Young children's age-of-acquisition estimates for spoken words

AMANDA C. WALLEY
University of Alabama, Birmingham, Alabama

and

JAMIE L. METSALA
Ontario Institute for Studies in Education, Toronto, Ontario, Canada

This study was concerned with the impact of stimulus familiarity on young children's ability to recognize spoken words and make explicit judgments about them. In Experiment 1, 5-year-olds made age-of-acquisition (AOA) estimates for a set of words that were very similar to estimates made by older children and adults. In Experiment 2, young children's picture recognition, mispronunciation detection, and vocabulary monitoring performance all varied systematically with these AOA estimates and with a stimulus-type (intact-mispronounced) manipulation. Subjective AOA estimates (whether from children or from adults) proved to be a better predictor of performance than did two objective familiarity measures and subjective imageability. These results point to considerable metalexical knowledge on the part of young children or explicit sensitivity regarding their own vocabulary knowledge. In addition, the results lend some support to the notion that actual AOA contributes to subjective AOA estimates.

This study provides information about young children's sensitivity to the familiarity of spoken words. Previous studies have not systematically varied the familiarity of words presented to children for recognition (e.g., Cole & Perfetti, 1980; Elliott, Hammer, & Evan, 1987; Walley, 1987, 1988). This is surprising, since it is presumably the growth in knowledge of particular words that underlies the various familiarity (e.g., frequency) effects observed for adult recognition (see, e.g., Luce, 1986).

Adults' subjective age-of-acquisition estimates (estimates of when they first learned various words; henceforth, AOA estimates) are, like other subjective judgments, reliable and sometimes better predictors of word recognition and lexical access than are traditional, objective frequency measures (see, e.g., Brown & Watson, 1987; Carroll, 1971; Carroll & White, 1973b; Cirrin, 1983; Gilhooly & Gilhooly, 1979; Gordon, 1985; Lachman, Shaffer, & Hennrikus, 1974; Loftes & Suppes, 1972). Objective word frequency has typically been defined with reference to norms such as those of Kučera and Francis (1967), which are based on materials written for adults. Kolson (1961) provides a juvenile word frequency count that is derived from kindergartners' productions in the home and at school. The subjective judgments of adults are perhaps empirically better and conceptually more valid measures of lexical familiarity,

because such objective frequency counts may reflect certain sampling biases (Carroll, 1971; Gordon, 1985; Lachman et al., 1974) and underestimate the receptive vocabulary knowledge of children. Some investigators have collected picture labeling and vocabulary definition data from children as a means of validating adults' AOA estimates (Gilhooly & Gilhooly, 1980; Lyons, Teer, & Rubenstein, 1978; Winters, Winter, & Burger, 1978), but the developmental trend for the effect of AOA on verbal performance has not been a central concern.

One exception to the lack of evidence regarding the growth of word familiarity is a study by Cirrin (1984). He examined lexical decisions of subjects from several age groups for auditorily presented words that varied according to several subjective and objective frequency measures. In multiple regression analyses, adults' subjective AOA estimates were found to be the best predictor of kindergarteners', first graders', and adults' latencies for correct word responses (cf. Gilhooly & Logie, 1982). Another predictor of children's performance was the Kolson (1961) juvenile frequency count.

A second exception to the lack of developmental evidence regarding the growth of word familiarity is a study by Walley and Mestala (1990), who employed the listening-for-mispronunciations task developed by Cole (1973). In this task, it is assumed that to detect a mispronunciation, listeners first recognize the intended word from the acoustic phonetic input and any available higher level constraints. An error might also be detected from a more general failure to make sense of the input, but this is seldom the case, at least for adult listeners (Cole, Jakimik, & Cooper, 1978). In Walley and Metsala's (1990) study, 5- and 8-year-olds and adults were presented with mis-

pronounced and intact words from three adult-defined AOA categories in either constrained or unconstrained contexts, and they were asked to judge whether the words were "mispronounced" or "intact." The AOA categories were labeled *early*, *current*, and *late* to reflect their presumed status for the youngest subjects; 5-year-olds were expected to know the early words, to be in the process of learning the current words, but not to know the late words very well. Subjects' hit rates ("mispronounced" responses to mispronounced words) were generally high across AOA levels. False alarms ("mispronounced" responses to intact words) increased with AOA, especially those of younger children. That is, when presented with either *galaxy* (a late word) or *talaxy* (its mispronounced counterpart), children tended to respond "mispronounced," whereas errors for early words were primarily misses (e.g., incorrect responses of "intact" to *proceries*, not "mispronounced" to *groceries*). Thus, children's sensitivity to mispronunciations, as indexed by d' from signal detection analysis, was substantially influenced by AOA. These results obtained for both constrained and unconstrained contexts.

In the studies by Cirrin (1984) and by Walley and Metsala (1990), nonwords and mispronunciations were constructed in a similar manner. Specifically, one phonetic segment in familiar words or in words varying in AOA was changed to yield a phonologically permissible nonword. Minimally, children's lexical decision and mispronunciation detection responses to these items imply that the age at which one learns a word influences how readily it is recognized (see also Coltheart, Laxon, & Keating, 1988; Gilhooly, 1984). For example, words learned early may have more complete or robust structural representations in memory than words learned later (Brown & Watson, 1987; Carroll & White, 1973b; Gilhooly, 1984). Young children's performance might further be interpreted as indicating that they possess considerable metalexical knowledge. That is, children appear to have conscious access to information regarding the status of spoken items (in terms of their familiarity) within their own lexicons. They also seem to be most aware of the mismatch between the presented, mispronounced, and intended forms of more familiar words. Their performance in the listening-for-mispronunciations task might even be indicative of an ability to notice localized, segmental errors in words, and thus of phonological awareness. Yet, is it appropriate to attribute such metalinguistic knowledge to young children?

In support, lexical monitoring does have developmental primacy relative to other standards for the monitoring and evaluation of comprehension (see Baker, 1985). The spontaneous repairs that 2- and 3-year-olds make for their own speech also point to the existence of early metalexical, and perhaps even metaphonological, knowledge—that is, the ability to detect and fix errors of form (Clark, 1978). The toddler's preference for certain sounds in production is similarly suggestive (see, e.g., Menn, 1983). However, the extent of young children's lexical monitoring ability may depend heavily on task demands. Baker (1985) suggests that sometimes children fail to identify

nonsense words as "difficult to understand," because they assume that the items are words and are reluctant to admit their ignorance. She has found that even 5-year-olds can make effective use of a lexical standard in evaluating spoken passages with nonsense words, when asked to "find the mistakes," rather than to identify words that are "difficult to understand"—when *stimulus* as opposed to *perceiver* inadequacies are stressed. Still, their performance is nowhere near ceiling, and they often challenge the "wordness" of real words (i.e., make false alarms). Even third graders are less likely to report visually presented short, as opposed to long, nonsense words as posing potential comprehension difficulties, perhaps because they assume that the short "words" should be familiar (Baker, 1989). Thus, stimulus characteristics such as word length and perceived familiarity also influence children's lexical monitoring performance.

The present study provides further information about the extent of young children's metalexical knowledge. We asked children themselves to make AOA estimates for spoken words (Experiment 1). We then considered how their AOA estimates were related to performance in the listening-for-mispronunciations task, as well as in a vocabulary monitoring task and a picture recognition task (Experiment 2). We focused on the performance of 5-year-olds, whose age was about the same as that of the youngest subjects in previous related research. Also, 5-year-olds have not yet had extensive exposure to written language, so that this could not exert a major influence on their performance.

EXPERIMENT 1

Can children make the sort of explicit, quantitative judgments about lexical familiarity represented by the subjective AOA measure, and how might their estimates compare with adults'? If children's and adults' estimates prove to be systematically related, this would suggest early metalexical knowledge, because, minimally, children's basic spoken word recognition varies with adults' AOA estimates (Cirrin, 1984; Gilhooly & Gilhooly, 1980; Lyons et al., 1978; Walley & Metsala, 1990; Winters et al., 1978).

Why is there this relation between adults' AOA estimates and recognition performance? Adults' estimates might reflect the residence time of words in memory and thus cumulative or lifespan experienced frequency. However, this account, by which a multiplicative function of frequency and AOA should be a better predictor of performance than either variable alone, has not been supported in studies of picture naming and word completion in adults (Carroll & White, 1973b; Gilhooly & Gilhooly, 1979). Furthermore, AOA is a better predictor of word-naming speed than length of residence, when the two variables are unconfounded, and thus the age at which words are learned may be the critical factor mediating subjective estimates (Gilhooly, 1984).

The suggestion is not that adults possess literal, autobiographical information about their own language-learning histories (information about the actual age at which they

have acquired particular words), but that their estimates reflect more general lexical awareness that is, nevertheless, sensitive to the age at which words were acquired, perhaps because words learned early rather than later are somehow better consolidated in memory (Carroll & White, 1973b; Gilhooly, 1984). Brown and Watson (1987) make a similar claim in proposing that the phonological output representations of early as opposed to later learned words are more complete, because, as vocabulary acquisition proceeds, less storage capacity is available. Adults' estimates might additionally reflect explicit, pragmatic knowledge about the "age appropriateness" of various words or what words people of different ages are likely to know (see also Coltheart et al., 1988; Gilhooly & Watson, 1981). Determining whether or not young children can make AOA estimates and how their estimates compare with those of older children and adults should lend differential support to either the experienced frequency or the age-of-acquisition account of subjective AOA estimates.

Method

Subjects. Twenty preschoolers (mean age = 5 years, 0 months; range = 4,2 to 5,9) and 20 older children (mean age = 7,10; range = 6,7 to 9,5) were solicited through local day-care centers. All were native English speakers, with no speech or hearing disorder.

Materials. The 96 test words, which were those used by Walley and Metsala (1990), had been rated by 17 adults on a 9-point AOA scale (after Carroll & White, 1973a).

Procedure. All the test words were read in their intact forms to the children. After each word was presented, the children were asked to estimate the age at which they had first learned it, or, if they did not know the word, the age at which they thought they would learn it. They made their estimates by pointing to 1 of 9 pictures of a child, depicted as increasing in age in terms of size, body proportions, amount of hair, and so forth. The intended age of the child was displayed under each picture in numerical form, and the nature of the scale was explained to the children in a familiarization phase. The children's responses in testing were subsequently recoded according to the 9-point AOA scale used by Walley and Metsala (1990), on which a rating of 1 = acquired at age 2, 5 = acquired at age 6, 9 = acquired at age 13 or later. Half of the children at each age rated one half (16) of the words in each of the three AOA categories (early, current, and late) formed from adults' ratings; the other half rated the remaining 48 words, presented in random order. The children were tested individually in a 30-min session.

Results

The mean AOA ratings (and standard deviations) for each test word obtained from adults by Walley and Metsala (1990) and from the children in this experiment are shown in the Appendix. Also shown are two objective frequency measures, one based on the Kolson (1961) count for words spoken by kindergarteners in the home and school, the other on the Kučera and Francis (1967) count for materials written for adults. Frequencies were combined for singular, plural, and possessive forms, with a value of .5 assigned to words not in these counts (19%, 56%, and 56% of the early, current, and late words in the juvenile count; 6%, 12%, and 38% of the words in the adult count). Frequencies were converted to Standard Frequency Index (SFI) values (see Carroll, 1970).

The average AOA ratings of each age group were similar in absolute terms, particularly for words originally categorized as early. These are shown in Table 1, together with mean SFI Kolson and mean SFI Kučera and Francis (1967) values. A two-way analysis of variance (ANOVA) of items (words) for a mixed design revealed main effects of AOA and age [$F(2,93) = 225.74$; $F(2,186) = 29.72$] and a significant AOA × age interaction [$F(4,186) = 15.55$]. All reported effects were significant at or beyond the level of $\alpha = .05$. Post hoc (Newman-Keuls) comparisons indicated that for early words, 5-year-olds', 8-year-olds', and adults' AOA estimates did not differ significantly; for current words, young and older children's estimates were similar, but lower than adults'; for late words, older children's estimates were lower than young children's, and young children's estimates were, in turn, lower than adults'. All AOA comparisons were significant; that is, each age group rated early words as acquired before current words, and current words as acquired before late words. Thus, except for the reversal in young as opposed to older children's estimates for late words, AOA estimates increased systematically with age, and with Walley and Metsala's (1990) categorization of the words based on adults' estimates.

We also examined the extent of age-related similarities in subjects' AOA ratings for individual words. The Pearson's r correlations between (1) young children's and adults' ratings, (2) older children's and adults' ratings,

Table 1
Mean Age-of-Acquisition Ratings and Objective Frequency Values
(With Standard Deviations) for the Test Words

Source	Early Words		Current Words		Late Words	
	M	SD	M	SD	M	SD
5-year-olds	2.29	.08	3.81	.25	6.23	.19
8-year-olds	2.51	.10	3.74	.11	5.82	.23
Adults	2.22	.10	4.49	.11	7.00	.15
SFI Kolson	52.67	1.54	45.39	1.35	39.60	.31
SFI Kučera & Francis	49.85	1.39	49.18	1.28	43.49	1.14

Note—The early, current, and late categories were originally formed on the basis of adults' age-of-acquisition estimates (Walley & Metsala, 1990). SFI, Standard Frequency Index.

and (3) young and older children's ratings were .88, .90, and .91, respectively. These correlations remained high (r = .80, .85, and .86), even when those between a given age group's ratings and the Kolson SFI values for the words were partialed out. (For young children, older children, and adults, these correlations were −.61, −.60, and −.67, respectively; these were higher than those between subjective AOA and the Kučera and Francis SFI values.)

Discussion

The ability to make AOA estimates has not been examined before from a developmental perspective. In the present experiment, both 5- and 8-year-olds' estimates for a set of 96 words increased systematically with those of adults. Moreover, children's and adults' estimates were similar in absolute terms, especially those for words originally categorized as early in Walley and Metsala's (1990) study. By inference, these results point to the existence of early metalexical knowledge, since adults' AOA estimates are significant predictors of both adults' and children's performance in various recognition tasks; when adults say that they have learned a word earlier than another word, the "earlier" word is more likely to be recognized by children and adults.

The present results are also suggestive with regard to the basis for adults' AOA estimates. The similarity in children's and adults' AOA estimates for early words suggests that these estimates are not a simple reflection of cumulative or lifetime experienced frequency, since adults have necessarily encountered these words more often than children have. Instead, these estimates might reflect *actual* age of acquisition (Carroll & White, 1973b; Gilhooly & Gilhooly, 1979). However, when children (and adults) rate a word such as *cartoon* as having been acquired early (around age 3), they do not *remember* having learned this particular word at that age. Rather, because they learned and consolidated this word at an early age, subjects *perceive* it as being very familiar and therefore infer that they must have learned the word at an early age. When children rate a word such as *quartet* as one that is acquired late (around age 10), presumably they realize that they do not know the word and therefore conclude that they will learn it later or that it is a word that only older people know. Of course, the smaller differences in children's as opposed to adults' ratings for the three AOA categories are consistent with the notion that subjective AOA is influenced by the *relative* experienced frequency of words (this being less in the case of children).

There were also age-related differences in AOA estimates. Although all subjects judged late words as acquired after current words and current words as acquired after early words, both 5- and 8-year-olds judged current and late words as being acquired earlier than did adults. Perhaps children simply possess less extensive metalexical knowledge and thus overestimate their own lexical knowledge. However, their estimates for current and late words might, for example, additionally reflect more recent exposure to these words than adults' estimates do.

If so, their estimates might actually prove to be equally good, if not better, predictors of children's recognition performance. Some preliminary evidence that bears on this question is provided in Experiment 2.

EXPERIMENT 2

Two new groups of 5-year-olds were presented with intact and mispronounced versions of the early, current, and late words used in Experiment 1. Each group performed one of two primary tasks: a listening-for-mispronunciations task (i.e., they judged whether the words were "intact" or "mispronounced") or a vocabulary monitoring task (i.e., they made "know" or "don't know" judgments). A picture recognition task followed the primary one.

The main question was how children's and adults' AOA estimates would compare as predictors of performance in these tasks. Examination of the pattern of responses across tasks should also provide information about the extent of young children's metalexical knowledge. In particular, how will children respond in the vocabulary monitoring task to less familiar, current, and late words— words to which they often respond "mispronounced" in the listening-for-mispronunciations task, regardless of whether items are mispronounced or intact (Walley & Metsala, 1990)? A high proportion of "know" judgments (or random responding) would suggest that children are unaware of whether or not they actually know these words. In contrast, a low proportion of "know" judgments would support the notion that children's "mispronounced" responses to these words reflect awareness that the words are unfamiliar. Also, how will children respond to mispronounced as opposed to intact versions of the test items? Children might claim not to know *all* mispronounced words, regardless of AOA status. Alternatively, they might be better able to "see through" the distorted versions of early words and recognize the intended forms. The picture recognition task should provide an additional index of recognition for individual test items in the listening-for-mispronunciations task and a means of evaluating the accuracy of children's judgments in the vocabulary monitoring one.

Method

Subjects. Thirty-two preschoolers (mean age = 4 years, 8 months; range = 3,8 to 5,11) were solicited, as was described for Experiment 1. The data of 6 other children were not analyzed, because they almost never used the "mispronounced" or "know" response for any of the early, current, and late, mispronounced words.

Materials. The test words were those used in Experiment 1. In Walley and Metsala's (1990) study, these words were selected so that they comprised three categories—early, current, and late, the mean adult AOA ratings (and standard errors) for which were 2.22 (.10), 4.59 (.11), and 7.00 (.15), respectively. The 32 words in each category had been divided into two lists (A and B), which were balanced with respect to word length, the identity and location in a word of target segments, and the stress of the carrier syllable, as were the words across AOA categories (see the Appendix).

Two audiotapes were prepared by a practiced female speaker. On Tape 1, List A items were mispronounced, and List B items

were intact; the reverse was true on Tape 2. Mispronunciations were created by changing a stop or fricative/affricate to another similar segment to yield a phonologically legal nonword. The changes across AOA categories were perceptually equivalent, according to adult ratings (Singh, Woods, & Becker, 1972). Since changes were limited to one phonetic segment of polysyllabic words, mispronounced words generally had unique solutions. The test items were recorded in four blocks of 24 trials, each block having an equal number of mispronounced and intact early, current, and late words in random order at 3-sec intervals.

A set of four pictures accompanied each of the auditory items in the picture recognition task, which involved re-presentation of one block of 24 trials from the primary listening-for-mispronunciations or vocabulary monitoring task. One picture depicted the target word; the other three depicted foils (one from each AOA category). These foils were selected without replacement from the 72 items in the three nonrepeated blocks. The pictures in a set, which were glued in random positions on pieces of cardboard, were created from illustrations in children's books, and their suitability was independently confirmed by three adults.

Procedure. The children were tested individually in a 30-45 min session. For both primary tasks, the children were told that some of the words they were about to hear contained mistakes or "were said wrong." They were also told that some words were difficult, so that they might not know them. The children were asked to indicate which words were correct or mispronounced, or which words they knew or did not know, by pressing one of two buttons on a response box. Pressing one button illuminated a green, happy face and corresponded to an "intact" or "know" judgment; pressing the other button illuminated a red, sad face and corresponded to a "mispronounced" or "don't know" judgment.

The children were randomly assigned to one of four block presentation orders (counterbalanced across subjects) for Tape 1 or Tape 2 in either primary task. Each tape began with six pairs of practice words—two pairs from each AOA category, each pair consisting of an intact and then a mispronounced version of a word. After responding to each of the intact practice words in the vocabulary monitoring task, the children were asked to tell the experimenter something about the word. On the basis of this response, they were given appropriate feedback for their "know" or "don't know" responses, as well as for their subsequent responses to the mispronounced version of the word. The intended, intact version of mispronounced words was emphasized in this familiarization phase for both primary tasks.

After the primary task, the children completed the picture recognition task, in which the last 24 auditory stimuli were re-presented. On each trial, the child was asked to point to the picture of the word that the speaker said or was trying to say. The experimenter manually recorded the child's responses in testing for this task and the primary one. No specific feedback was given. Auditory stimuli were presented over matched and calibrated Sennheiser HD430 headphones at a comfortable listening level via a portable Uher tape recorder.

Results

Listening-for-mispronunciations task. Proportions of hits ("mispronounced" responses to mispronounced stimuli) and false alarms ("mispronounced" responses to intact stimuli) were calculated on the basis of each child's responses for items in each AOA category (see Figure 1, top panel, for group means). A 2 (stimulus type) × 3 (AOA) repeated measures ANOVA of "mispronounced" responses revealed main effects of stimulus type and AOA $[F(1,15) = 57.35; F(2,30) = 17.11]$ and a significant two-way interaction $[F(2,30) = 34.56]$. According to post hoc analyses, hits were higher for early than for current and

late words; false alarms increased for each AOA level, and all stimulus type comparisons were significant. There was also a significant negative correlation between the mean AOA ratings and the numbers of children who correctly responded "mispronounced" to the 96 test words across tapes (Pearson's $r = -.28$); the correlation between AOA and the numbers of children who correctly responded "intact" to the words was higher ($r = -.67; z = 3.56$). Thus, for individual test items, correct responses decreased with increasing AOA, and this relation was greater for intact stimuli, in agreement with the ANOVA.

Further analyses, based on signal detection theory, indicated that children were more likely to discriminate mispronounced from intact versions of early as opposed to current, and of current as opposed to late, words [mean $d' = 2.32, .99,$ and $.52; F(2,30) = 40.81$]. They were more biased to respond "mispronounced" to both current and late items as opposed to early words (mean $\beta = 1.16$ and .88 vs. 2.24; $F(2,30) = 12.78$). Nevertheless, the raw data indicate that children did, to some extent, distinguish between the mispronounced and intact forms of even late words.

We next examined the picture recognition results for children who performed the listening-for-mispronunciations task. Proportions of correct picture responses were calculated as a function of stimulus type and AOA (see Figure 1, bottom panel). A two-way ANOVA revealed main effects of stimulus type and AOA $[F(1,15) = 11.27; F(2,30) = 23.88]$. Correct pictures were chosen more often for intact than for mispronounced words ($M = .70$ vs. .54) and for early as opposed to current and current as opposed to late words (.81, .63, and .42). Correct responses in each cell of the stimulus design were greater than the level expected by chance (.25). Early, current, and late foils comprised 17%, 34%, and 49% of children's incorrect picture selections, respectively. The main effect of stimulus type and the lack of a two-way interaction indicate that children were more likely to select the correct picture for intact words, even late ones. This pattern of performance is consistent with the results of the listening-for-mispronunciations task.

Vocabulary monitoring task. Proportions of "know" responses for mispronounced and intact words in a given AOA category were calculated (see Figure 2, top panel). A two-way ANOVA revealed main effects of stimulus type and AOA $[F(1,15) = 65.22; F(2,30) = 98.67]$ and a significant stimulus × AOA interaction $[F(2,30) = 20.82]$. "Know" responses decreased with each increase in AOA for both mispronounced and intact words. Stimulus type comparisons were significant for early and current words, but not for late ones. As in the listening-for-mispronunciations task, there was a higher correlation between the mean AOA ratings for the 96 test words and the numbers of children who responded "know" for intact as opposed to mispronounced words across tapes ($r = -.77$ vs. $-.41; z = 3.97$).

The picture recognition results for children who first performed the vocabulary monitoring task are shown in Figure 2 (bottom panel). A two-way ANOVA again re-

LISTENING—FOR—MISPRONUNCIATIONS

O——O "MP"/MP (Hits)
●——● "MP"/I   (False Alarms)

PICTURE RECOGNITION TASK

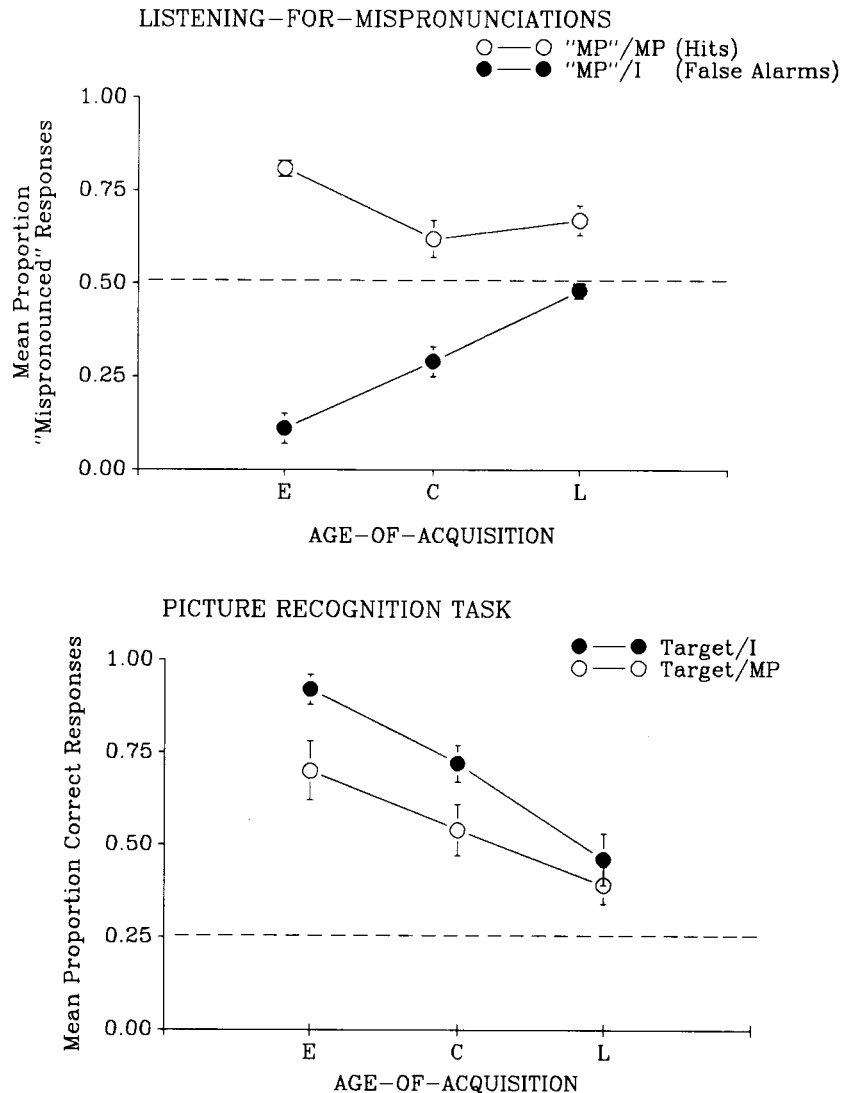●——● Target/I
O——O Target/MP

Figure 1. Mean proportions of "mispronounced" responses as a function of stimulus type and age of acquisition (AOA) in the listening-for-mispronunciations task (top panel). Hits and false alarms represent correct and incorrect "mispronounced" ("MP") responses, given a mispronounced (MP) or an intact (I) word. Mean proportions of correct picture (target) selections as a function of stimulus type and AOA in the picture recognition task (bottom panel). Vertical bars indicate standard errors of the means. The broken lines parallel to the x-axes represent chance performance.

vealed main effects of stimulus type and AOA [$F(1,15)$ = 24.20; $F(2,30)$ = 28.27]. Correct pictures were chosen more often for intact than for mispronounced words (M = .71 vs. .51) and for early and current than for late words (.80 and .68 vs. .37). Correct responses were all above chance (.25), except those for mispronounced late words. (Errors were primarily selections of late foils; early, current, and late foils comprised 16%, 36%, and 48% of children's errors, respectively.) Thus, although these children did not respond differentially to mispronounced and intact late words in the vocabulary monitoring task (they responded "don't know" to both), they were better able to recognize the intact items in the picture recognition task.

Children's and adults' AOA estimates as predictors of performance. Because 5-year-olds' and adults' AOA estimates were highly correlated with each other (r = .88), multicollinearity would likely be a problem in any regression analysis involving both variables as predictors; that is, they would not make independent contributions in predicting performance in the three tasks of the present experiment. Therefore, we examined these subjective measures separately as predictors of the children's performance in several analyses. Each analysis also included the Kolson (1961) SFI and Kučera and Francis (1967) SFI values described in Experiment 1, as well as an imageability measure, since there is evidence that this can in-

VOCABULARY MONITORING TASK
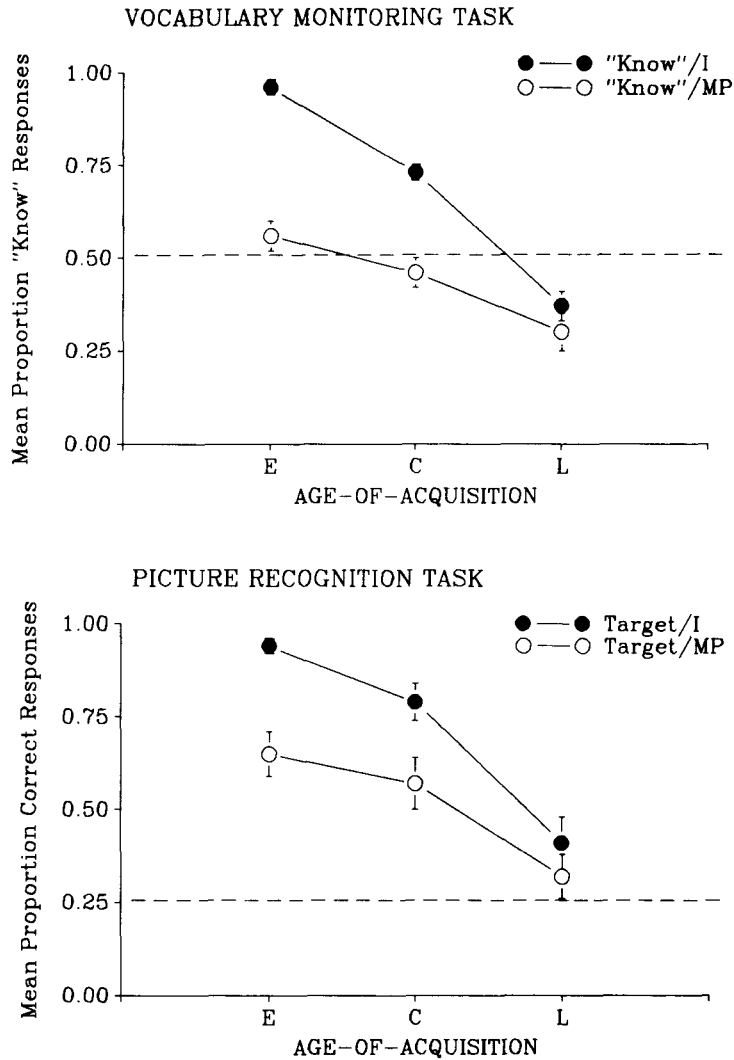

PICTURE RECOGNITION TASK

Figure 2. Mean proportions of "know" responses as a function of stimulus type and age of acquisition (AOA) in the vocabulary monitoring task (top panel). Mean proportions of correct picture (target) selections as a function of mispronounced (MP) and intact (I) stimuli and AOA in the picture recognition task (bottom panel). Vertical bars indicate standard errors of the means. The broken lines parallel to the x-axes represent chance performance.

fluence children's verbal performance (see Coltheart et al., 1988). Imageability ratings were obtained from 20 adults by using a 7-point scale (after Toglia & Battig, 1978). The mean ratings (and standard errors) for the early, current, and late words were 5.38 (.14), 4.77 (.16), and 4.40 (.16), respectively (higher ratings reflecting greater imageability). (Average ratings for each word are shown in the Appendix.) Table 2 shows the correlations between these potential predictor variables.

Forward, stepwise regression analyses were conducted, using the .05 level for inclusion of either children's or adult's AOA ratings, as well as imageability ratings and the two objective frequency measures. The criterion variables of interest were correct rejections and hits in the listening-for-mispronunciations task, "know" responses for intact and mispronounced words in the vocabulary

monitoring task, and correct picture selections for intact and mispronounced words in the picture recognition task (across the two subject groups). In each case, the number of children making a particular response to the 96 test

Table 2
Correlations Between Potential Predictor Variables in Experiment 2

	YAOA	AAOA	IMAG	KOL	KF
YAOA		.88*	-.46*	-.61*	-.33*
AAOA			-.47*	-.67*	-.35*
IMAG				.31*	-.15
KOL					-.47*

Note—YAOA, younger children's age-of-acquisition estimates; AAOA, adults' AOA estimates; IMAG, adults' imageability ratings; KOL, Kolson (1961) frequency (Standard Frequency Index); KF, Kučera and Francis (1967) frequency (Standard Frequency Index).   *p < .01, based on 94 df.

Table 3
Correlations Between Performance in the Listening-for-Mispronunciations, Vocabulary
Monitoring, and Picture Recognition Tasks and Predictor Variables, in Experiment 2

Predictor Variable	Listening-for-Mispronunciations		Vocabulary Monitoring		Picture Recognition	
	"I"/I (CR)	"MP"/ MP (Hit)	"K"/I	"K"/MP	T/I	T/MP
YAOA	−.68*	−.26	−.81*	−.49*	−.69*	−.55*
AAOA	−.63*	−.28*	−.78*	−.39*	−.62*	−.47*
IMAG	.23	.19	.32*	.38*	.33*	.38*
KOL	.55*	.36*	.63*	.35*	.39*	.27*
KF	.43*	.01	.41*	.18	.15	.07

Note—Shown for each task is the type of response, given either an intact or a mispronounced stimulus. "I"/I, an "intact" response, given an intact word (CR, or correct rejection); "MP"/MP, a "mispronounced" response, given a mispronounced stimulus (hit); "K," a "know" response; T, selection of the correct, target picture. YAOA, younger children's age-of-acquisition estimates; AAOA, adults' AOA estimates; IMAG, adults' imageability ratings; KOL, Kolson (1961) frequency (Standard Frequency Index); KF, Kučera and Francis (1967) frequency (Standard Frequency Index). *p < .01, based on 94 df.

words was examined. Table 3 shows the correlation between each predictor and criterion variable.

Table 4 shows the results from more detailed regression analyses, including the order in which predictor variables were entered into each analysis (which was based on the amount of variance individually shared with the criterion variable). The results were similar across criterion variables, regardless of whether children's or adults'

AOA estimates were included in a given analysis. Children's AOA estimates were significant predictors for five out of six of the criterion variables, as were adults' AOA estimates. On the average, children's estimates accounted for approximately 43% of the variance in performance, corresponding to an average multiple correlation coefficient of .66; on the average, adults' estimates accounted for about 35% of the variance in performance, correspond-

Table 4
Summary Results of the Regression Analyses in Experiment 2

Response Type	Predictor	β	Cumulative R^2	Change R^2	F
	Listening-for-Mispronunciations Task				
Correct Rejections ("I"/I)	(a) YAOA	−.64	.466	.466	62.41*
	KF	.06	.512	.046	8.88*
	(b) AAOA	−.54	.399	.399	44.49*
	KF	.06	.448	.049	8.35*
Hits ("MP"/MP)	(a) KOL	.06	.131	.131	14.21*
	(b) Same as (a); neither YAOA nor AAOA were significant.				
	Vocabulary Monitoring Task				
"K"/I	(a) YAOA	−.86	.664	.664	88.36*
	KOL	.06	.692	.028	8.64*
	(b) AAOA	−.76	.612	.612	59.44*
	KOL	.06	.634	.022	5.43*
"K"/MP	(a) YAOA	−.52	.244	.244	30.36*
	(b) AAOA	−.26	.150	.150	16.63*
	IMAG	−.57	.199	.049	5.68†
	Picture Recognition Task				
Target/I	(a) YAOA	−.45	.482	.482	87.61*
	(b) AAOA	−.37	.381	.381	57.76*
Target/MP	(a) YAOA	−.33	.303	.303	40.96*
	(b) AAOA	−.21	.221	.221	27.53*
	IMAG	−.27	.254	.033	4.05†

Note—"I"/I, an "intact" response, given an intact word; "MP"/MP, a "mispronounced" response, given a mispronounced stimulus (hit); "K," a "know" response. YAOA, young children's age-of-acquisition estimates; AAOA, adults' AOA estimates; IMAG, adults' imageability ratings; KOL, Kolson (1961) frequency (Standard Frequency Index); KF, Kučera and Francis (1967) frequency (Standard Frequency Index). (a) Models employing children's AOA estimates. (b) Models employing adults' AOA estimates. *p < .01. †p < .05.

ing to an average multiple correlation coefficient of .59. Objective frequency and imageability typically made much smaller contributions.

Discussion

Consistent effects of lexical familiarity, defined by adults' AOA estimates, and of the mispronunciation manipulation were observed for children's responses in the listening-for-mispronunciations, vocabulary monitoring, and picture recognition tasks (see also Cirrin, 1984; Walley & Metsala, 1990). Specifically, the more familiar a word, the more likely it was to be correctly judged as intact or mispronounced, the more likely it was to be judged as "known," and the more likely its pictured referent was to be selected correctly. Also, children were more likely to respond correctly to intact as opposed to mispronounced items in the listening-for-mispronunciations task, more likely to judge these items as "known," and more likely to select the correct picture referent. These effects were attenuated for the least familiar, late words. Five-year-olds' AOA estimates were slightly better than adults' in predicting these results. Together with the pattern of performance across tasks, this attests to considerable metalexical knowledge on the part of young children.

Consider performance across the three tasks in this experiment. Correct and thus differential responding to intact and mispronounced stimuli in the listening-for-mispronunciations task was most apparent for early items. For the same mispronounced stimuli, children were also most likely to select the correct picture referent, other children to judge these items as "known." This suggests that children are most able to "see through" the distorted forms of these items (i.e., to note the error and the intended word at the same time). The fact that children's "mispronounced" responses for both current and late mispronounced words (hits) were lower than those for early mispronounced words also suggests that children do not respond "mispronounced" to words simply because they are unfamiliar. Rather, they are most likely to detect mispronunciations of early words as such. They do evidence a bias to respond "mispronounced" to less familiar words, presumably because this seems most appropriate in the context of the task, where the focus is really on identifying errors. However, children seem to realize that these words are unfamiliar. This interpretation is supported by their selection of late foils in the picture recognition task, when they made an error, as well as by their responses in the vocabulary monitoring task.

There do appear to be differences in the abilities tapped by the three tasks. With the contextual support afforded by the picture recognition task, children were better able to recognize intact as opposed to mispronounced late words—words that presumably are not yet fully or robustly represented in their lexicons. Differential responding to intact as opposed to mispronounced late words was also observed in the listening-for-mispronunciations task, but the children did not distinguish between these items in the vocabulary monitoring one. In this sense, the former task

may serve as a more sensitive measure of metalexical knowledge than the latter, and also as a fairly accurate index of basic word recognition ability.

The children's vocabulary monitoring judgments were, nevertheless, generally in agreement with picture recognition performance and thus accurate. Their lexical monitoring ability may not always be apparent, because they are reluctant to admit that they do not know words (Baker, 1985). However, in our vocabulary monitoring task, the children might have been willing to do so, because they were told that the words varied in familiarity (which they did) and that some contained errors (which they did). As we have seen, children can even make quantitative judgments regarding lexical knowledge and familiarity that resemble those of adults. Their AOA estimates were, like adults', a significant predictor of performance in the three tasks of this experiment. The one exception consisted of predicting hits or "mispronounced" responses to mispronounced words. Presumably this was due to a lack of variation in these data; as noted above, the children tended to respond "mispronounced" for early items, because they recognized the error, and they tended to respond "mispronounced" to a somewhat lesser extent for current and late items, because they often perceived the latter as unfamiliar. (Responding "intact" to these items was, therefore, likely deemed task-inappropriate.) Perhaps the extent of the children's ability to make accurate AOA estimates is attributable to the fact that they did not have to make an all-or-none dichotomous judgment, as in the vocabulary monitoring task.

GENERAL DISCUSSION

Little attention has been given to children's recognition or more explicit knowledge of spoken words varying in familiarity. In Experiment 1, 5-year-olds' estimates of the age at which they had learned or would learn various words were very similar to estimates previously obtained from adults. In Experiment 2, children's performance in the listening-for-mispronunciations, vocabulary monitoring, and picture recognition tasks varied systematically with lexical familiarity, as defined by adults' AOA estimates, replicating and extending the results of previous research (Cirrin, 1984; Walley & Metsala, 1990). The finding that children's own AOA estimates were on a par with adults' as predictors of performance in these three tasks is perhaps not very surprising, given the results of Experiment 1. However, this does not mitigate its importance.

The present results help to document the origins of familiarity effects on spoken-word recognition and metalexical knowledge. At least by age 5, when children say that they have learned one word earlier than another, they are more likely to distinguish mispronounced and intact versions of the word, more likely to claim to know the word, and more likely to correctly select a picture of the word. Apparently this is because some words are actually acquired or entered into the lexicon earlier than others, rather than being experienced more often or having

resided in the lexicon longer (see also Carroll & White, 1973b; Gilhooly, 1984; Gilhooly & Gilhooly, 1979). This follows from the fact that both children and adults give earlier AOA estimates for some words, such as *balloon*, than they do for others, and that children are indeed more likely to recognize these words (see also Gilhooly & Gilhooly, 1980; Lyons et al., 1978; Winters et al., 1978). However, more extensive examination of children's AOA estimates for larger samples of words from each AOA category, and of their relation to performance, is needed. In particular, would a similar relation hold for the late category alone? Investigation of children's AOA estimates for words in situations in which episodic frequency is manipulated and of their relation to recognition performance might be informative too.

Our results also have practical implications. Specifically, they indicate that children's own AOA estimates could serve as the basis for selecting materials for developmental investigations and as an alternative to norm-based, objective measures and to time-consuming, a priori empirical determinations of lexical familiarity. The use of either children's or adults' subjective AOA estimates for this purpose might be advantageous when, for example, the test words of interest are from a relatively restricted frequency range, according to objective measures, as were the words used in the present study (see Walley & Metsala, 1990; see also Gordon, 1985), or when normative information about the words is not available. The Kolson (1961) norms, for example, are now quite old and may not include some words that are known by children today and that researchers might wish to employ for various reasons.

In conclusion, we have presented evidence indicating that young children can make AOA estimates that resemble those of more advanced language users. In comparison with other, objective measures of word familiarity and one other subjective measure (imageability), these AOA estimates were superior in predicting children's picture recognition, mispronunciation detection, and vocabulary monitoring performance. This finding, together with the pattern of performance observed across these three tasks, suggests that by about 5 years of age, children possess considerable metalexical knowledge or explicit knowledge about which words are familiar and unfamiliar. The observation of such knowledge may be enhanced in tasks in which the child is asked to focus on stimulus characteristics, rather than on inadequacies in his or her own knowledge base.

REFERENCES

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 155-205). New York: Academic Press.

Baker, L. (1989). Developmental change in readers' responses to unknown words. *Journal of Reading Behavior, 21*, 241-260.

Brown, G. D. A., & Watson, F. L. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition, 15*, 208-216.

Carroll, J. B. (1970). An alternative to Juilland's usage coefficient for lexical frequencies, and a proposal for a Standard Frequency Index (SFI). *Computer Studies in Humanities & Verbal Behavior, 3*, 61-65.

Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning & Verbal Behavior, 10*, 722-729.

Carroll, J. B., & White, M. (1973a). Age of acquisition norms for 220 picturable nouns. *Journal of Verbal Learning & Verbal Behavior, 12*, 563-576.

Carroll, J. B., & White, M. (1973b). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*, 85-95.

Cirrin, F. M. (1983). Lexical access in children and adults. *Developmental Psychology, 19*, 452-460.

Cirrin, F. M. (1984). Lexical search speed in children and adults. *Journal of Experimental Child Psychology, 37*, 158-175.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's conception of language* (pp. 17-43). New York: Springer-Verlag.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics, 13*, 153-156.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64*, 44-56.

Cole, R. A., & Perfetti, C. A. (1980). Listening for mispronunciations in a children's story: The use of context by children and adults. *Journal of Verbal Learning & Verbal Behavior, 19*, 297-315.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*, 1-12.

Elliott, L. L., Hammer, M. A., & Evan, K. E. (1987). Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults. *Perception & Psychophysics, 42*, 150-157.

Gilhooly, K. J. (1984). Word age-of-acquisition and residence time in lexical memory as factors in word naming. *Current Psychological Research & Reviews, 3*, 24-31.

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition, 7*, 214-223.

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

Gilhooly, K. J., & Logie, R. H. (1982). Word age-of-acquisition and lexical decision making. *Acta Psychologica, 50*, 21-34.

Gilhooly, K. J., & Watson, F. L. (1981). Word age-of-acquisition effects: A review. *Current Psychological Reviews, 1*, 269-286.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory & Language, 24*, 631-645.

Kolson, C. (1961). *The vocabularies of kindergarten children*. Pittsburgh, PA: University of Pittsburgh Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lachman, R., Schaffer, J., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of Verbal Learning & Verbal Behavior, 23*, 613-625.

Loftes, E., & Suppes, P. (1972). Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning & Verbal Behavior, 11*, 770-777.

Luce, P. (1986). Neighborhoods of words in the mental lexicon. In *Research on speech perception* (Tech. Rep. No. 6). Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.

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

Menn, L. (1983). Development of articulatory, phonetic and phonological capabilities. In B. Butterworth (Ed.), *Language production* (Vol. 2, pp. 3-50). London: Academic Press.

SINGH, S., WOODS, D. R., & BECKER, G. M. (1972). Perceptual structure of 22 prevocalic English consonants. *Journal of the Acoustical Society of America*, 52, 1698-1713.

TOGLIA, M. P., & BATTIG, W. R. (1978). *Handbook of semantic norms.* New York: Erlbaum.

WALLEY, A. C. (1987). Young children's detections of word-initial and -final mispronunciations in constrained and unconstrained contexts. *Cognitive Development*, 2, 145-167.

WALLEY, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development*, 3, 137-165.

WALLEY, A. C., & METSALA, J. L. (1990). The growth of lexical constraints on spoken word recognition. *Perception & Psychophysics*, 47, 267-280.

WINTERS, J. J., JR., WINTER, L., & BURGER, A. L. (1978). Confidence in age-of-acquisition estimates and its relationship to children's labeling performance. *Bulletin of the Psychonomic Society*, 12, 361-364.

APPENDIX
The 96 Test Words in Experiments 1 and 2

Word	OS–MP	Category	AAOA		OAOA		YAOA		IMAG		KOL	KF
			M	SD	M	SD	M	SD	M	SD		
antenna	t–b	*C	5.23	2.05	3.60	1.26	4.90	2.73	5.40	1.98	39.29	51.46
anvil	v–č	L	8.29	0.77	7.50	1.58	7.00	1.41	3.85	2.56	39.29	40.00
balloon	b–t	E	1.29	0.59	2.00	1.05	2.90	0.57	6.05	1.47	53.10	51.14
banana	b–k	E	1.47	0.87	1.80	1.03	2.00	0.94	6.40	1.14	53.10	46.99
bandana	b–t	*L	7.00	1.22	6.30	1.83	6.60	2.54	6.00	2.73	39.29	36.99
banjo	ǰ–š	C	4.76	2.28	4.10	1.37	5.80	2.53	5.40	2.01	39.29	43.01
barbeque	b–k	C	4.18	1.07	3.40	0.70	2.20	0.63	5.60	3.23	52.30	52.04
bathtub	b–k	*E	2.05	1.08	2.00	1.05	2.10	1.10	6.20	1.32	56.78	46.99
battery	b–g	*C	5.00	2.06	4.00	1.25	3.20	1.87	5.10	2.22	39.29	53.22
beaver	v–š	*C	3.64	1.69	3.90	1.45	2.50	0.85	4.80	1.96	39.29	44.77
bikini	b–t	*C	4.12	1.87	3.90	1.37	5.50	2.76	6.15	1.04	39.29	40.00
blackboard	b–g	*C	3.52	1.01	3.60	1.90	2.50	0.97	5.85	1.31	51.85	43.01
breakfast	f–s	*E	1.65	1.11	2.20	1.40	2.50	0.71	5.20	2.02	59.21	57.40
brother	ð–č	E	1.65	0.49	2.70	1.57	2.60	0.97	4.85	2.08	63.80	61.00
cactus	t–b	L	6.42	2.03	3.80	1.32	4.00	2.75	6.10	1.52	39.29	36.99
café	f–ž	L	7.71	1.16	5.40	1.51	6.60	2.01	3.95	1.47	39.29	53.98
carnival	k–d	C	4.82	2.65	4.80	1.32	4.20	2.78	4.85	1.56	39.29	49.54
carpenter	p–d	*C	5.71	1.40	4.30	1.57	5.20	2.57	4.05	1.93	54.06	51.14
cartoon	t–g	*E	1.94	0.94	2.30	1.16	2.20	1.03	4.95	1.67	45.31	49.54
cassette	s–ǰ	*L	6.18	2.04	4.50	0.97	5.70	2.58	5.90	1.48	39.29	36.99
cellophane	s–v	*L	6.88	1.36	6.70	1.83	7.40	1.71	3.55	2.09	39.29	40.00
cereal	s–ǰ	E	2.64	1.17	2.30	0.82	2.30	1.06	4.75	1.94	51.33	53.22
chimney	č–v	C	3.94	1.09	3.40	1.26	2.70	0.48	5.90	1.29	55.92	50.00
circle	s–č	E	2.06	0.83	1.80	1.03	1.70	1.06	6.85	0.49	58.43	59.64
cocoon	k–d	*C	5.18	2.13	3.70	1.89	3.50	1.84	4.60	1.54	39.29	47.44
commercial	k–b	C	3.94	2.22	3.00	1.41	2.40	1.84	4.20	2.12	39.29	58.51
crayon	k–b	E	1.91	0.83	1.60	0.97	1.90	0.74	5.70	1.95	64.48	40.00
croquet	k–d	*L	6.71	2.02	6.40	1.43	6.50	2.59	3.05	1.79	39.29	36.99
cubicle	b–k	L	7.00	1.32	7.50	0.97	7.30	1.57	4.40	1.70	39.29	36.99
dessert	z–f	E	1.94	0.83	2.20	1.23	1.80	0.79	4.65	1.72	51.85	49.54
detective	t–b	C	5.29	1.76	3.60	1.27	5.30	3.06	3.90	1.89	50.08	58.57
diploma	p–g	L	7.71	0.85	6.20	2.44	7.50	1.78	5.60	1.67	39.29	36.99
dynamite	d–k	*L	5.76	1.71	4.90	1.29	5.70	2.91	5.55	1.79	49.29	46.99
factory	f–v	*C	5.23	2.02	3.90	1.60	4.00	2.36	3.95	1.82	52.30	57.48
family	f–v	*E	2.82	1.74	2.90	1.45	2.30	0.68	5.20	1.94	56.28	66.07
fingerprint	f–ǰ	C	4.82	1.94	3.30	1.57	2.10	1.20	6.10	0.97	50.76	48.45
flashlight	f–g	C	3.94	1.82	3.20	1.23	2.40	1.08	6.00	1.86	53.76	49.03
flower	f–s	*E	1.24	0.44	2.30	0.95	2.10	0.99	6.05	0.94	64.45	53.80
fortress	f–g	L	5.94	1.03	6.20	2.20	5.80	2.86	4.30	1.78	39.29	49.03
galaxy	g–t	L	7.47	0.80	4.80	1.23	5.40	2.68	4.75	1.77	39.29	50.00
giraffe	ǰ–s	*E	2.47	1.23	3.00	1.49	2.00	1.56	6.45	1.00	51.34	36.99
groceries	g–p	*E	2.82	1.74	2.70	1.25	1.90	0.99	4.80	1.40	51.34	50.41
gymnastics	ǰ–f	L	6.00	1.17	3.10	1.37	3.80	2.57	4.55	1.76	39.29	51.76
hamburger	b–t	E	2.64	1.54	2.30	1.06	2.00	1.05	5.65	1.78	39.29	50.00
hotel	t–b	C	4.18	1.07	3.50	1.84	4.00	3.02	4.95	1.88	42.30	61.67
Indians	d–p	*C	3.71	1.61	3.40	1.71	3.50	1.08	5.30	1.89	57.99	59.54
kayak	k–d	*L	7.23	1.44	7.10	1.45	6.10	3.04	5.20	1.54	39.29	36.99
lumber	b–k	*C	4.94	2.08	4.90	1.97	5.50	2.99	4.55	1.79	54.06	55.44
machine	š–v	*C	4.29	2.17	2.90	1.29	2.60	1.58	3.60	1.90	62.56	61.96
magician	ǰ–f	*C	4.29	1.26	3.20	1.32	2.70	1.49	4.75	1.71	39.29	48.45

APPENDIX (Continued)

Word	OS–MP	Category	AAOA		OAOA		YAOA		IMAG		KOL	KF
			M	SD	M	SD	M	SD	M	SD		
mechanic	k–b	*L	6.35	1.17	4.30	1.42	5.40	2.72	3.95	1.96	39.29	47.78
medicine	t–k	*E	2.47	1.12	3.10	1.53	2.60	0.97	3.90	2.22	45.31	55.44
musician	z–f	C	5.47	1.23	4.50	1.65	4.50	2.42	4.55	1.88	39.29	58.13
octopus	t–s	C	4.05	1.60	4.00	2.06	3.20	1.62	6.35	1.14	39.29	40.00
pajamas	ĵ–f	E	1.94	0.97	2.10	1.10	1.90	0.99	4.65	2.06	55.73	46.02
palette	p–d	L	7.58	1.37	7.30	0.82	7.30	1.49	3.20	2.07	39.29	46.99
parcel	p–g	C	4.82	1.74	4.00	1.49	3.30	1.95	3.30	2.03	39.29	43.01
pavilion	p–g	*L	7.17	1.13	7.30	0.68	7.70	1.06	4.05	1.79	39.29	47.78
penguin	g–t	C	4.35	2.15	3.10	1.52	2.20	1.55	6.15	1.35	39.29	36.99
perfume	f–š	C	4.35	2.06	3.70	1.57	2.20	1.55	4.15	1.93	57.08	50.41
photograph	t–p	E	2.82	1.85	4.50	1.58	4.10	2.64	5.25	2.10	42.30	55.31
piano	p–g	*E	3.00	1.54	3.10	1.37	2.50	0.53	6.60	0.68	62.04	55.91
pizza	p–g	*C	3.53	1.54	1.80	0.79	2.20	1.03	6.20	1.32	50.09	44.77
podium	d–k	*L	8.00	0.94	7.20	1.03	6.90	1.60	5.20	1.82	39.29	40.00
policeman	p–d	*E	2.17	0.81	2.70	1.49	2.40	0.84	5.75	1.71	53.10	55.44
potato	t–b	*E	2.00	0.94	2.20	1.23	2.10	0.88	5.95	1.70	62.35	54.77
prescription	p–d	L	7.52	1.18	5.60	0.97	4.80	2.25	4.20	2.14	39.29	48.45
propeller	p–d	*C	4.35	0.86	3.90	1.66	5.60	3.24	4.75	1.86	39.29	43.01
puppet	p–d	*E	2.00	1.11	2.60	0.97	2.10	0.88	4.60	1.57	50.75	50.79
quartet	t–b	*L	7.64	1.40	6.50	1.43	7.40	1.51	3.85	1.81	39.29	49.54
rabbit	b–t	E	1.82	0.88	2.10	1.29	1.80	0.92	5.80	1.79	39.29	52.04
raincoat	k–d	E	2.65	1.05	3.00	1.16	2.40	0.84	5.45	1.79	57.87	43.01
reflector	f–g	*L	7.47	0.72	5.00	1.33	6.90	2.02	3.95	1.70	39.29	49.03
reporter	p–d	L	6.94	0.75	4.50	1.58	5.90	2.51	4.20	1.58	39.29	57.32
safari	f–č	L	6.11	1.90	4.70	0.95	6.40	2.22	4.35	1.81	39.29	43.01
saltshaker	š–v	*E	3.47	1.62	3.30	1.42	2.80	1.87	5.05	2.16	39.29	36.99
sandbox	s–ĵ	E	2.47	1.28	2.80	1.23	2.40	0.84	5.00	1.62	39.29	36.99
sardine	s–v	*C	5.00	2.02	3.80	1.32	5.10	2.77	4.50	2.16	39.29	43.01
scalpel	p–g	*L	8.24	0.75	7.10	0.99	7.20	1.75	4.95	1.88	39.29	36.99
scorpion	p–d	*L	6.25	0.58	5.40	1.35	6.50	2.51	5.40	1.57	39.29	36.99
sculpture	č–ž	*L	6.64	1.58	4.40	1.84	4.30	2.83	5.00	1.38	39.29	52.55
spaghetti	g–t	E	2.23	0.97	2.00	1.05	2.40	1.08	5.95	1.67	56.45	40.00
strawberries	b–k	*E	2.29	0.92	2.10	1.45	2.00	1.05	6.45	1.10	59.79	43.01
sunshine	š–v	*E	2.53	0.94	2.40	1.27	2.20	1.23	5.85	1.31	54.61	49.03
surgeon	s–g	*L	5.94	1.34	6.20	0.92	6.40	2.27	5.05	1.88	39.29	50.79
syringe	s–v	*L	8.29	0.69	7.50	1.08	7.30	2.11	5.65	1.76	39.29	40.00
table	t–p	*E	1.12	0.30	2.60	1.08	2.30	0.95	5.45	2.19	68.44	63.86
tattoo	t–b	L	6.18	1.51	4.40	2.07	5.30	3.27	5.25	1.48	39.29	36.99
tobacco	t–g	C	5.47	1.81	5.30	2.11	7.40	1.58	4.15	2.03	39.29	52.79
toboggan	t–g	E	2.88	1.05	3.10	1.79	3.00	2.45	5.35	1.81	39.29	36.99
trapeze	p–d	L	6.18	1.92	4.90	2.02	5.20	2.66	4.95	2.01	39.29	36.99
triangle	t–g	E	2.71	1.36	2.50	1.65	2.10	0.88	6.70	0.73	39.29	46.99
trombone	t–b	C	5.12	1.87	4.20	1.48	5.00	2.79	5.00	2.05	39.29	36.99
tuxedo	t–g	L	6.53	1.70	5.60	1.71	5.90	2.69	6.25	0.91	39.29	36.99
valet	v–θ	L	8.82	0.39	7.80	0.79	7.10	1.97	3.00	1.75	39.29	43.01
vampire	v–š	*C	5.64	1.66	3.70	0.82	4.50	2.55	5.85	1.27	39.29	40.00

Note—OS, original segment (also underlined); MP, mispronunciation; E, C, and L, early, current, and late categories based on adults' AOA (AAOA) ratings (Walley & Metsala, 1990). List A items are marked with an asterisk. OAOA and YAOA, older and younger children's AOA ratings; IMAG, adults' imageability ratings; KOL and KF, Kolson (1961) and Kučera and Francis (1967) SFI values.