

Confidence in age-of-acquisition estimates and its relationship to children's labeling performance

JOHN J. WINTERS, JR., and LARAINÉ WINTER
E. R. Johnstone Training and Research Center, Bordentown, New Jersey 08505

and

AGNES LIN BURGER
New York University, New York, New York 10003

College students estimated the age-of-acquisition (AA) of the names of 456 colored picturable objects and gave their confidence in those estimates. From these stimuli, 414 with U values ≤ 1.0 were selected for analyses. When compared with males, females estimated that the picture names had been acquired at an earlier age, and they were more confident in their AA estimates. The earlier the estimated AA, the more confident both groups were in their estimations. Labeling performance of the same stimuli by 5-year olds (Winters & Brzoska, 1975) was related to adults' estimated AA. Words that were labeled most efficiently by children were estimated to have been acquired earlier by adults. Sex differences were interpreted in terms of different strategies of estimating AA.

Studies investigating the age of acquisition (AA) of words require adult subjects to estimate the age at which they first experienced a word. The age range of estimates extends from age 2 through adolescence. There are two general purposes for these investigations. One is to determine the degree of relationship between AA and other variables, such as familiarity, frequency, imagery, concreteness, and meaningfulness (Gilhooly & Hay, 1977; Stratton, Jacobus, & Brinley, 1975). The other is to determine the effects of AA on the retrieval of those words from semantic memory (Carroll & White, 1973a, 1973b; Lachman, 1973; Lachman, Shaffer, & Hennrikus, 1974).

Generally, it has been found that all of the above variables are associated significantly with AA, and that the AA and uncertainty (U) of a word are related to the retrieval of information from lexical store. Words estimated to have been acquired earlier are retrieved faster than words estimated to have been acquired later; words (names of visual displays) for which there is greater consensus (low U) are retrieved faster than words or names for which there is less consensus (high U). The theoretical consequence of such investigations is that, if the variables that can best define the speed of retrieval of items from memory can be determined,

The authors are grateful to Derril McGuigan from Rider College and Allan Waterman from Trenton State College for their cooperation. We also thank Burt Holland for his assistance in analyzing the data and Herman Spitz for his critical reading of the manuscript. This study was supported in part by National Institute of Child Health and Human Development Grant HD 07329.

then more efficient models of semantic memory can be constructed.

During language development, as words enter that component of memory that contains the "dictionary" or list of names designating nameable patterns, the lexicon, they enter a steady-state storage system that is relatively stable (Lachman, 1973). Assuming a common sequence of words entering into the lexicon of native-speaking Americans, it should be possible to determine that sequence of word inclusion in the lexicon of this population during development.

In an attempt to determine that sequence, using labeling performance as an indicator of word acquisition, Winters and Brzoska (1975) tested children and adolescents on a task that required they label each of 480 colored pictures. A positive relationship between age and labeling performance was found. Labeling performance by their 5-year-olds was related to the AA estimates given by Carroll and White's (1973a) adult subjects ($r = .70$) for 117 items, indicating that adult estimates of AA are valid to the extent that they represent labeling performance by 5-year-old children for those words. A comparison between adults' AA estimates and 5-year-olds' labeling performance on a larger corpus is needed to verify this relationship and to examine further the degree of agreement between these two groups as to the sequence of word inclusion into the lexicon.

The finding that U also influences speed of retrieval of information from lexical store (Lachman, 1973; Lachman et al., 1974) suggests that normative data using AA estimates consider this variable. Inasmuch as

one purpose for collecting AA data is to use them in reaction time studies that investigate speed of retrieval in words from the lexicon, and U influences retrieval time, a measure of U should be obtained for each item.

Both Carroll and White (1973a) and Lachman et al. (1974) state that other variables may introduce a source of bias in AA estimates. One that has not yet been investigated is the subject's confidence of those estimates. Our subjects were requested to give a confidence (C) rating with each AA estimate and then, upon completion of the task, rate the difficulty of different aspects of the task.

METHOD

Four hundred and fifty-six slides were selected from the original pool of 480 colored slides of picturable objects (Winters & Brzoska, 1975). Twenty-four slides were excluded because of possible ambiguity. There were 76 slides in each of six Carousel trays. The adult subjects were tested in groups, and each group viewed one tray of slides. There were 298 subjects (approximately 50 per tray) with approximately equal sex distribution for each tray.

As each slide was projected onto a large screen at the front of the testing room, the subjects were instructed (1) to write the name of each picture without concern for spelling, (2) to estimate the age at which the word was first learned (Carroll and White's, 1973a, instructions and their nine-level age code were used), and (3) to indicate the confidence in that age estimation using a 7-point scale, with 1 indicating least confidence and 7 indicating most confidence in that estimate. The subjects then responded to six questions concerning the difficulty of various aspects of the task, again using a 7-point scale.

RESULTS

The labeling values for Winters and Brzoska's (1975) 5-year-olds (5YO), which represent the proportion of children who correctly labeled each item, were transformed to their arcsins. The measure for U was the same as that used by Lachman (1973). Preliminary analyses indicated that the U values for the 456 stimuli ranged from 0 to 2.56. As one purpose of this study was to obtain a list of items for which there was a high level of consensus as to the name of each picture (low U), 42 pictures with U values greater than 1.0 were eliminated from further analyses.

Intergroup comparisons on AA, C, and U revealed that females and males did not differ on U, but did on AA and C. The females estimated they had acquired the list of words approximately 5 months earlier than the males [$t(412) = 10.12, p < .01$] and had greater confidence in their ratings [$t(412) = 3.16, p < .01$]. Intergroup reliability for AA estimates was .93, for C ratings, .65, and for U, .61 ($p < .01$).

Correlations among the four variables of interest, including 5YO labeling values, were determined. Cohen's (1969) criteria of effect size indicated that the relationship between U and the other variables was small ($r < .30$), between C ratings and 5YO values was

Table 1
Correlations Between AA, C, 5YO, and U by Sex

	Females			Males		
	C	5YO	U	C	5YO	U
AA	-.75†	-.58†	-.12*	-.54†	-.57†	-.02
C		.47†	-.14**		.35†	-.13**
5YO			-.27†			-.19†

* $p < .05$ ** $p < .01$ † $p < .001$

medium ($.29 < r < .50$), and between AA estimates and 5YO values, as well as C ratings, was large ($r > .49$). C ratings and AA estimates had the highest association. The correlations among the four variables by sex are given in Table 1.

The greatest diversity between the two groups was in the relationship between their AA estimates and C ratings. Relationships between those ratings for the two groups when the AA estimates were in the lower or upper third of the AA range indicated that, in the lower third, the correlations were significant [$r(136) = -.66$ and $-.45, p < .01$, for the females and males, respectively] and differed significantly [$z = 2.55, p < .05$]. In the upper third of the range, only the females produced a significant correlation [$r(136) = .28, p < .01$], but did not differ significantly from the males. A comparison of the C ratings in the lower and upper thirds of the AA range for both groups revealed that the females had higher C ratings than males [$F(1,274) = 3.80, p < .05$], the C ratings were higher in the lower than in the upper third of the range [$F(1,274) = 531.37, p < .01$], and the females were more confident of their estimates when those estimates were in the lower third of the AA range [$F(1,274) = 38.37, p < .01$]. Thus, the disparity between sexes in their confidence as to their AA estimates was due largely to the greater confidence the females had for words learned at an earlier age.

A stepwise regression analysis, using AA as the dependent measure and allowing the variables C, 5YO, and U to enter the analysis in the order in which they contributed to the variance, revealed that for females, C entered first ($R = .75$), 5YO entered second ($R = .80$), and U entered last ($R = .80$). For the males, 5YO entered first ($R = .57$), C entered second ($R = .68$), and U entered last ($R = .69$). The contribution of the first two variables to the total variance was significant for both groups ($p < .01$).

The six questions on the posttest questionnaire dealt with overall difficulty of the task, difficulty in the use of Carroll and White's (1973a) AA code and the C rating scale, the degree to which they guessed at their AA estimates, whether the spoken or written word was used more often as the guide for determining acquisition of the words, and whether there was more confidence in AA estimates for words judged to have been acquired earlier or later in life.

The results indicated that neither group found the

task difficult and both assessed it to be easier than would be expected by chance [$t(297) = 6.67, p < .01$]. Both groups judged the AA code to be no more difficult to use than would be expected by chance. Males found it more difficult to apply the C scale than did females [$t(296) = 2.83, p < .01$]. More guessing as to their AA estimates occurred in both groups than would be expected [$t(297) = 2.87, p < .01$]. To give their estimates of AA, both groups relied more on when a word was first spoken than when it was first written [$t(297) = 15.75, p < .01$]. Both groups were inclined to be more confident in their earlier AA estimations, but females indicated they were more confident than males in their estimates for words acquired earlier in life [$t(296) = 4.67, p < .01$]. This is in agreement with the findings that, although both groups had more confidence in their AA estimations of words acquired earlier in life, the females exhibited an even greater relationship between confidence and earlier acquired words.

DISCUSSION

The relationship between AA estimates and U was small and U contributed negligibly to the variance of AA (less than 1%). Names of pictures that were estimated to have been acquired early (from approximately 2 years of age) were as diverse as names of pictures estimated to have been acquired late (to approximately 10 years of age). A higher consensus as to the name of a picture when there was greater confidence as to the age at which the picture's word was acquired (U vs. C) might have been expected, indicating greater agreement as to the pictures' names when the subjects were more confident of their AA ratings. The restricted range of U values did attenuate the associations between U and AA estimates and C ratings to some extent. However, when all 456 stimuli are considered, in each comparison the effect size was still small. Lachman et al. (1974) reported that both AA and U contribute independently to lexical retrieval time. This is yet another situation in which this relative independence is demonstrated.

The high relationship between AA estimates and C ratings implies that response latencies to earlier or later acquired words should parallel response latencies when there is greater or less confidence as to when the word was acquired with the greater confidence (lower AA estimates), resulting in shorter response latencies. Studies investigating response latency to AA estimates have not as yet taken into account subjects' confidence in those estimates, nor have studies comparing AA estimates with other variables considered subjects' confidence in the ratings (cf. Gilhooly & Hay, 1977; Stratton et al., 1975). Until confidence in ratings is accounted for, it will remain uncertain as to whether semantic retrieval time or the relationships among the several variables reflect measurements of those variables of interest or the confidence of the values upon which the measurements are based.

The relationship between the AA estimates given by our adults and the labeling values of 5-year-old children supports the notion that the sequence of word acquisition for these words during development is similar to adults' estimates of that sequence. The relationship was not as high as found previously ($-.70$) when 5YO values were compared with other adults' estimates as to that sequence (Carroll & White, 1973a). This lower relationship could be due either to the different AA estimates given by the adults in this study or to the restricted distribution of AA ratings caused by the elimination of stimuli that had high AA estimates as well as high U values.

A comparison between AA estimates given by the adults in the present as well as Carroll and White's (1973a) study on the remaining 106 items common to both studies produced an r of $.90$. It is unlikely that interstudy response variability caused the lower relationship in this study between AA estimates and 5YO values. When the 5YO values and the AA estimates given by our and by Carroll and White's (1973a) adults were compared on the 106 items, the resulting r s were $-.60$ and $-.56$, similar to those found here for all 414 stimuli. Thus, the elimination of higher AA estimates apparently produced the lower relationship found here between the empirical sequence of word acquisition and the estimated sequence.

The findings that the females' lower AA estimates and higher confidence in those estimates, as well as the fact that their C ratings contributed more to the variance of AA than 5YO values, requires explanation. It is possible that the females learned the words in this list earlier than did the males and were more accurate in their retrieval, giving them greater confidence in their judgments. This explanation is unlikely, however, as (1) there was no sex difference for the 5-year-old children on the entire list (Winters & Brzoska, 1975), indicating that those females had not acquired these words earlier than males, and (2) the association between AA estimates and 5YO values were almost identical for both females and males. If retrieval by the females had been more accurate, their sequence of estimates of AA would have complied more with the ability of children to label the items.

On the other hand, the females' greater exposure to their or other's children could have resulted in a greater awareness of words spoken by children, and would explain the sex differences found here. If this is the reason, greater confidence in the AA estimates should occur because experience with the words is more recent. AA estimates, although lower, may be more veridical. A higher relationship between C ratings and 5YO values should also exist for the females if the ratings are based on the vocabulary of today's children. It was found that females, in fact, did have a significantly higher relationship than males when C ratings were compared with AA estimates and 5YO values (see Table 1).

Our adult males apparently did not or could not adopt the strategy of using contemporary children's language as a guide to the same extent as did our females and, for that reason, found it more difficult to apply the confidence rating scale. The finding that the females had greater confidence than males for those words estimated to have been acquired earlier suggests further that confidence in their AA estimations was not consistent throughout the range of age estimations. Here, as elsewhere (cf. Carroll & White, 1973a), it is apparent that sex influences AA estimates. It is proposed here that the difference may be due to the strategy of making estimations.

REFERENCES

- CARROLL, J. B., & WHITE, M. N. Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 563-576. (a)
- CARROLL, J. B., & WHITE, M. N. Word frequency and age of acquisitions as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 1973, 25, 85-95. (b)
- COHEN, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.
- GILHOOLY, K. J., & HAY, D. Imagery, correctness, age-of-acquisition, familiarity, and meaningfulness values for 205 five-letter words having single-solution anagrams. *Behavior Research Methods & Instrumentation*, 1977, 9, 12-17.
- LACHMAN, R. Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, 1973, 99, 199-209.

LACHMAN, R., SCHAFFER, J. P., & HENRIKUS, D. Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical access time. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 613-625.

STRATTON, R. P., JACOBUS, K. A., & BRINLEY, B. Age-of-acquisition, imagery, familiarity, and meaningfulness norms

for 543 words. *Behavior Research Methods & Instrumentation*, 1975, 7, 1-6.

WINTERS, J. J., JR., & BRZOSKA, M. A. Development of lexicon in normal and retarded persons. *Psychological Reports*, 1975, 37, 391-402.

(Received for publication July 29, 1978.)