A Lexical Model of Learning to Read Single Words Aloud

Roman Taraban and Carolyn Beth Taraban
Department of Psychology
Texas Tech University
Lubbock, TX 79409-2051
tirmt@ttacs.ttu.edu

Abstract

Three principles governing the operation of the lexical pathway in a model of reading single words aloud were applied to the question of learning, as measured by times to initiate correct pronunciations. I. At the lexical level, a target word activates a ***neighborhood*** of orthographically similar entries in the lexicon. II. At the phoneme level, the correct phonemes in the phonemic spelling of the word compete with the other active phonemes. III. At the naming level, the pronunciation is composed of a conjunction of phonemes. These principles were tested using the data from a 4-year-old beginning reader (LT), resulting in a goodness-of-fit R^2 = .44. When a rule pathway using grapheme-phoneme correspondences was added to the lexical pathway, the goodness-of-fit was comparable (R^2 = .46). When single entries were accessed along the lexical pathway, instead of word neighborhoods, and grapheme-phoneme correspondences were accessed along the rule pathway, as in standard dual-route models, the goodness-of-fit R^2 fell to .27. Although the model-fitting supported the importance of neighborhood activation and failed to support the importance of rules, grapheme-phoneme correspondences were overtly used by LT in the initial trials with words and when feedback indicated an errorful pronunciation. Thus, rule application may be relatively slow in normal fluent word naming, but may still play a strategic role in attempts to initially decode letter strings or to correct errors.

Two related questions are central to an understanding of reading single words aloud. One concerns what gets activated in a mental lexicon, which is a person's store of knowledge about words. Another is whether a lexical pathway is sufficient for the task. Dual-route theorists (e.g. Baron & Strawson, 1976; Coltheart, 1978; Forster & Chambers, 1973) proposed that a person can "look up" the pronunciation of a familiar word along a lexical route. However, because most people have no difficulty reading pseudowords, like *gok*, for which they have no lexical entry, it seemed that individuals also had a store of rules available, which they accessed and applied along a rule route. These rules relate graphemes (the letter or letters that spell phonemes) to phonemes (the simplest units of spoken sound). The suggestion that rule knowledge is part of the mechanism for pronunciation is plausible, particularly for beginning readers. Children are often taught spelling-sound correspondences in reading programs and children and adults use this knowledge readily (Coltheart & Leahy, 1992; Siegler, 1988). An awareness of phonemes and an understanding that letters map into sound are also strong predictors of reading development (see Rayner & Pollatsek, 1989, for a summary).

Others have disputed the need for a rule route, arguing that a lexical route is sufficient. On this account, a target word does not simply access a single entry in the lexicon, as in standard dual-route models. Rather, the target word activates "neighborhoods" of words in the lexicon, based on their orthographic overlap (shared letters) with the input; the activated entries are then synthesized into a pronunciation (e.g. Glushko, 1979; McClelland & Rumelhart, 1981; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987). Single-route lexical models have been successful in explaining many of the empirical results that dual-route models can explain, like the difficulty in pronouncing exception words (words that do not follow the typical spelling-sound correspondences in English, like *have*) compared to regular words (words that do conform to the typical correspondences, like *fish*), as well as results that dual-route models cannot easily explain, like the difficulty in pronouncing regular words with exception neighbors (so-called regular-inconsistent words, like *gave*).

The two questions posed at the outset were addressed using mathematical analyses of pronunciation latencies that were collected while teaching a 4-year-old to read her first set of words. Three models, depicted in Figure 1, were tested: a *lexical model* in which a target word activated a word neighborhood in the lexicon; the *standard dual-route model*, in which the target word was the only word accessed on the lexical route and in which grapheme-phoneme correspondence rules were accessed on a rule route; and a *modified lexical model*, which added the rule route from the standard dual-route model to the lexical model. The lexical model was motivated by findings with adult subjects reported in Taraban and McClelland (1987). Although Taraban and McClelland rejected the idea that a rule route was necessary, as stipulated by dual-route models, it was not clear that this was the case throughout development, particularly during early reading instruction. The 4-year-old subject in this experiment knew the grapheme-phoneme correspondences for regular consonant sounds prior to learning any words. Thus, in this case, there were strong reasons to hypothesize that further rule learning would occur along with new lexical learning, which could be neighbor learning (the modified lexical model) or single whole word learning (the standard dual-route model). The lexical model is described in detail next. The equations for testing all the models are provided in the Appendix.

The Lexical Model

Before describing the experiment, it may be useful to focus on the cognitive components of the lexical model depicted in Figure 1 and described mathematically in Figure 2. First, consider the equation for the activation of exemplars (i.e. entries) at the **lexical level**. Given a target word L, the activation of each exemplar (a_e) in the lexicon is a multiplicative function of its match to the input. Given weight strengths between letter detectors and exemplars in the range $.5 < w < 1.0$, with .5 representing a neutral value, a letter match in position i contributes w_i to the activation of the exemplar, and a letter mismatch contributes $1 - w_i$ (See Taraban & Palacios, 1993, for details of this activation function). Activation of the i th phoneme (p_i) in word L is represented at the **phoneme level** as the sum of the activations of all exemplars whose pronunciations match the correct phoneme in the i th position, divided by the sum of the activations of all the exemplars (cf. Luce, 1959; McClelland, 1991; McClelland & Rumelhart, 1981; Medin & Schaffer, 1978; Taraban & Palacios, 1993).[1] The composition of the pronunication at the **naming level** is conjunctive, that is, it consists of

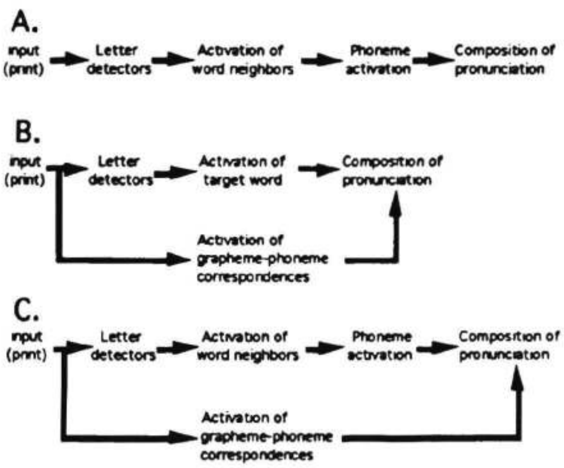

Figure 1. A depiction of (A) the lexical model, (B) the standard dual-route model, and (C) the modified lexical model.

[1] The specific set in the denominator of the expression is an artifact of our learning set (described later). That is, every exemplar has an i th letter, thus the full set of alternatives is represented by the activations of all the exemplars (the set E); if this were not the case, then the denominator would be a subset of the activations of all the exemplars.

p_1 AND p_2 AND p_3 (for three phonemes) which is implemented through the multiplicative function. Later, we verify that the appropriate operator is AND, not OR, or some intermediate operator.

Naming Level

$$n_L = \prod_{i=1,k} p_i$$

(where n is the activation of the name for word L, and k is the number of phonemes in the phonemic spelling)

Phoneme Level

$$p_i = \frac{\sum_{e \in E_i} a_e}{\sum_{e \in E} a_e} \qquad i = 1, 2, \ldots k$$

(where p_i is the activation of the correct phoneme in position i, E_i is the set of exemplars whose pronunciation matches the correct phoneme in the i-th position, k is the number of phonemes in the phonemic spelling)

Lexical Level

$$a_e = \prod_{i=1,n} \omega_{e,i} \qquad e \in E$$

$$\omega_{e,i} = \begin{cases} w_i, & \text{if } l_i \text{ matches the i - th letter of e;} \\ 1 - w_i, & \text{otherwise} \end{cases}$$

(where a_e is the activation of exemplar e, E is the set of exemplars, word L = l_1, l_2, l_3, and n is the number of letters in word L)

Figure 2. The computation of pronunciations in the lexical model.

The Experiment

The subject (LT) was the 4-year 5-month old daughter of the experimenters. At the beginning of the expeirment, LT could visually identify the letters of the alphabet and the typical sound that each consonant made. The rules relating letters to sounds constituted a major part of what LT knew about reading. The learning set consisted of 36 three-letter words[2] that were presented on a computer screen and that were read

[2] The words were: bag, bib, bud, bug, bun, but, den, did, dig, dim, dip, gag, god, got, gum, leg, lip, lot, mad, man, map, mat, men, mug, nag, nap, not, pad, pen, pig, pot, tag, tap, ten, top, tub. These words were chosen for a priming manipulation that is not discussed here. LT was familiar with the meaning of the words.

aloud for speed and accuracy. A single-subject design was appropriate because it allowed for precise control over the frequency of presentation of specific words and for a test of the various models when all the lexical entries in visual memory were known.

The data were collected over 54 meetings. The newness of the task and the age of the subject did not allow the immediate collection of pronunciation latencies using a computer. Rather, for the first 12 meetings, words were presented on index cards, printed in one-inch capital letters. Error data were collected but latencies were not. Throughout the experiment, LT was given explicit feedback after each trial about the accuracy of her response. During the first few meetings, LT would sometimes ask about the sound associated with a letter in a word, and was told what the sound was. Within six meetings, she always generated a pronunciation before receiving any feedback, which from that point forward was simply "Right" or "Wrong, try again" (unless the error was reading the word backwards, in which case she was told explicitly that this was the error). Whenever an error was made, LT usually got the pronunciation right on the second attempt. Beginning with the seventh meeting, LT was encouraged not to vocalize before naming the word (This was done to prepare her for the shift to the timed computer paradigm). At the computer, trials were initiated by LT by pressing a switch interfaced with the computer after a fixation mark appeared on the screen. The switch closure started a timer accurate to 1 millisecond. LT pronounced the word into a microphone that was also interfaced with the computer, which stopped the timer. If LT named the word incorrectly, the experimenter verbally indicated that the pronunciation was incorrect and LT tried again. There was a computer-controlled delay of 5 seconds after the correct pronunciation during which the word remained on the screen. The screen was then replaced with a fixation mark.

In order to avoid empty cells for any of the items in the statistical analyses, the reading times and error probabilities for sets of six consecutive meetings were pooled item-by-item and the means for each item were computed. These 252 means (36 items X 7 pooled meetings) were also used for model fitting. An analysis of variance using pooled meeting as the factor for the seven pooled meetings at the computer showed a significant effect in the analysis of error probabilities (F (6, 210) = 4.46, p < .001) and in the analysis of times to initiate correct pronunciations (F (6, 210) = 39.13, p < .001). The mean times in milliseconds and the error probabilities (including the off-line sessions *pre1* and *pre2*) are summarized in Figure 3. These results show a consistent decrease in times to initiate pronunciation and a reduction in errors, with practice. The most interesting anecdotal data from this experiment involved LT's attempts in the earliest trials to first identify the sound of each letter in the word and then to repeat the combination of sounds quickly over and over until she recognized a word. In a related fashion, if she made an error at the computer, she would sometimes appear to guess but would usually analyze the sounds, letter by letter, before trying another pronunciation. These observations suggested a strong role for the rule pathway. The fits of the data to alternative models, presented next, provided a test of this possibility.

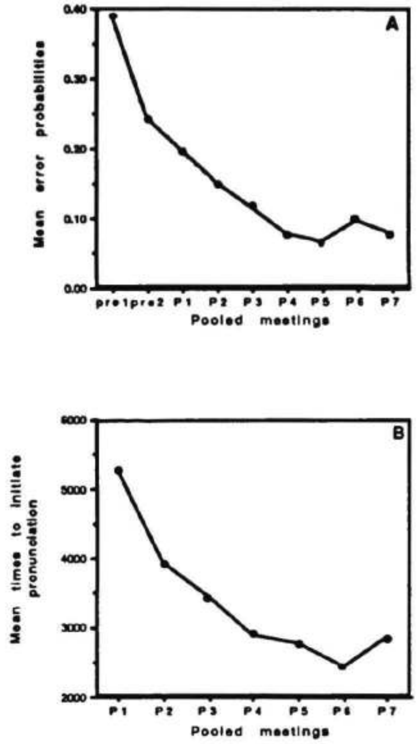

Figure 3. (A) Mean error probabilites; (B) Mean times (ms) to initiate correct pronunciations.

The Analyses

The purpose of the analyses was to assess the roles of the lexicon and rules in learning pronunciations. The data were the times to initiate correct pronunciations.[3] For each of the three models that was fit, learning was interpreted as a change in associative strengths of lexical and rule knowledge. These weight changes were expressed as the natural log of the pooled meeting, as shown in Eq 1. At pooled meeting m,

$$w_{i,m} = c_i + r_i \ln(m), \qquad m = 1, 2, \ldots 7 \qquad (1)$$

[3] Only correct pronunciations were modeled because it was not clear how to relate times for errors to the learning expressed in Eq 1 and to the models in the Appendix. More detailed process models could model all the latencies. Error probabilities could have been modeled but we chose not to because for any given trial "correct" (1) and "incorrect" (0) are crude measures of performance.

where c_i is a constant, r_i is a learning rate parameter, and w_i represents the associative strength of a letter to an exemplar (lexical models), a whole word to a lexical entry (standard dual-route model), or a letter to a phoneme (standard dual-route model).

First, consider learning in the *lexical model.* For a particular target word, like *pot*, the lexical representation for *pot* would have a relatively high activation level, and close neighbors (that share two letters) like *got* and *lot* would be relatively more active than distant neighbors, like *but* and *pad*, or non-neighbors, like *dim*. This is because close neighbors match the target in more letter positions than distant neighbors (See Figure 2). In the lexical model, the associative strengths of these word neighborhoods changed with practice. In the *standard dual-route model*, in contrast, a single lexical entry was strengthened for any particular target, as well as the grapheme-phoneme correspondence rules for the letters in the target (see the Appendix for details). The *modified lexical model* combined the neighborhood learning of the lexical model and the rule learning of the standard dual-route model.[4]

In order to test the conjunctive nature of pronunciation at the naming level in the lexical model, we additionally tested a version that incorporated a tuneable negation operator (Neg) (Oden, 1992). The negation operator is based on De Morgan's Law:

$$\text{A AND B} = \text{NOT (NOT A OR NOT B)} \qquad (2)$$

In the range of v_j values tested, which was -1.0 to 1.0, the tuneable operator acts like an OR (at -1.0) or AND (at 1.0) at the boundary values and like a fuzzy operator in between (see Oden, 1992, for details). The operator was implemented using Eq 3 for the p_i values from Figure 2, with negation beginning with p_1 and applying from left to right, requiring six v_j parameter estimates.

$$\text{Neg}(p_i) = p_i^{vj} / (p_i^{vj} + (1-p_i)^{vj}) \qquad (3)$$

In order to assess how well the models accounted for the data, a goodness-of-fit R^2 was calculated for each of the models, as $R^2 = 1$ - *Residual SS / Corrected SS*, which indicates how much better the model fit the data compared to simply using the mean. The *standard dual-route model* (R^2 = .27) did considerably worse than the *lexical model* (R^2 = .44), supporting the validity of neighborhood activation and learning, as well as the remaining components in Figure 2. The tuneable negation operator improved the fit only marginally (R^2 = .45), which supported the use of conjunctive combination at the naming level rather than an operation between logical AND and OR. Finally, the fit of the *modified lexical model* (R^2 = .46) was comparable to the fit of the lexical model.

Two findings argued against the importance of rules: the poor fit of the dual-route model compared to the lexical model and the comparable fit of the modified lexical model compared to the lexical model. If rules were of any consequence, there should have been more of a discrepancy in the fits for the lexical and modified lexical models.

A comparison of the lexical model with neighborhood activation to the dual-route model without neigborhood activation argues for the importance of neighbors. If LT were simply learning about the target word, then the R^2 for the standard dual-route model should have been comparable to, or perhaps even better than, the R^2 for the lexical model. The mathematical expression for the dual-route model (See the Appendix) presumed single whole word learning by assigning a single weight to the lexical pathway, on the assumption that on a given trial only one lexical representation was strengthened and that from meeting to meeting each lexical representation was strengthened by roughly the same amount. A related question is whether the lexical model simply fit a global learning curve better than the dual-route model. If this were the case, then the actual times would be randomly distributed among the *slower* and the *faster* times that the lexical model predicted. To examine this possibility, the predicted times to initiate pronunciations were divided into slower and faster times, for each pooled meeting, according to a median split. An examination of Table 1 shows that the model consistently predicted the actual slower and the faster times. This is important, because it indicates that the times themselves are based on more than a strengthening of unrelated lexical items, as would be the case if the lexical model simply fit a global learning curve.

Table 1. Predicted (Pred) and actual means (LT), in milliseconds, for slower and faster pronunciations, based on a median split of predicted times to initiate pronunciation.

	Slower Words		Faster Words	
Pooled Meeting	Pred	LT	Pred	LT
1	5472	5578	4411	4767
2	4661	4350	3478	3472
3	4144	3872	2917	2917
4	3783	3311	2517	2450
5	3489	3250	2178	2267
6	3250	2833	1917	2011
7	3039	3294	1683	2389

[4] This model is similar to a version of a dual-route model suggested in Coltheart, Curtis, Atkins, and Haller (1993), which incorporates a lexical pathway that accesses neighborhoods of words, not just the target word.

Overall, the analyses provided support for the three principles embodied in the levels in Figure 2. **I**. At the lexical level, a target word activates a "neighborhood" of orthographically similar entries in the lexicon. **II**. At the phoneme level, the correct phonemes in the phonemic spelling of the word compete with the other active phonemes. **III**. At the naming level, the pronunciation is composed of a conjunction of phonemes. These three principles help to explain essential aspects of the learning that took place. In order to provide a sense of the overall goodness-of-fit for the reading times, the predicted times from the lexical model are plotted against the actual times in Figure 4 for the 252 fitted data points.

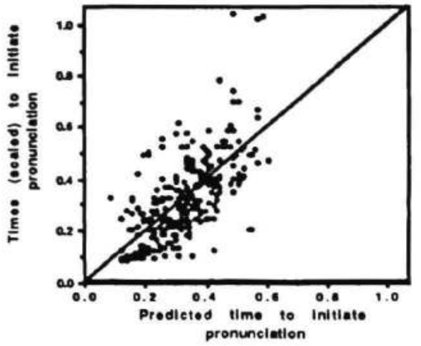

Figure 4. Times predicted by the lexical model versus actual times to initiate pronunciations. The actual times were multiplied by 10^{-4} prior to model fitting.

Discussion

LT's knowledge of letter-sound rules at the outset of the experiment and the anecdotal data from the experiment suggested that there might be an important role for rules in the cognitive mechanism that she relied on for reading single words aloud. LT overtly used letter-sound rules in the initial trials and when she made an error. However, the fits of the models to the reading time data for correct pronunciations indicated that letter-sound rules were not important. How can we resolve this apparent paradox?

It would be incorrect to conclude from the success of the lexical model in fitting LT's data that the prerequisite skills commonly associated with starting to read were unimportant, particularly some specific knowledge of how letters map into sound. This knowledge allows a child to confront the task of decoding letter strings into speech with some hope of success. We question the suggestion, however, that rules play a central role in generating pronunciations for words after a person gains some fluency in reading. From this it would follow that rules are not strengthened as part of the weight strengthening (error correction) process in fluent reading but lexical neighborhoods are. Rudimentary spelling-sound mappings could still provide the means of bootstrapping the creation of lexical representations that are strengthened through practice.

Our observations suggested that LT's knowledge of letter-sound correspondences supported an overt back-up strategy for correcting errorful pronunciations. Her reliance on these rules was not surprising, as these correspondences constituted part of what she had learned about written codes prior to the experiment. On this interpretation, rules can be used strategically when novel letter strings are confronted or when the person is aware of an error in pronunciation. A promising way of thinking about this restricted use is in terms of a backup strategy that can be applied independently when retrieval along the lexical pathway either fails or is intuited as likely to fail. The strategy choice paradigm in Siegler (1988) should be helpful in further exploring this possibility.

Appendix

The equations for the specific models (Eqs 4, 5, 6) were developed by reasoning as follows. A particular cognitive component, like the lexical route in Figure 1, modulated times to initiate pronunciation. Because the equation for any particular component quantified the fit of the component's output to the correct pronunciation, the contributions of the (scaled) cognitive components were *subtracted* from a baseline value (the β_0's or β_1's). Basically, the better the fit, the faster the response time. In all cases, learning was quantified as a change in associative strengths (w), as specified in Eq 1.

Equation 4 was used to fit reading times to the lexical model:

$$RT_L = \beta_0 - \beta_D n_L \qquad (4)$$

where RT_L is the predicted time to initiate pronunciation for word L, β_0 is a baseline parameter, β_D is the scaling parameter for the word name activations, and n_L is defined in Figure 2. The second model tested was the standard dual-route model, in which one component consisted of changes to the associative strengths of whole words and another component consisted of changes in the associative strengths of grapheme-phoneme correspondence rules. In Eq 5, the first additive factor represents the lexical lookup route. Because LT saw all the words with equal frequency, the strengthening of all the lexical representations was assumed to be roughly the same, as reflected in the w_D parameter; the second factor represents the rule route:

$$RT_L = (\beta_0 - \beta_D w_D) + (\beta_1 - \beta_R \Pi_{i=1,k} f_{L,i} w_i) \qquad (5)$$

where $f_{L,i}$ is the frequency of the grapheme-phoneme rule for the correct phoneme in the i-th position in the phonemic spelling of word L (see Coltheart et al., 1993), k is the number of phonemes in the phonemic spelling, and w_i is the associative strength of rules in phoneme position i. Values for $f_{L,i}$ are constants that were computed directly from an examination of the learning set. The third model combined the lexical model and the rule component from the standard dual-route model:

$$RT_L = \beta_0 - \Pi_{i=1,k} (\beta_D p_{L,i} + \beta_R f_{L,i} w_i) \quad (6)$$

where $p_{L,i}$ is the i-th phoneme in the phonemic spelling of word L, and represents the lexical route, as defined in Figure 2, and $f_{L,i}$ w_i represents the rule route as defined in Eq 5.

The best-fitting parameters for the three models were found using an iterative search algorithm. The algorithm is available in the non-linear regression procedures in SPSS 4.0; also see Gill, Murray, Saunders, & Wright, 1984.

Acknowledgements

We would like to thank Gary Fireman, Philip Marshall, and three anonymous reviewers for comments on an earlier version of this paper.

References

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 386-393.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). San Diego, CA: Academic Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review, 100*, 589-608.

Coltheart, V., & Leahy, J. (1992). Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 718-729.

Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Gill, P., Murray, E., Saunders, M., & Wright, M. (1984). Procedures for optimization problems with a mixture of bounds and general linear constraints. *ACM Transactions on Mathematical Software*, *10*, 282- 296.

Glushko, R. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 674-691.

Luce, R. (1959). *Individual choice behavior*. New York: Wiley.

McClelland, J. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23*, 1-44.

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Nosofsky, R. (1992). Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (pp. 149-167). Hillsdale, NJ: Erlbaum.

Oden, G. (1992). Direct, incremental learning of fuzzy propositions. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 48-53). Hillsdale, NJ: Erlbaum.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Seidenberg, M., Waters, G., Barnes, M., & Tanenhaus, M. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23*, 383-404.

Siegler, R. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59*, 833-851.

Taraban, R., & McClelland, J. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, *26*, 608-631.

Taraban, R., & Palacios, J. (1993). Exemplar models and weighted cue models in category learning. In G. Nakamura, R. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation (Vol. 29): Categorization by humans and machines* (pp. 91-127). San Diego, CA: Academic Press.