltema & Holkema.

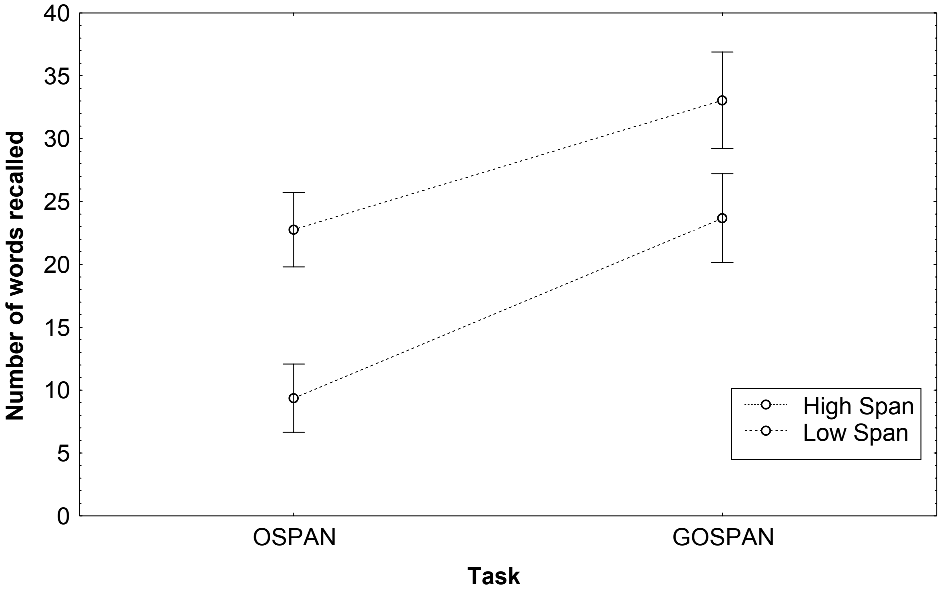
Verschueren, N., De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Working memory capacity and the nature of generated counterexamples. *Proceedings of the 24th Annual Conference of the Cognitive Science Society, 914-919*. Mahwah, NJ: Erlbaum.

Vingerhoets, G. (1993). A comparative study on word and digit span tasks in 4 and 7 year old children. *Psychologica Belgica, 33*, 99-113.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology, 49A*, 51-79.

Figure Captions

Figure 1. Mean OSPAN and GOSPAN score (number of words recalled) for participants classified as High and Low Spans. Vertical lines depict 95% confidence intervals of the means.



40
35
30
25
20
15
10
5
0
Number of words recalled
High Span
Low Span
OSPAN
GOSPAN
Task

Appendix

Table A1

Dutch words used in the OSPAN and GOSPAN operation-word strings, their word frequency (per 1 million words), and English translation.

Dutch word	English translation	Frequency[a]	Frequency[b]
altaar	altar	10	10
appel	apple	7	14
arm	arm	97	150
azijn	vinegar	5	9
bakker	baker	30	12
balans	balance	10	13
ballon	balloon	8	4
bank	bank	47	81
beker	mug	18	9
berg	mountain	22	21
blad	sheet	82	51
bloed	blood	82	121
bord	plate	30	38
bos	forrest	57	58
boter	butter	40	23
brief	letter	235	114
dak	roof	37	43
dorp	village	103	98
droom	dream	30	84
duivel	devil	13	37
eiland	island	43	52
emmer	bucket	5	13
feest	feast	50	40
fiets	bike	53	42
film	film	148	80
fles	bottle	33	74
fontein	fountain	8	6
foto	photo	47	55
gebak	cake	12	4
geld	money	253	276

gezin	family	132	86
gids	guide	8	19
glas	glass	73	124
gordijn	curtain	17	15
gras	grass	33	61
grendel	bolt	7	5
hart	heart	140	183
haven	harbour	43	29
hemd	shirt	20	21
hond	dog	67	107
hoofd	head	370	515
hotel	hotel	72	78
huis	house	507	541
kaart	card	45	53
kamer	room	222	318
kanaal	canal	25	16
karton	cardboard	7	7
kasteel	castle	33	30
kerk	church	368	170
ketting	chain	10	12
kind	child	445	454
kist	box	18	29
klas	class	42	40
kleed	robe	5	21
klok	clock	48	26
knop	button	23	17
koffie	coffee	122	111
kogel	bullet	8	16
koning	king	83	87
kraan	crane	7	10
kreet	yell	25	21
leger	army	95	62
lepel	spoon	13	11
leraar	teacher	43	35
lichaam	body	153	264
loper	picklock	5	4
lucht	air	137	185
maïs	corn	7	8
majoor	major	12	25
man	man	962	876
mantel	coat	18	16

massa	mass	43	48
melk	milk	22	51
mens	human	398	436
menu	menu	7	6
meter	meter	140	108
nacht	night	170	191
nota	note	52	33
ober	waiter	20	11
olie	oil	37	45
ontbijt	breakfast	8	27
oven	oven	15	11
paard	horse	55	99
papier	paper	73	78
piloot	pilot	15	11
plan	plan	143	137
planeet	planet	12	19
pop	doll	38	11
punt	point	163	144
radio	radio	65	49
regen	rain	50	53
rol	roll	143	183
rooster	grid	12	4
slot	lock	63	70
snor	moustache	10	14
spiegel	mirror	27	43
spijker	nail	7	5
spoor	track	33	49
stal	stable	23	14
steen	stone	38	58
stem	voice	180	265
stoel	chair	100	117
stof	dust	68	71
suiker	sugar	43	39
tafel	table	160	189
trap	stairway	85	90
verf	paint	15	26
vinger	finger	32	47
viool	violin	5	9
vogel	bird	62	35
vork	fork	12	10
vulkaan	volcano	7	4

wagen	car	60	191
wand	side	45	25
wens	wish	38	40
werk	work	547	496
wijzer	pointer	30	9
woord	word	335	282
wortel	carrot	27	13
zak	bag	37	66

[a]Word frequency in Uit den Bogaert (1975, a 600 000 words sample rescaled to frequency per million words). [b]Word frequency in the Celex database (Baayen, Piepenbrock, & van Rijn,1993).

Footnotes

[1] We selected short, concrete, and uncompounded words from different semantic categories. We avoided clear a priory associations between words to be presented in the same set. (e.g., JOB-OFFICE-DESK could not be in the same set). Given these criteria we tried to select words with the highest possible word frequencies. As in Vingerhoets (1993), word frequencies could not be lower than five per million (Uit den Bogaert, 1975). The selected words are presented in Table A1.

[2] Engle et al. (1999) reported a correlation of .51 between the OSPAN and Reading Span tests and a correlation of .47 between the OSPAN and Counting Span task.