CHAPTER

4

ACCURACY LEVEL, A_L

If it is agreed that reading achievement can be replaced by the more precise construct of efficiency level, E_L, as was advocated in the preceding chapter, then it becomes important to analyze the primary factors that influence E_L. In the causal model, accuracy level is one of the two factors at Echelon 2 that are causal for E_L—along with rate level which will be described in detail in the next chapter.

As was noted earlier in Chapter 1, the theoretical construct called A_L is very similar to the traditional concept of reading level. Rauding accuracy level, or A_L, is simply an upgraded version of the reading level concept. Therefore, A_L is a very traditional concept with a new name, a more precise theoretical definition, and a new technique for measuring it, as will be explained later.

High accuracy level, A_L is purported to be a necessary, but not sufficient condition for high reading achievement. Individuals cannot become very good readers unless they have a high accuracy level, A_L. However, it is possible to have high A_L but still not have high reading achievement (high E_L) because rate level, R_L, is low.

This chapter contains a detailed definition of the A_L construct, explains how it relates to similar traditional concepts, explains how it can be measured, and gives a brief summary of empirical evidence supporting its existence.

Theoretical Construct

Accuracy level, A_L, is the highest level of text difficulty in GE units, that individuals can accurately read, when they are reading at their rauding rate, R_r. The operational definition of what it means to read text accurately has conventionally been taken to be 75% (e.g., see Betts, 1946), but this cutoff criterion has been operationally refined on the basis of empirical data to be 64% (Carver, 1990b). For example, suppose individuals with a rauding rate of 150 Wpm (R_L = 4.5; R_r = 9 standard sentences per minute; see Appendix D) are given passages to read at varying levels of difficulty. If their accuracy of text comprehension in this situation, A_r, is .80 for a fifth-grade level passage but is only .60 for a sixth-grade passage, then their accuracy level is at the fifth-grade level (e.g., A_L = 5.5), because the fifth-grade level of passage difficulty is the most

difficult level that these individuals can accurately read (A > .64) at their own rauding rate.

Remember that the words in texts that are relatively easy, $A_L > D_L$, have been practiced so often that they are recognized at the rauding rate of the individual. For example, if an individual has an accuracy level at grade four (A_L = 4), then it is assumed that at least 99% of the words in text at $D_L = 3$ and below are accurately recognized and successfully processed at the rauding rate (see Carver, 1994b). All of these words that an individual can accurately process at the rauding rate will be called "raudamatized words." Therefore, all of the words in texts at $D_L < A_L$ should be raudamatized words; those words will be symbolized as A_LWords. This means that a measure of A_L and a measure of A_LWords should correlate almost perfectly because the A_L construct and the A_LWords construct are inherently very closely related, such that the number of raudamatized words, A_LWords, is an indicant of A_L.

Related Traditional Concepts

Introduction. In this section, the concepts of reading level, automaticity, autonomous lexicon, and unitized words, will be related to the A_L construct.

Reading Level. The concept of reading level was explicated by Betts (1946) when he described the independent, instructional, frustration, and capacity levels. The lowest level of Betts' was the independent level—also referred to as the basal level. For example, if an individual had an independent level at second grade, then this student should be able to read material at the second-grade level of difficulty, such as a book, without help from a teacher, parent, or anyone else and achieve at least 90% comprehension with 99% accurate pronunciation of the words in the text. The next higher level was the instructional level, which should be higher than the independent level; it is the teaching level, and is the level that is most similar to accuracy level, A_L. For example, if an individual has a third-grade instructional level, or $A_L = 3$, then this individual should be taught using materials at the third-grade level because they are challenging but not too difficult. The individual is supposed to achieve around 75% comprehension when reading at the instructional level, with 95% accuracy of pronunciation. The frustration level should be higher than the instructional level; it is to be avoided. For example, if an individual has a fourth-grade frustration level, then this is the lowest level of difficulty at which this individual is unable to understand. For this individual, material at the fourth-grade level is too difficult and frustrating because less than 50% will be comprehended. The highest level of all should be the capacity level, or hearing level. It is the highest level of material difficulty that a student can understand when listening to someone else read the material aloud. For example, indi-

viduals who have a fifth-grade capacity level, can comprehend 75% of the material from a book at fifth-grade level when someone reads it aloud to them.

Informal reading inventories are often based upon the above criteria of Betts. These inventories often ignore rate measurement, and none of the criteria presented by Betts involved rate. For example, Fuchs, Fuchs, and Deno (1982) conducted a study of the reliability and validity of curriculum-based informal reading inventories (IRIs), and none of the seven criteria they investigated for determining reading level involved a rate requirement. Another example comes from Cadenhead (1987), who traced the history of the concept of reading level in his effort to eliminate the concept, and rate was not involved. Although IRIs theoretically do not involve rate, the measurement of a student's reading level by a classroom teacher may involve rate in a manner that makes the result more similar to efficiency level, E_L. Furthermore, IRIs traditionally involve oral reading with all the measurement problems that accompany this practice (see Carver, 1990a). Therefore, when reading level is measured by use of an IRI, it may reflect efficiency level, E_L, more than accuracy level, A_L, because it may involve both accuracy and rate via oral reading rate and accuracy of word pronunciation. Nevertheless, the traditional concept of reading level ordinarily does not denote or connote a rate dimension. So, the concept of instructional reading level can be used to help communicate the meaning of A_L, in spite of the fact that some measures of reading level using IRIs may correlate higher with E_L than A_L, and some may correlate relatively low with both because these informal reading measures are often unreliable.

In summary, Betts' instructional reading level is conceptually similar to rauding accuracy level, A_L; however, Betts' uses a 75% comprehension criterion, and rauding theory uses a 64% comprehension criterion because empirical data suggests that 64% of passages will be comprehended when the accuracy level matches or equals the difficulty level ($A_L = D_L$), or the relative difficulty is zero ($D_L - A_L = 0$). From one standpoint, rauding theory has provided a rather precise definition of reading level which can be operationally applied, that is, the traditional concept of instructional reading level means rauding accuracy level. The construct A_L is an upgraded version of the reading level construct because it specifies that individuals must be reading at their own rauding rate—normal or typical silent rate—when determining the most difficult level of text that can be read accurately.

Automaticity. Raudamatized words, as defined earlier, would probably be considered "automatized" words by many researchers. However, the concept of automaticity that has been advanced by LaBerge and Samuels (1974) is associated with a general ability that has been learned by the individual rather than a property of a particular word with respect to a particular individual. Furthermore, their concept of automaticity is defined directly in terms of a "lack of attention," whereas raudamatized words are defined in terms of rate, not attention.

The criterion that LaBerge and Samuels used for deciding if a skill had reached automaticity was whether "it can complete its processing while attention is directed elsewhere" (p. 295). They thought that automaticity was a general skill that an individual learned which could be applied to all the words an individual read. That is, they contended that "... during the development of automaticity the person may either attempt to reorganize smaller units (e.g., words) into larger units (e.g., word phrases) or he may simply stay at the word unit level" (p. 316). Thus, for LaBerge and Samuels each word did not become automatized for the individual so that each individual had some words that were automatized and some that were not. Instead, each individual had varying degrees of automaticity for all words and those who do not develop automaticity do not succeed in becoming "fluent" readers. They carried this idea of automaticity for larger and larger units (letters to words to phrases) even further by using it to explain fast reading, even speed reading, as the following quotations attest:

> For example, when the child reads text in which the same vocabulary is used over and over again, the repetitions will certainly make more automatic the perceptions of each word unit, but if he stays at the word level he will not realize his potential reading speed. If, however, he begins to organize some of the words into short groups or phrases as he reads, then further repetitions can strengthen those units as well as word units. In this way he can break through the upper limit of word-by-word reading and apply the benefits of further repetitions to automatization of larger chunks. (p. 315)
>
> We do not know specifically how to train a child to organize codes into higher units although some speed-reading methods make claims that sheer pressure for speed forces the person out of the word-by-word reading into larger units. (p. 315)

It seems clear from the above quotes that LaBerge and Samuels were offering a theory as to why some individuals are able to comprehend accurately while reading fast (are "fluent" readers) while others are slow (and not "fluent"). Supposedly, the fast readers have learned to chunk the information contained in text into larger units (from letters to words to phrases) by practicing this process whereas the slower readers have not learned to do this, probably because they never practiced this process by forcing themselves to process larger and larger chunks. This idea that the reader needs to achieve a phase or stage of automatic reading has recently been articulated by Stahl (1997). He said that "the transition from accurate to automatic word recognition occurs over a number of years, conventionally from the end of first grade to the end of third grade" (p. 17). Notice that this idea seems to suggest that typical readers by the fourth grade can rapidly read all words, seemingly without practice on them, because they have become automatic readers. Furthermore, this idea seems to suggest that typical readers in second and third grade are not able to

recognize any words automatically even with hundreds or thousands of practice trials on such words as "the," "and," "you," etc.

In summary, the general idea of automaticity has been upgraded in the causal model by the A_L construct and raudamatized words, because the original concept of automaticity does not seem to be valid.

Autonomous Lexicon. Fries (1963) seems to be the first to talk about the "automatic recognition" of words. He says that "as great skill develops they become automatic habits and the recognition response itself sinks below the threshold of attention" (p. 177). He has elaborated upon this automatic word recognition as follows:

> The major spelling patterns of present-day English are fortunately few in number, but for those the reader must develop, through long practice, high-speed recognition responses. These responses must become so habitual that practically all the clues that stimulate them eventually sink below the threshold of attention leaving only the cumulative comprehension of meaning. (p. xvi)

Perfetti (1991a) has defined the autonomous lexicon as consisting of all the printed words an individual can recognize automatically, or quickly. Therefore, conceptually there is no difference between the autonomous lexicon and A_LWords as defined earlier. That is, the lexicon of raudamatized words and the autonomous lexicon are synonymous in concept. The only justification for creating a new term, "raudamatized" to replace "autonomous" or "automatized" words, is that these latter words carry the unwanted connotations of limited attention capacity and a general skills affecting all words. Translated into rauding theory and the causal model, all the words in relatively easy material, $A_L > D_L$, should be in an individual's autonomous lexicon because they should be accurately and quickly recognized, and comprehended, at the individual's rauding rate, R_r. Any treatment that successfully increases A_L should also produce a corresponding, or proportional, increase in the size of the autonomous lexicon, or A_LWords, and any treatment that successfully increases the size of the autonomous lexicon, or A_LWords, should produce a corresponding, or proportional, increase in A_L.

Given the above connections among A_LWords, A_L, and the autonomous lexicon, it will be helpful to note what Stanovich (1980) said about automaticity and reading that is also relevant to raudamatized words, or A_LWords. He has articulated the relationship between automaticity and reading as follows:

> ... during the time that automaticity is developing, and even after a word is fully automatized, recognition time continues to decrease. The latter point is often lost in discussions that center on the automaticity concept itself, even though there is ample evidence in the

> literature documenting the fact that recognition time continues to decrease after words have become fully automatized. (p. 60)

In the above research referred to by Stanovich, the criterion for reaching automaticity was the point where attention was no longer directed toward recognition of the word but could be directed elsewhere, such as to comprehending the complete thought in the sentence. Notice that these ideas of Stanovich about automaticity make it clear that for him, automaticity refers to a particular word for a particular individual. These ideas articulated by Stanovich may be readily translated into the following: An automatized word is not necessarily a raudamatized word because further practice on an automatized word may be needed to get it up to the rauding rate; however, a raudamatized word is necessarily an automatized word because a raudamatized word has been practiced to a rate well beyond the point where it can be recognized without attention. Raudamatized words are likely to be recognized at a faster rate than the slower rate associated with automatized words.

In summary, the size of the autonomous lexicon as conceived by Perfetti (1991a) should be equivalent in concept to A_LWords; therefore, measures of the size of the autonomous lexicon should also correlate highly with measures of the A_L construct.

Unitized Words. In 1983, Ehri and Wilce further clarified the connection between automaticity and reading by testing their theory (Ehri & Wilce, 1979b) that skilled word recognition develops in the following three phases:

> In Phase I, unfamiliar words become familiar and are recognized *accurately* by readers directing their attention to component letters as they map sounds. During Phase 2, as a result of more practice, familiar words come to be recognized *automatically* as wholes without attention and without deliberate processing of component letter-sound relations. In Phase 3, the *speed* of processing familiar words increases to a maximum as the components involved in stimulus recognition and response production become consolidated or "unitized" in memory. (p. 3)

Notice that Ehri and Wilce make the same kind of distinction that Stanovich (1980) made; an automatized word is processed at a slower rate than a "unitized" word. The concept of a unitized word used by Ehri and Wilce is very similar to the concept of a raudamatized word, so these two concepts need further elaboration.

There is a major problem with using the word "unitized" to describe the point at which the speed of word recognition no longer increases due to practice or experience. Ehri and Wilce had the idea that each unitized word could be recognized at exactly the same rate or speed. Yet, this is not possible during rauding. Carver (1976) has shown that when reading rate is measured using actual words per minute, then shorter words are read faster; however, when rate

is measured in standard length words per minute, then rate is constant at the rauding rate, R_r. This means that shorter words are processed faster than longer words during rauding. Therefore, the concept of a unitized word is misleading because there is not likely to be a constant amount of time needed per actual word. Instead, as discussed in Chapter 2, there is likely to be a certain amount of time needed per standard length word. So, words that have reached asymptote on a learning curve will not be described as "unitized," mainly because that would erroneously suggest a unitary rate of word recognition for a particular individual, no matter how long the word was in letters, phonemes, or syllables. Yet, it should not go unnoticed that the earlier concept of a unitized word advanced by Ehri and Wilce is very similar to the raudamatized concept introduced earlier in this chapter.

In summary, all of the A_LWords and all of the words in the autonomous lexicon should be raudamatized words that Ehri and Wilce (1983) would call unitized words, except all raudamatized words are only recognized at the same rate, or a unitary rate, if rate is measured in standard length words.

Relevant Tests

Introduction. There are two tests that would seem to provide highly valid measures of A_L, and they will be described in some detail. Then, a few other measures of A_L will be briefly acknowledged.

Degrees of Reading Power (DRP). The DRP test includes passages that increase in difficulty level along with questions which must be read and answered without a time limit. Therefore, this test provides an indicant of accuracy level because the nature of the test follows closely the definition of accuracy level. This test does not provide GE scores that have been scaled to reflect A_L; so, it provides an indicant of A_L instead of an indirect measure. However, it should be noted that Carver (1990c) has rescaled the scores on this test into GE units so these scores may provide a valid direct measure of A_L, though their validity for providing absolute levels of A_L has not been investigated.

The DRP, described above, measures reading ability in a manner very similar to what would be expected if it had been explicitly designed to measure the A_L construct. That is, it requires items to be answered covering passages of increasing difficulty, and the examinees are given the following instructions:

> You are not expected to read at the same rate as other people or to answer the same number of questions. As you work on this test you will find that the passages become harder to read... . You will be given as much time as you need.

Accuracy Level Test (ALT). The ALT is a reading vocabulary measure of A_LWords; therefore, it is an indicant of A_L. It has been used previously to investigate rauding theory; it is the measure noted in the causal model, from Chapter 1. This test was originally developed in 1987, with 100 vocabulary items spanning the range of 0.0 to 15.8 GE units. It is also purported to be an indirect measure of accuracy level, A_L, because it involves a measure of A_L-Words scaled in GE units. It was later revised (Carver, 1994a) by including 20 more items at the top of the test; now, it measures across 120 items which cover the range between 0.0 and 19.0 GE units.

The idea that vocabulary tests are indicants of reading level, or A_L, has been articulated (a) by Sternberg (1987) who said that "... one's level of vocabulary is highly predictive of one's level of reading comprehension" (p. 90), and (b) by Anderson and Freebody (1981) who said that "an assessment of the number of meanings a reader knows enables a remarkably accurate prediction of this individual's ability to comprehend discourse" (p. 77). The more printed words known by an individual, the more likely the individual will be able to comprehend more difficult texts that contain less frequently used words. Furthermore, individuals who can comprehend text at higher levels of difficulty are more likely to know the meaning of more words that occur less frequently. Therefore, a measure of reading vocabulary should constitute an indicant of A_L.

The ALT is somewhat different from traditional reading vocabulary tests which ordinarily include alternative wrong answers (foils) that are somewhat plausible, so that many of the examinees have to use their reasoning ability to determine which answer is best. Instead, the correct answer to a vocabulary item on the ALT is always a synonym that is an even easier word than the stem, that is, at a higher frequency of usage, and is the only one of the three alternatives that is close in meaning to the stem. This procedure for developing the ALT helped assure that a higher score on the test would reflect a proportionally greater number of words known; i.e., a greater number of A_LWords. In fact, in the manual for the test, a table is provided for transforming a raw score on the ALT into an estimate of words known, A_LWords, based upon the sampling from the Carroll, Davies, and Richman (1971) ranked list.

The manual for the ALT also contains reliability estimates (Carver, 1994a). The standard error of measurement is around 0.7 GE units in each of grades 3 through 12, plus college and graduate school, and the average alternate form reliability coefficient is .85 in grades 3 through 12; the corresponding reliability coefficients in college and graduate school are comparable or higher.

One way to evaluate the validity of the ALT, as an indicant of A_L, is to relate it to the RELT as a measure of E_L. Carver (1987b) reported that the ALT correlated .73 with the RELT—with grade level in school partialled out, for 107 students in grade 3 to graduate school. This is a very high correlation considering that E_L is made up of both A_L and R_L, in roughly equal proportions; the corresponding partial correlation for an indicant of R_L in this same research was .77.

The validity of these GE scores for A_L has also been established by asking 568 students in grades 3 through 12 and graduate school, to read passages once at varying difficulty levels, D_L, and then to make a judgment or estimate of their percentage of comprehension of the passage (Carver, 1990b). These percentages constitute a measure of rauding accuracy, or A_r. Notice that A_r represents the accuracy of comprehension, A, that accompanies the rauding rate when the individuals are instructed to read the passage once at their normal reading rate. These students were also given the ALT, so that the relative easiness of the passage, $A_L - D_L$, could be related to the estimates of accuracy of comprehension, or A_r. For the means of rauding accuracy at 17 levels of relative difficulty, the relationship between A_r and $A_L - D_L$ was almost perfect (r = .992). The equation of the straight line that fit these data was as follows: A_r = .04 ($A_L - D_L$) + .64, which is Equation 2-2, referenced earlier. So there is a high relationship between the percentage of a passage that was accurately comprehended, A_r, and the relative easiness of the passage, $A_L - D_L$, when the indicant of A_L was a reading vocabulary test. Any measure of A_r for texts at a constant level of difficulty, D_L, would automatically provide an indicant of A_L because variations in A_r would be linearly related to A_L according to Equation 2-2.

Other Standardized Tests. There are several standardized tests of reading comprehension that are administered individually and are untimed or unspeeded; these tests should provide indicants of accuracy level, A_L. One example is the Woodcock Reading Mastery Tests (WRMT) - Passage Comprehension. On this measure, an individual must read short paragraphs which increase in difficulty. Each of these paragraphs have had one word deleted and replaced with a blank. The individual must figure out what word goes into the blank, with more than one word being scored as correct because synonyms or equivalents of the correct word are equally correct. It is likely that this particular indicant of A_L would load higher on fluid intelligence, Gf, than would the ALT, because more abstract reasoning is likely to be involved in figuring out what word best fits into the puzzle of the missing word. That is, information needs to be gathered from all the context surrounding the missing words, possible words need to be generated, and a decision needs to be made regarding the best word.

There are two other tests that can be mentioned that are similar to the WRMT- Passage Comprehension, and would provide indicants of A_L. They are the Peabody Individual Achievement Tests (PIAT) - Passage Comprehension and the Woodcock-Johnson Tests of Achievement-Passage Comprehension.

Other Tests. A time-limited word identification test would also provide a measure of raudamatized words, or A_LWords, and would constitute an indicant of A_L. On this type of test, words that vary in frequency of usage are each pre-

sented for a brief period of time, such as 1 or 2 seconds. For example, Olson, Wise, Ring, and Johnson (1997) recently administered such a test to poor readers in grades 2 through 5; they described the experimental test as "... a difficulty-ordered list of words presented individually for 2 sec. on the computer screen" (p. 245). On this type of test, if the word presented was a known word that had been practiced until it was overlearned, then it should be quickly recognized and pronounced correctly during the brief time limit. Words presented that are not known or had not been overlearned from practice, probably would not be pronounced correctly because there would not be enough time to develop a reasonable guess. Therefore, the number of correctly pronounced words on this type of test should be an indicant of the size of the lexicon of known words, A_LWords, and an indicant of A_L.

The instructional level from an informal reading inventory, IRI, would ordinarily provide an indirect measure of A_L. These IRIs are often constructed by teachers from a basal reading series which require the student to read aloud from successively higher GE levels with comprehension sometimes assessed at each level. Ordinarily, the students read increasingly difficult material at an oral rate of their own choosing, until they no longer can accurately comprehend or accurately pronounce the words. These tests are not standardized and are not likely to be reliable, but it can be seen that they would be measuring a level of reading comprehension in GE units that is somewhat similar to accuracy level, A_L, in that both require increasingly difficult text to be accurately read and both yield GE units.

Another kind of test providing an indicant of A_L is the score on an untimed or unspeeded measure of reading comprehension. It may be remembered that when an individual is allowed to read a passage once, the accuracy of text comprehension is called rauding accuracy, or A_r, and it is equal to .04 (A_L - D_L) + .64 (from Equation 2-2, or Equation C-7). If everyone is given the same text to read in this situation, then the measure of text comprehension, A_r, would reflect only individual differences in A_L because D_L would be constant in Equation 2-2. That is, a measure of A_r in this situation should correlate almost perfectly with A_L, as noted earlier. In actual research situations, the tested individuals are often given as much time as they want to read so some may read the text more than once. However, when individuals are allowed to read text longer than the time they need to read it once (see Equations C-3 and C-4 in Appendix C), the increase in the accuracy of text comprehension is not so large, relatively, that such reading comprehension measures, A, are rendered invalid as indicants of A_L. This means that reading comprehension tests that happen to have ample time limits so that almost everyone can finish the test without hurrying or feeling time pressure will correlate higher with A_L than they will with R_L or E_L. As the time limits on the comprehension test are shortened, then rate level, R_L, becomes more of a factor so the test starts to become a

better indicant of efficiency level, E_L. If the test has a very short time limit then it could become much more of an indicant of R_L than A_L or E_L.

Concluding Comments. The close connection between the number of printed words known (reading vocabulary, or A_LWords) and the ability to accurately comprehend text of increasing difficulty, A_L, is assumed to involve a causative connection. That is, it is assumed that it would be impossible to gain one GE on the ALT test without also gaining 1 GE on a more direct measure of A_L such as the DRP. For example, a program designed to increase vocabulary knowledge might be thought to provide an artifactual increase in ALT with no increase in the DRP. Yet, it seems unlikely that individuals could learn and remember the meanings of the hundreds of words necessary to increase ALT one GE without gaining the associated verbal knowledge that would also allow text at the next GE to be accurately comprehended on the DRP. Conversely, it is assumed that it would be impossible to gain one GE on a more direct measure of A_L, such as the DRP, without also gaining one GE on the ALT. It seems unlikely that a person would be able to accurately comprehend text at the higher level of difficulty on the DRP without learning the thousands of concepts in the form of words that would also result in a one GE gain on the ALT.

The above assumptions seem to be endorsed by several other researchers. For example, Curtis (1987) contended that vocabulary growth increases comprehension ability and that better reading comprehension enhances vocabulary. Also, Stanovich (1986) has argued for a reciprocality between vocabulary and reading deficits, such that a problem in reading will lower vocabulary development.

Summary of Theory

Accuracy level, A_L, is a very important theoretical construct because it is one of only two factors that are directly causal with respect to high and low reading achievement, or E_L. A_L is actually rauding accuracy level, which is defined as the most difficult text, in GE units, that an individual can read accurately when the text is read at the individual's rauding rate. A_L is similar to the traditional concept of reading level, or Betts' (1946) instructional level. The Degrees of Reading Power test measures the most difficult text that an individual can accurately comprehend under unspeeded conditions so it provides an indicant of A_L.

The number of words that can be accurately and quickly pronounced is a measure of A_LWords, or raudamatized words. Reading vocabulary is a measure of A_LWords that is so closely connected to the accuracy of text comprehension that it is also an indicant of A_L. Also, text that is relatively easy, $A_L > D_L$, contains words that can be accurately processed at the rauding rate, so these

words in relatively easy texts are raudamatized words. Therefore, the size of the lexicon of raudamatized words, called A_LWords, is likely to be almost exactly the same as the size of the *autonomous* lexicon as conceptualized by Perfetti (1991a), or the number of unitized words as conceptualized by Ehri and Wilce (1983). The lexicon of raudamatized words may not be exactly the same as the lexicon of automatized words, because words can become automatized (not requiring much attention) at a lower rate than raudamatized words. Automaticity and automatized words have been tied so closely to attention capacity and to an attribute of the reader (LaBerge & Samuels, 1974), that it is likely to be more confusing than illuminating to talk about automatized words, or automaticity, in connection with the causal model and rauding theory.

The A_L construct integrates measures of reading level, reading vocabulary, and the number of words on a time-limited word identification test. For example, a factor analysis of scores on the Passage Comprehension test from the Woodcock Reading Mastery Tests, the Vocabulary test from the Iowa Test of Basic Skills, and the time-limited word identification test used by Olson et al. (1997) should result in one factor—the A_L factor. There are many unspeeded standardized tests of reading comprehension that should provide indicants of accuracy level, A_L, such as the aforementioned Degrees of Reading Power test and the Woodcock Reading Mastery Tests—Passage Comprehension. At present, the best indirect measure of A_L is probably the Accuracy Level Test, ALT, which is a reading vocabulary test that seems to be reliable, valid, and quickly administered. There are no standardized direct measures of A_L available, at present, although the DRP test is close to meeting this criterion.

Summary of Evidence

Measures of reading vocabulary, A_LWords, and unspeeded measures of the accuracy of comprehension, A_L, are highly related (Anderson & Freebody, 1981; Beck, McKeown, & Omanson, 1987; Carroll, 1993; Holmes, 1954; R. L. Thorndike, 1973; Thurstone, 1946). The correlation between perfectly reliable measures of reading vocabulary tests, A_LWords, and measures of the accuracy of comprehension, A_L, can be estimated to approach 1.00 under optimum conditions (Sticht, Hook, & Caylor, 1982). These data suggest that the level of reading comprehension is very similar to the level of reading vocabulary, A_LWords, which in turn supports the existence of the theoretical construct called rauding accuracy level, A_L, and supports the use of a reading vocabulary test scaled into GE units as an indirect measure of A_L.

The validity of the ALT as an indicant of A_LWords and the DRP as an indicant of A_L has been researched by correlating scores on these two tests with each other. In grades 7 and 8, the ALT correlated .80 with scores on the Degrees of Reading Power, DRP, test (Carver, 1992a). The ALT and the DRP

also have been factor analyzed along with other tests; the ALT loaded on an accuracy factor about equally with the DRP (.92 for DRP vs. .90 for ALT).

Finally, evidence which seems to support the existence of raudamatized words comes from the eye movement research of Rayner and Duffy (1986). They found that fixation times for low-frequency words were higher than fixation times for high-frequency words even when word length was controlled. This result can be explained as follows: The low-frequency words probably were raudamatized whereas the high frequency words probably were not raudamatized.

Implications

Researchers need to know what they are measuring when they administer a standardized test of reading comprehension, or a reading achievement test. For example, Stone and Brady (1995) administered the Passage Comprehension subtest of the Woodcock Reading Mastery Test, and then generalized about "reading achievement." In a general sense, it is true they were measuring achievement in reading. However, they were not measuring E_L, because this Woodcock test is untimed and unspeeded so that it does not involve rate, or R_L. Therefore, Stone and Brady were likely to be measuring accuracy level, A_L. We need to know relatively precisely what we are measuring from a theoretical standpoint, so our theories are more likely to be testable, and therefore more readily refutable. Researchers need to be more aware of the difference between A_L and E_L, so that they can be more precise about whether they are measuring A_L or E_L; this distinction is important because C_s is a distal cause of E_L but C_s is neither a distal or proximal cause of A_L.

It would seem to be almost impossible to investigate lawful behavior in reading, without the A_L construct. Reading level and instructional level, are similar in concept to A_L but A_L is defined much more precisely; for example, reading level contains no qualifications with respect to rate and it is often determined under oral reading conditions. If the A_L construct is accepted, then it seems best to fuse it with (a) the concept of reading vocabulary size, (b) the concept of the accuracy of text comprehension under unspeeded conditions, and (c) the concept of time-limited word identification. Furthermore, if the A_L construct is accepted, then it is almost impossible to avoid accepting the lexicon of raudamatized words, or A_LWords.

Forget Me Nots

Rauding accuracy level, A_L, is a theoretical construct that tries to incorporate all of the good ideas that have traditionally been associated with the following concepts: (a) reading level, (b) level of reading comprehension, and (c) level of reading vocabulary. These earlier concepts have been upgraded by the A_L construct, which is more precisely defined both from a theoretical and an operational standpoint. Accuracy level, A_L, is the most difficult level of text that individuals can read accurately when they read at their own typical rate. For example, an individual may have a fourth-grade accuracy level, $A_L = 4$, as measured by a reading vocabulary test.