

**A VALIDATION STUDY OF THE AGE-OF-ACQUISITION NORMS
COLLECTED BY
GHYSELINCK, DE MOOR, & BRYLSBAERT**

Wendy DE MOOR, Mandy GHYSELINCK & Marc BRYLSBAERT
Ghent University

Ghyselinck, De Moor & Brylsbaert (2000) collected age-of-acquisition (AoA) norms for 2,816 Dutch four- and five-letter nouns based on student ratings. To assess the validity of these ratings, we presented a sample of the words to children from kindergarten and the last year primary school. Overall, the validity data are in agreement with the rating data, so that the Ghyselinck et al. measures can be used for further research on the effects of AoA. In addition, the rated AoA norms correlate with young children's actual spoken language use, as assessed on the basis of the CHILDES data base.

In recent years, it has become clear that the age at which words are acquired, is a significant variable in word and picture processing: Words learned earlier in life are processed faster than words learned later in life (e.g., Gerhand & Barry, 1998; Morrison & Ellis, 1995). The effect of AoA has largely been ignored because it is heavily confounded with frequency of occurrence, and because of the limitations of the existing AoA measures. As for the latter, it has been objected that AoA measures are only available for a limited number of words, and that they are mainly based on ratings from a small number of students. As usual, the situation is worse for non-English languages than for English.

In Dutch, the only source of AoA estimates have been teacher ratings collected by Kohnstamm, Schaerlaekens, de Vries, Akkerhuis, and Frooninckx (1981; see also Schaerlaekens, Kohnstamm, & Lejaegere, 1999, for a recent update). These authors asked a representative sample of teachers from kindergarten and the first year primary school which words should be known by 6-year olds. Although the teacher ratings form an interesting source of information on AoA and have been used successfully in previous studies (Brylsbaert, 1996; Brylsbaert, Lange, & Van Wijnendaele, 2000; Brylsbaert, Van Wijnendaele, & De Deyne, in press), they are limited both in number and in scope (because they only provide estimates for the age of 6 years). Therefore, Ghyselinck, De Moor, and Brylsbaert (2000) recently

This research was made possible by a BOF-project granted to the authors by the Research Council of Ghent University. The authors thank the three anonymous reviewers for their helpful comments on earlier versions of the manuscript.

Correspondence concerning this article should be addressed to Wendy De Moor, Department of Experimental Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent. Electronic mail may be sent to Wendy.DeMoor@rug.ac.be.

collected AoA measures for 2,816 Dutch four- and five-letter nouns, and showed that these measures were reliable, at least for the Dutch speaking region of Belgium (due to their very high correlation with ratings previously compiled at the University of Leuven).

The present study was designed to assess the validity of the Ghyselinck et al. AoA measures, as it might be argued that retrospective student estimates, despite their high reliability, have limited validity because they are biased by other variables. For instance, students' ratings may be influenced by word frequency, because people tend to underestimate the AoA of frequent words and to overestimate the AoA of rare words (Brysbaert, 1996). In English, several of these validation studies about AoA-ratings have been run, generally with reassuring results, but to our knowledge, no such study has been done in Dutch.

Gilhooly and Gilhooly (1980) were among the first to examine the validity of AoA measures obtained from student ratings. In a first study, they selected 40 words from the Crichton Vocabulary Scale and 13 words from the Mill Hill Vocabulary Scale. For each word an objective AoA measure was available since the words in the scales were ordered according to the frequency with which children between 5 and 11 years (Crichton Vocabulary Scale) or children between 5 and 16 years (Mill Hill Vocabulary Scale) were able to explain the meanings of the words. Seventy undergraduate students were asked to indicate on a 9-point rating scale when they thought they first had learned these words. Using Pearson correlations, Gilhooly and Gilhooly found that rated AoA was the best predictor of the Mill Hill rank position ($r = 0.93$), followed by Thorndike-Lorge frequency ($r = -0.77$). Multiple regression analysis indicated that rated AoA was the only variable that made a significant independent contribution in predicting the Mill Hill rank position. In their second validation study, Gilhooly and Gilhooly presented 48 words for which Gilhooly and Hay (1977) had collected AoA ratings, to children of different ages and tested the children on the meaning of these words. According to the authors (pp. 108-109): "A generous criterion was used in assessing the answers since we were interested in whether each word was known rather than in the ability to give detailed dictionary answers." A word that was not correctly responded to by at least 50% of an age group was further tested in an older group. For each word, the average age at which 50 % of the children knew the word was determined and taken as an objective measure of AoA. The results showed that rated AoA was the best predictor ($r = 0.84$) of the objective AoA estimates, followed by Thorndike-Lorge frequency ($r = -0.68$). Multiple regression analysis again indicated that rated age was the only variable that made a significant independent contribution to the prediction of objective AoA. On the basis of these results, Gilhooly and Gilhooly (1980) concluded that adult ratings are valid indices

of AoA. Other evidence supporting the validity of rated AoAs has been obtained by Carroll and White (1973), Lyons, Teer, and Rubenstein (1978), Jorm (1991), Walley and Metsala (1992), and Morrison, Chappell, and Ellis (1997).

Morrison et al.'s (1997) study is the largest attempt ever made to collect objective AoA measures. In this study, 280 British children were asked to name 297 pictured objects. Two AoA scores were calculated. First, a logistic regression analysis was used to estimate the age at which 50% of the children would be able to name a picture correctly. Second, the AoA of a word was defined as the age at which 75% of the children named the item correctly. These objective norms were compared with rated AoA norms (based on 20 participants), and both yielded a correlation of .75. On the basis of this finding, Morrison et al. (1997) concluded that "objective measures should be used when available, but where not, our data suggest that adult ratings provide a reliable and valid measure of real word learning age" (pp 528).

The English findings suggest that adult ratings are fairly good estimates of when children acquire words. However, to our knowledge no such study has been done in Dutch yet. Therefore, if the AoA measures of Ghyselinck et al. are to be used in psycholinguistic experiments, a representative sample of them should be checked against the real performance of children. Such a validation study has its own limitations (the sampling of children, the way in which the knowledge is tested, the scoring, cohort differences, etc.), but provides invaluable information about the correspondence of the student ratings with the actual performance they are supposed to measure. Two empirical validation studies were run. In the first study, 80 children of kindergarten were tested on the meaning of 230 supposedly early- and 24 late-acquired words. In the second study, 172 children of the last year of primary school were tested on the meaning of 410 words that were either rated as early-acquired or late-acquired. The reasons for using these two different age groups are to examine (1) whether words rated as early acquired are indeed known by 6-year olds, and (2) to what extent words rated as late acquired are known by 12-year olds. Following Morrison et al. (1997), we expected early acquired words to be known by at least 75% of the children of 6 year, and we expected significantly lower percentages of children (both of 6 year and of 12 year) to know the late acquired words (preferentially not more than 25%).

In addition to the empirical validation studies, we also compared the Ghyselinck et al. ratings with spoken language of young children, as estimated on the basis of the CHILDES data base. This measure allowed us to check whether the scope of our AoA-measures is limited to word *definitions*, or can be generalised to word use in spontaneous speech.

Word Samples in Experiments 1 and 2

Since our primary concern is to have a set of words that enables us to use fully factorial designs in future research on the effects of AoA and frequency, the sample of words tested in the two validation studies was selected from an orthogonal combination of rated AoA and frequency. For all 2,816 words from the Ghyselinck et al. study, rated AoA was plotted against frequency. Next, four quadrants were formed in such a way that each quadrant contained at least 30 words. We sought to separate the quadrants as far as possible. Due to the scarcity of early acquired low-frequency words and late acquired high-frequency words, the borders of the quadrants had to be set at 7.5 years (upper border early acquired words) and 10 years (lower border late acquired words) with respect to AoA. The borders for frequency were chosen at 1.5 (upper border low frequency words) and 2.5 (lower border high frequency words). On the basis of these quadrants, we selected words that were likely to be useful in future experiments about the effects of frequency and AoA (see Figure 1 for the values of the actual words sampled)

Experiment 1 (with 6-year olds) mainly included words from the lower left quadrant (early-acquired, low-frequency words) and matched words (for AoA) from the upper left quadrant (early-acquired, high-frequency words). The stimulus set of the second study (with 12-year olds) contained stimuli from all four quadrants.

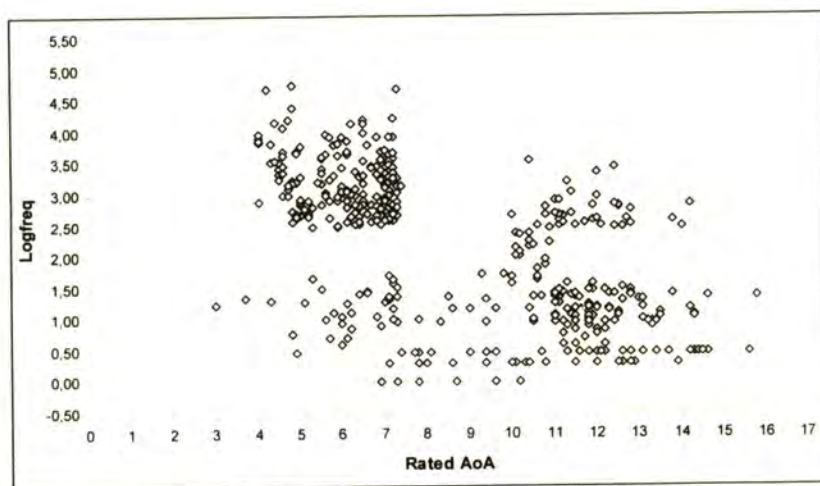


Figure 1. Rated AoA plotted against logarithmic frequency for all the words used in experiment 1 and 2.

Experiment 1

Method

Word Sample. We selected 230 early-rated words and 24 late-rated words from the Ghyselinck et al. data base. The early-acquired words had an estimated AoA below 7.5 years; the late-acquired words had an AoA of 10 years and more. These words were distributed over four lists, which were matched on AoA and frequency. Three lists contained 58 early-rated words and 6 late-rated words, and one list contained 56 early-rated words and 6 late-rated words. Our main objective was to see to what extent the early-acquired words were mastered by 6-year olds. The late-acquired words were mainly added as a check of the validity of our procedure (i.e., to ensure that the cues provided to the children were not of such a nature that the children could easily guess the meanings of the words without actually knowing them). In order not to discourage the children by presenting them too many words they did not know, and given the main purpose of the experiment, the number of late-acquired words was limited to 6 items per list. Each child got only one of the four lists.

Procedure. The children were tested individually in a quiet room. They were told that the experimenter would read a set of words and that their task was to explain the meaning of the words. They were further told that if they did not know the meaning of a word, the experimenter would read four sentences from which they could select the one that in their view provided the right description of the word. The definitions of the words and the three wrong alternatives were selected from the *Van Dale Junior Woordenboek*. It took on average 20 minutes for the children to complete the list. Two different orders of the words were used.

Participants. Eighty children (43 girls and 37 boys) from kindergarten (mean age 5.6 years, ranging from 5 to 6 years; testing in the months April and May) participated in the study. The children were drawn from 4 different schools in and around the city of Ghent. All children were native speakers of Dutch and none was bilingual. We took care that none of the participating children followed special courses for Dutch.

Scoring. All responses were tape-recorded. Three independent judges (graduate students experimental psychology) listened to the tapes and indicated for each word whether the child knew the word right away, after the presentation of a cue (the four alternatives), or not at all. The judges were instructed to focus on whether they thought the child understood the word,

and not on whether the child was able to give a detailed dictionary definition (e.g., when presented with the word 'oksel' [armpit] many children pointed to their armpit – this was mentioned by the experimenter on the tape – and accordingly the word was scored as known). The judges were unanimous in their evaluations on 4,642 of the 5,088 ratings (90%). The judges were paid for their participation.

Results

A full matrix of all words with their rated AoA and logarithmic frequency and the evaluation of the three judges is available on the internet (<http://allserv.rug.ac.be/~hmaessen/vakgroep/> go sub *Research, available documents, data, etc.*). For each word, we determined (1) the percentage of children that knew the word without a cue, (2) the percentage of children that knew the word after a cue was given (the four alternatives), (3) the percentage of children that gave a wrong answer to the four alternatives, and (4) the percentage of children that refused to give an answer. These percentages are calculated on all observations per word (i.e., the number of children times 3 judges). Because it turned out that many cued answers were given by children who were too shy to say the answer right away, we will base our objective AoA measure on the sum of correct cued and uncued responses (see also Morrison et al., 1997).

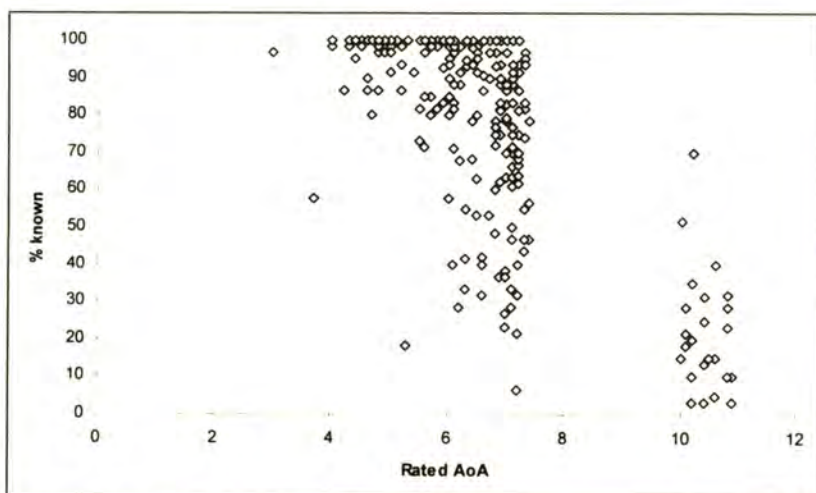


Figure 2. Scatter plot of rated AoA versus the percentage of 6-year olds that knew the meaning of 230 early rated and 24 late rated words.

Figure 2 shows a scatter diagram of rated AoA versus the percentage of children that knew what the word meant.

Words rated as early acquired. Of the 230 early-rated words, 55 (24%) were known by less than 75% of the children. Most of these words had a rated AoA of more than 6 year, which is not surprising, given that the children in this study had a mean age of 5.6 years. Of the 95 words with a rated AoA of 6 years or younger, only 5 items failed to reach the 75% criterion. These were the words 'toet' (58%), 'luier' (18%), 'zaal' (73%), 'getal' (71%), and 'klank' (58%). The reason why the word 'toet' was not known by 75% of the children probably is that the students had interpreted the word as an onomatopoeia [toot], whereas the first definition of the word (and the one we asked for) is an informal word for 'face', which is rarely used in the Dutch-speaking part of Belgium. The low percentage of the word 'luier' [nappy] probably reflects the fact that the word has become old-fashioned, because it refers to a linen, washable nappy, whereas nowadays most families use disposable nappies, which are called 'pampers'. Despite the high percentages of 'zaal' [large room] and 'getal' [number], they fell short of the 75% criterion. It may be noted that the word 'zaal' is mostly used in compound words (e.g. 'eetzaal' [dining-room]). As a result children may be more aware of the meaning of the word when it is used in a compound word. During the session, many children confused the word 'getal' [number] with the definition of 'letter' [an alphanumerical letter]. For example, when children were asked to explain the meaning of the word 'negen' [nine] they often said this was a letter, confusing the concept of letter with the concept of number. Given the high percentages of children that knew the words 'negen' [nine] and 'zeven' [seven], it is clear that the children had a notion of what numbers are but they could not explain the concept yet. The low percentage of the word 'klank' [sound] may be due to the fact that other words (e.g., 'geluid', 'toon') are used to refer to the meaning of this word.

Words rated as late acquired. Of the 24 words rated as late acquired (AoA of 10 years or older), 8 words were known by more than 25% of the children. These were the words 'cocon' (52%), 'list' (29%), 'motel' (70%), 'sauna' (35%), 'wals' (31%), 'sfinx' (40%), 'fuif' (32%) and 'kiosk' (28%). A closer examination of the results showed that these high percentages were mainly caused by the percentage of children that knew the word after a cue (the four alternatives) was given. Except for the word 'cocon' [cocoon], no word reached a percentage of more than 25% when no cue was given. This might indicate that the words rated as late acquired are not actively known by 6-year olds but that some children were able to select the right alternative when cued with four alternatives. The high percentages in the uncued condition for

the word 'cocon' were to a large extent due to the fact that in one of the classes children had recently learned about the butterfly. As a result, many of the children knew very well what a 'cocon' was.

Regression analyses. Finally, we repeated Gilhooly and Gilhooly's (1980) analysis and looked at the correlations between rated AoA, log frequency, and percentage of children that knew the word. In this analysis, the 5 early-acquired words that failed to reach the 75% criterion were omitted. The correlation between rated AoA and percentage known was .75 ($N = 249$), the correlation between log frequency and percentage known was .26. A multiple regression analysis showed that only rated AoA had significant predictive power for percentage known (57% of the variance explained). Log frequency added less than 1% of the variance explained ($t(246) = 1.12$, $p > .20$). Similarly, log frequency explained only 0.8% of the variance in AoA ratings, when percentage known was partialled out, although the contribution this time was significant at the .05 level ($t(246) = -2.17$, $p < .05$).

Discussion

Experiment 1 showed that the AoA norms collected by Ghyselinck et al. agree very well with the actual performance of children at the age of six. Only for a very small percentage of words (some 5%) there is a discrepancy, which most probably is due to cohort differences. The clearest example of this is the word 'luier' which was 15 years ago still quite common, but now is virtually replaced by the word 'pamper' (although 88% of the teachers in Schaerlaekens et al., 1999, still think the word 'luier' ought to be known by 6-year olds). Other examples are objects that now figure in movies and cartoons for children (e.g., 'sfinx'). So, despite the limitations of both the ratings (i.e., retrospective estimates by students) and the current validation measures (i.e., being able to define a spoken word in a way that convinced the independent judges), both measures to a large extent assessed the same variable, which can minimally be described as the order in which children acquire words.

Experiment 2

Method

Word sample and procedure. Stimulus materials consisted of 410 words coming from the four quadrants indicated in Figure 1. For the early-acquired

words there was an overlap with the words from Experiment 1, but this overlap was not complete. The 410 words were distributed over four lists (2 lists of 103 words and 2 lists of 102 words) matched for AoA and frequency. The words were printed in a random order, 10 words per page. Together with a word, three possible explanations of the word and the option 'I don't know what a ... is' were presented. The definition of the test word and the two alternative definitions were selected from the *Van Dale Junior Woordenboek*. Three different permutations of the words and two different permutations of the alternatives were used.

The lists were randomly assigned to the children in a class, one list per child. The children filled in the list under supervision of the teacher and the experimenter. The children were instructed to indicate for each word the correct meaning by ringing one of the alternatives. The children got as much time as they needed to complete the list. It took on average 40 minutes to do so. It may be remarked that by using this procedure we assessed the meaning of written words instead of spoken words. However, because there was no indication that a significant proportion of the children had reading difficulties and because the children could work at their own pace (and ask for help if they could not read a particular word), we do not think there are a priori reasons to believe that a written version of the test would yield other results than a spoken version of the test.

Participants. Participants were 172 children from the sixth grade of primary school. The children were drawn from rural and city schools in and around Ghent. All the children were native speakers of Dutch and none was bilingual. We took care that none of the children followed special courses Dutch. For privacy reasons, we were not allowed to ask the age of the children, but it can be estimated around 11.5 years (testing time: April – May).

Results

An alphabetical listing of the 410 validated words together with their rated AoA and the percentage of children that correctly indicated the meaning of the word is available on the internet (<http://allserv.rug.ac.be/~hnaessen/vakgroep/> go sub Research, available documents, data, etc.). For each word, we mention (1) the percentage of children that correctly indicated the meaning of the word, (2) the percentage of children that did not know the meaning of the word and ringed the option 'I don't know what a ... is', and (3) the percentage of children that chose the wrong alternative. As only a few children chose a wrong alternative when

they did not know the word, we did not use a guessing correction (e.g., the number of correct answers minus half the number of wrong answers; these measures can easily be computed on the basis of the data on the web if authors have a preference for using them). A scatter plot of rated AoA versus percentage of children that correctly indicated the meaning of the words is shown in Figure 3.

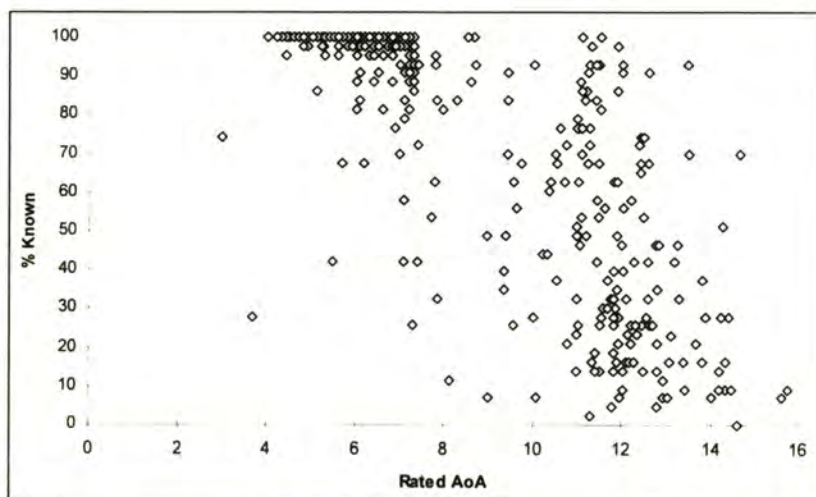


Figure 3. Scatter plot of rated AoA versus the percentage of children from the sixth grade of primary school that correctly indicated the meaning of 410 early-rated and late-rated words.

Words rated as early acquired. Of the 96 words with an AoA of 6 years or younger, only 4 words were known by less than 75% of the children, namely 'tuut' (74.4%), 'toet' (27.9%), 'kwak' (41.9%), and 'slab' (67.4%). If the words with a rated AoA of 7 or younger were included, two more items did not reach the 75% criterion: 'brei' (69.8%) and 'kadee' (67.4%). The words 'tuut', 'toet', and 'kwak' are all best known as onomatopoeias, rather than as nouns (see also Experiment 1). Our data, therefore, indicate that one should be frugal when considering this type of words as potential stimulus materials in an experiment on word recognition. Words as 'slab' [bib] and 'brei' [knitting] are words that are generally used in a longer form, such as 'slabbetje' and 'breiwerk'. As a result, these words are probably less known to the children than their longer versions. No explanation was found why the word 'kadee' [rogue] did not reach the 75% criterion, although it may be that its use is declining and mainly used by older speakers.

Words rated as late acquired. Of the 230 words with a rated AoA higher than 7 years, only 81 words (35%) were known by more than 75% of the 12 year old children. There was a general decrease of percentage known as a function of rated AoA, going from 86% for AoAs between 7.1 and 8.0, to 63% for AoAs between 8.1 and 9.0, and less than 30% for AoAs larger than 9.

Regression analyses. The correlation between AoA and the percentage of children that knew the word, was 0.81 ($N = 404$) when the six 'suspicious' early-acquired words were discarded. The correlation between log frequency and the percentage of children that knew the word, was .63. A multiple regression analysis indicated that both AoA and log frequency were predictors of the percentage known, respectively accounting for 65% of the variance ($t(401) = -18.41$) and 3.4% of the variance ($t(401) = 6.64$, $p < .01$). Similarly, rated AoA was predicted by both percentage known (65% of the variance, $t(401) = -18.41$), and log frequency (1.8% of the variance, $t(401) = -4.66$).

Discussion

In general, the results of Experiment 2 corroborated those of Experiment 1, although the picture was slightly less clear than we had hoped for. In particular, frequency started to slip in, and words were less well known than they should have been according to their rated AoA (i.e., many words with a rated AoA below 12 were not known by 75% of our participants). On the basis of this single experiment, it cannot be decided to what extent the procedure (especially the group testing) accounted for the problems, or whether they reflect genuine difficulties with AoA measures.

One reason why AoA ratings may be influenced by frequency-of-occurrence for words that have been acquired after the age of 8, is that the order of acquisition becomes more heterogeneous for late-acquired than for early-acquired words, with larger interindividual differences for words acquired at the end of the primary school than for words acquired in kindergarten. Also, it is a general finding in serial recall that performance is better for the first and the last stimuli than for the middle stimuli. For instance, Crowder (1993) reported that American students are more accurate to recall the first presidents and the last presidents of the USA than the presidents in-between. These factors may explain why frequency had an influence on the AoA ratings of the students and on the actual performance of the children when later acquired words were included in the stimulus materials. Still, it should be noted that the contribution of frequency,

although significant, was very small in practical terms (less than 2% of the variance of the rated AoAs).

The multiple regression analyses only capture the order in which words have been acquired. Independent of this issue, it also looks like the students underestimated the AoA of many words learned after the age of 6. Morrison et al. (1997) also made this observation. Comparing rated AoAs with objective data, they found that raters underestimated the age at which they had learned late-acquired words. A possible explanation for this shift may be that the students used their AoA ratings of early-acquired words as a reference point against which they rated the late-acquired words. As a result they presumably gave late-acquired words an AoA rating close to the AoAs of early-acquired words, resulting in lower AoA ratings than suggested by the objective data. Another possibility is that the students in the study of Ghyselinck et al. (2000) gave a lower AoA to late-acquired words because they indeed learned these words at a younger age than indicated by the 12-year old children of Experiment 2. When we distributed our lists, we used no criteria to select pupils, but we handed the lists out to all the children of a class, regardless of their school results. Yet, only some 20% of these children will eventually go to university. Hence, the first-year university students who formed the basis of the Ghyselinck et al. norms, probably corresponded more or less with the upper fifth of the pupils in the primary school classes. A solution could be to lower the inclusion criterion for determining the objective AoA. Gilhooly and Gilhooly (1980), for instance, used a cut-off point of 50% to calculate the real AoA.

Study 3: Spontaneous Speech

An objection to our research thus far may be that we have validated the Ghyselinck et al. norms against word definitions only. Some authors (e.g., Snow, 1990) have argued that knowing words and being able to define words may be separate skills, because a child can have full acquisition of a word without being able to define the word. Although one would assume that there is a high correlation between using words and knowing what those words in general mean, it is interesting to examine whether the AoA ratings of Ghyselinck et al. also predict young children's word use in spontaneous speech over and above the effect based on 'adult' word frequencies.

For Dutch speaking children up to 4 years, there are several databases available on the internet. The best known is CHILDES (<http://childes.psy.cmu.edu/>). For the purpose of the present study, we selected 10 children (Arnold, Diederik, Maria, Gijs, Joost, and Katelijne – studied by Schaerlaekens – and Abel, Josse, Tomas, and Daan – data

collected by Bol, Krikhaar, & Wijnen). For each of the 2816 words studied by Ghyselinck et al., we looked whether we could find this word in the speech of the children (collapsed over the different files per child). Flexions of the words (i.e., plurals and diminutives, but not compound words) were included.

Figure 4 shows the data for all the Ghyselinck et al. words. A multiple regression analysis was run on those words that were present in the speech of at least one child ($N = 529$). The correlation between rated AoA and the percentage of children that used the word was $-.53$, against a correlation of $.38$ between word frequency and the percentage of children that used the word. In the multiple regression analysis, rated AoA accounted for 28% of the variance ($t(526) = -11.37, p < .01$), while word frequency accounted for another 3% ($t(526) = 4.71, p < .01$).

When the analysis was restricted to the 254 words of Experiment 1 (see Figure 5), 13 of the words with a rated AoA of 10 were present in the speech of one of the children. These were the words 'genie, kuif, shit, silo, moor, wals, speen, ditje, talk, kram, meute, shawl, drab'. At least four of these words, however, were mere repetitions of words previously said by the adult talking to the child. The correlation between rated AoA and percentage used was $-.49$, and the correlation between word frequency and percentage used was $.28$ (based on the 127 words that were present in the speech of at least one child). AoA accounted for 26% of the variance in the multiple regression

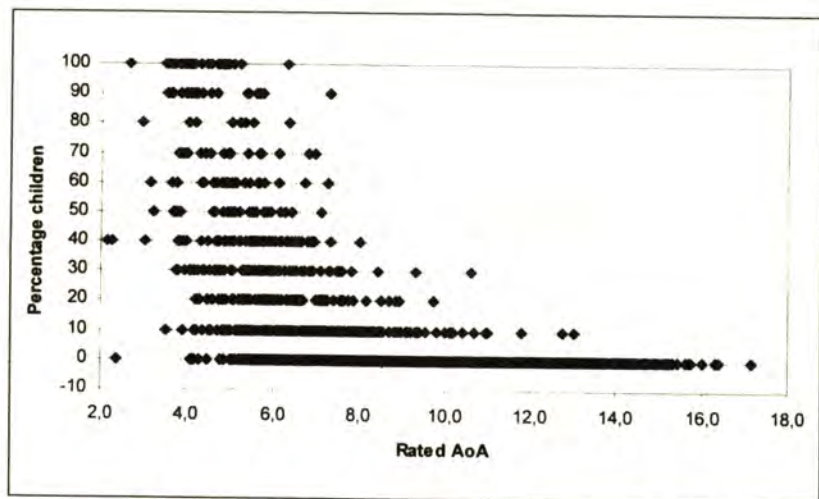


Figure 4. Percentage of children that used the 2816 words studied by Ghyselinck et al. versus rated AoA.

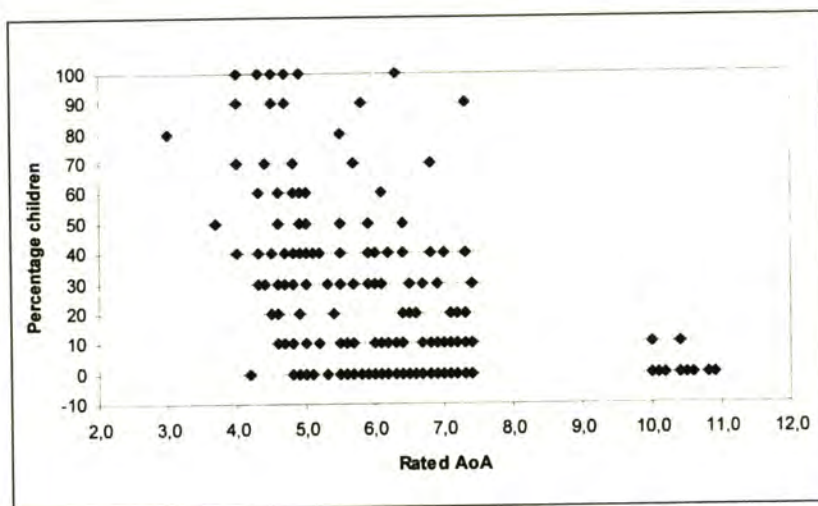


Figure 5. Percentage of children that used the 254 validated words plotted against the rated AoA of these words.

analysis ($t(124) = -5.73, p < .01$); word frequency accounted for another 2% ($t(124) = 2.25, p < .05$). If rated AoA was dependent variable, only the percentage of children was a significant predictor ($t(124) = -5.73, p < .01$), accounting for 24% of the variance; word frequency had no significant contribution ($t(124) = -1.22, p > .22$).

All in all, the comparison of rated AoAs and words used by young children confirm the thesis that students' estimates of age-of-acquisition have predictive value for the words that will appear in spontaneous speech of young children, and that this effect cannot be reduced to a frequency effect (based on the Celex frequency counts). At the same time, our analysis shows that the CHILDES data base can be used as another source to estimate the AoA of words (for these data, see our internet web site), although this measure too suffers from some limitations. One of the limitations is that, in general, words that are used frequently have a higher chance to figure in the speech excerpts. This implies that researchers are likely to run short of words if they are searching for low-frequency words. Also, it must be taken into account that the children on whom the studies are based, come from a restricted socio-economic segment of the society (many have parents with a university degree). Finally, some more research may be indicated to search for the best AoA measures on the basis of spontaneous child speech. Is one occurrence in the speech of one child enough to conclude that a word is used at that age? And what to do with words that are not present in the speech:

Are they not used, or did they not get a chance in the topics that figured in the talks?

Conclusion

In this study, an attempt was made to validate a sample of the AoA norms gathered by Ghyselinck, De Moor and Brysbaert (2000). The sample consisted of words that are potentially very useful in further experiments on the effects of word AoA and word frequency, because they allow an orthogonal variation of these variables. The results of the two empirical validation studies and the search in the CHILDES data base can be summarised as follows: (1) Except for a few outliers which can be accounted for, words rated as early acquired are indeed known by 6-year olds, (2) words that have been rated as late acquired are indeed late learned words, not even known by many 12-year olds, and (3) words rated as earlier acquired have a higher chance of being used by 4-year olds in spontaneous speech. Given these results and the fact that frequency had but a very small effect on the AoA ratings, once the objective AoA was partialled out, we can safely conclude that the AoA norms collected by Ghyselinck et al. are valid measures of the age at which a word is acquired and can be used in future research on the effects of AoA.

References

- Brysbaert, M. (1996). Word frequency affects naming latency in Dutch when age of acquisition is controlled. *European Journal of Cognitive Psychology*, 8, 185-193.
- Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from Dutch. *European Journal of Cognitive Psychology*, 12, 65-85.
- Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (in press). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*.
- Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 25, 85-95.
- Crowder, R. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, 21, 142-145.
- Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology*, 24, 267-283.
- Ghyselinck, M., De Moor, W., & Brysbaert, M. (2000). Age-of-acquisition ratings for 2816 Dutch four- and five-letter nouns. *Psychologica Belgica*, 40, 77-98.

Gilhooly, K. J., & Gilhooly, M. L. M. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology*, 71, 105-110.

Gilhooly, K. J., & Hay, D. (1977). Imagery, concreteness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behaviour Research Methods & Instrumentation*, 9, 12-17.

Jorm, A. F. (1991). The validity of word age-of-acquisition ratings: A longitudinal study of a child's word knowledge. *British Journal of Psychology*, 9, 437-444.

Kohnstamm, G. A., Schaerlaekens, A. M., de Vries, A. K., Akkerhuis, G. W., & Frooninckx, M. (1981). *Nieuwe streeflijst woordenschat*. Lisse: Swets & Zeitlinger.

Lyons, A. W., Teer, P., & Rubenstein, H. (1978). Age-at-acquisition and word recognition. *Journal of Psycholinguistic Research*, 7, 179-187.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age-of-acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology*, 50A (3), 528-559.

Morrison, C. M. & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21, 116-133.

Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen: Derde herziene versie gebaseerd op nieuw onderzoek in Nederland en België*. Lisse: Swets & Zeitlinger.

Snow, C.E. (1990). The development of definitional skill. *Journal of Child Language*, 17, 697-710.

Walley, A. C., & Metsala, J. L. (1992). Young children's age-of-acquisitions estimates for spoken words. *Memory and Cognition*, 20, 171-182.

Received September, 1999

Revision received March, 2000

Accepted May, 2000