Readability assessment of the textbooks used in the 2 year nursing program at San Bernardino Valley College ...

Rita Sturgeon

Follow this and additional works at: https://scholarworks.lib.csusb.edu/etd-project

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
Sturgeon, Rita, "Readability assessment of the textbooks used in the 2 year nursing program at San Bernardino Valley College ..." (1981). Theses Digitization Project. 266.
https://scholarworks.lib.csusb.edu/etd-project/266

This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
READABILITY ASSESSMENT OF THE TEXTBOOKS USED
IN THE 2 YEAR NURSING PROGRAM
AT SAN BERNARDINO VALLEY COLLEGE

Presented in Partial Fulfillment of the Requirements
for the Master's Degree in a Special Interest
at California State College
San Bernardino

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
A Special Interest

by
Rita Sturgeon
June 1981
READABILITY ASSESSMENT OF THE TEXTBOOKS USED
IN THE 2 YEAR NURSING PROGRAM
AT SAN BERNARDINO VALLEY COLLEGE

Presented in Partial Fulfillment of the Requirements
for the Master's Degree in a Special Interest
at California State College
San Bernardino

by
Rita Sturgeon, MA
June 1981
California State College
San Bernardino

Approved by
Advisor

Committee member

Committee member
ABSTRACT

San Bernardino Valley College's Board mandated a minimum 12th grade reading level be established as a prerequisite to acceptance into the Registered Nursing Program effective spring semester, 1982. Grade level is a difficult concept to define for either students or textbooks. Textbook grade level is termed "readability", and various readability formulas are explored and discussed. A readability process is selected and an assessment is done on the core textbooks used in the Registered Nursing Program. The results suggest that difficulties may still be encountered in textbook reading by the students who are accepted. A reading class designed as a textbook support system is recommended.
ACKNOWLEDGMENTS

to PJ

This thesis, the product of six month's labor, is complete only because of the tolerance, generosity, emotional support, and work of so many wonderful people.

A special acknowledgment must go to my four-legged roommate who supervised every word I wrote.
<table>
<thead>
<tr>
<th>Table of Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER I. PURPOSE OF ASSESSMENT</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER II. REVIEW OF LITERATURE</td>
<td>8</td>
</tr>
<tr>
<td>Definition of readability</td>
<td>8</td>
</tr>
<tr>
<td>History of grade level for students and textbooks</td>
<td>8</td>
</tr>
<tr>
<td>History of readability</td>
<td>12</td>
</tr>
<tr>
<td>Review of commonly used formulas</td>
<td>13</td>
</tr>
<tr>
<td>Lorge</td>
<td>13</td>
</tr>
<tr>
<td>Flesch</td>
<td>14</td>
</tr>
<tr>
<td>Dale-Chall</td>
<td>16</td>
</tr>
<tr>
<td>Spache</td>
<td>18</td>
</tr>
<tr>
<td>Fry</td>
<td>19</td>
</tr>
<tr>
<td>Botel</td>
<td>22</td>
</tr>
<tr>
<td>SMOG</td>
<td>24</td>
</tr>
<tr>
<td>Discussion of readability</td>
<td>26</td>
</tr>
<tr>
<td>CHAPTER III. SELECTION OF PROCEDURE</td>
<td>30</td>
</tr>
<tr>
<td>Choice of variables</td>
<td>30</td>
</tr>
<tr>
<td>Frequency of samplings</td>
<td>31</td>
</tr>
<tr>
<td>Rationale for choice of formula</td>
<td>32</td>
</tr>
<tr>
<td>Computerization</td>
<td>38</td>
</tr>
<tr>
<td>Chapter Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>CHAPTER IV. RESULTS OF COMPUTERIZED READABILITY</td>
<td>40</td>
</tr>
<tr>
<td>ASSESSMENT</td>
<td></td>
</tr>
<tr>
<td>CHAPTER V. SUMMARY</td>
<td>42</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>Recommendations</td>
<td>43</td>
</tr>
<tr>
<td>Bibliography</td>
<td>45</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>48</td>
</tr>
</tbody>
</table>
PURPOSE OF ASSESSMENT

The process by which San Bernardino Valley College selected applicants into its R.N. program was viewed by students and board members as being too subjective. A point system was developed whereby a student could receive a maximum of 4 points based upon a 4.0 GPA in the general requirements of state mandated classes. In addition, extra points could be earned by taking specific preparatory classes which included a possible point for taking a reading class.

The SBVC nursing program is a prestigious program with attractive job possibilities with desirable remuneration. Of the approximately 200 prenursing applicants yearly, only 30 will be accepted each semester into the program. The point for reading was developed in an effort to weigh admission less heavily on GPA, and to allow a competitive chance for persons with strong humanistic qualities and high motivation for nursing.

Student hopefuls expressed considerable concern. They appeared at the college board meetings to vociferously attack this point system with the assertion that it favored academically inferior students. Dr. Byron Skinner chaired a committee that was formed to investigate admission procedures of other schools. In November 1980, the committee's recommendation was to alter the point system with more, but not all, points based
on grade point average and to eliminate the point for taking a reading class. Additional study was recommended.

In January 1981 Dr. Skinner informed the board that no other way could be found to determine admittance. He was supported by the college President, Dr. Arthur Jensen, who assured the board that the point system with emphasis on GPA not only was in line with other schools but that students with lower GPA's were found to be more likely to fail the State licensing examination. The board recommended additional study.

In March 1981 the SBVC College Board adopted the revised proposal, "Preparation for Licensure as an R.N." to be effective beginning with the Spring Semester, 1982. Included in the general requirements on page 1, section 2 of this proposal is the establishment of a 12th grade reading equivalency determined by the Nelson/Denny test or any equivalent institution-approved test.

Does the establishment of a 12th grade reading level adequately prepare a student to manage the difficult reading load required in this stringent program? One method of assessment and the subject of this study will be to determine the grade level of the required textbooks used within the 2-year nursing program at San Bernardino Valley College.
REVIEW OF LITERATURE

Definition of Readability

Readability is a construct designed to measure a construct. Readability is determined either by a regression formula based on multiple correlations or plotted on a graph. It is designed to assign grade levels to books to match the grade levels of students. Both readability and grade levels are constructs.

History of Grade Levels for Students and Textbooks

Industrialism, the scientific method, and mandatory public education free to the masses all contributed in the late 1800's to the grave concern we have had: how to set standards for both products and processes. "Modern critics like to say pejoratively that educational scientists of those days equated efficiency with science."\(^1\) It is true that these early educational scientists tried to solve educational problems by means of experimental and statistical techniques, particularly the measurement of ability and achievement. This zeal for measurement brought forth an abundance of "facts" about the level of the student and the content of the textbook. Studies were undertaken to find out how students learn and to design new methods for overcoming their reading difficulties. As a result of these

studies, educators became overly concerned with being able to determine the grade level of students. It did not deter them a whit that grade level is just a construct.

Grade level is actually a statistical device at which the average examinee in a norm group had a given number right. Grade levels do not indicate the appropriate grade placement for a student. For instance, suppose that a student enrolled in the 11th grade took a test containing 10th, 11th, and 12th grade material. If the student receives a grade level of 13th grade, this cannot possibly mean that he can necessarily do 13th grade work, since he was not actually tested on 13th grade material. It merely tells us that he can perform 10th, 11th, and 12th grade material as well as the average students in the 13th grade can perform these tasks. Even if the student's grade level had fallen within the grade level range covered by the material in the test, it would not indicate appropriate grade level placement for him since his score would represent some unknown combination of successes and failures on tasks for three different grade levels.

Because of possible misinterpretations of grade level, the most recent edition of Standards for Educational and Psychological Tests and Manuals call upon test publishers to abandon or discourage the use of grade level equivalents.


3 Ibid, 76.
Yet, our educational system continues to measure students in terms of grade level. As a result it has become necessary to attempt to match the grade levels of the students with the grade levels of the books being utilized.

Readability refers to the relative difficulty of a piece of material. Readability formulas generally assess language variables in a written selection thereby generating an index of probable difficulty for the average reader. Readability values are commonly reported in terms of a grade level.

A major concern of any teacher is the choice of the right book for each student so that maximum achievement gain is possible.

The selection of a book for instructional purposes should be based upon careful consideration of two factors: the interest value of the book for the student, and the difficulty of the book in terms of the student's skills and knowledge.

Bradley's statement is a reflection of the thinking which educators finally came round to, which is that the "science" aspect of book difficulty must be tempered with the "art" aspect in order to give a readability indication. Teachers who are concerned with building a curriculum to meet the needs of mass education find that they are held accountable for their teaching methods and student performance in terms of demonstrated outcomes. The teaching-learning process

---

⁴G. Britton, & M. Lumpkin, "Computerized Readability Verification of Textbook Reading Levels," Reading Improvement, XIV, (Fall, 1977), 193.

⁵John Bradley, "Using Readability to Improve the Content Validity of Information Placement Tests," Reading Improvement, XIV, (Fall, 1976), 182.
becomes a function of numerous causal factors as to why some students don't learn ranging from the learner, the teacher, the methods, heredity, environment, race, sex, language, diet, drugs, and lack of material correctly identified as to reading difficulty. It is time that we investigate the possibility that some students don't learn because, although they can read, per se, they cannot read well enough to comprehend because the textbook is too difficult.

Perhaps it is time to look at the common teaching tools of textbooks more closely and focus on one vital aspect of textbooks . . . readability. It would seem that the scientific method of counting, weighing and measuring would be the only possible method of readability assessment.

A readability formula is simply a mathematical equation derived by regression analysis. This procedure finds the equation which best expresses the relationship between two variables, which in this case are: a measure of the difficulty experienced by people reading a given text, and a measure of the linguistic characteristics of that text. This formula can then be used to predict reading difficulty from the linguistic characteristics of other texts.

Readability formulas do attempt to combine the art of reading by validating their formulas with actual comprehension scores

---

6 G. Britton, & M. Lumpkin, "Computerized Readability Verification of Textbook Reading Levels," Reading Improvement, XIV, (Fall, 1977), 193.

acquired by human test subjects, with the science of a method of weighing and measuring linguistic characteristics.

History of Readability

Early attempts at measuring readability were even more cumbersome than they are today. Which characteristics of linguistics one might use to relate to what difficulties of reading were not as clearly defined in 1923 when Lively and Pressey devised the first workable readability formula. Their statistical formula was used by educational researchers and learning theorists only. It was not until 15 years later, in conjunction with America's beginning concern with our "reading problem" that a spate of formulas began appearing on the academic scene.

It is certainly a major problem in mass, free, public education how to tell whether a particular piece of writing is likely to be readable to a particular group of readers. There are three possible different solutions to this problem. The first one is to guess. Both teachers and writers have been making estimates of readability for a long time with a fairly adroit skill as a result of their experiences. The second solution is to give a comprehension test. A comprehension test constructed to predict readability must be built and refined with considerable care. With the large amount of reading material being published and available today, this is simply not practical. A third solution is needed.

Readability formulas have come to provide a third possible solution to the problem. A readability formula uses counts of language variables in a piece of writing in order to provide an index
of probable difficulty for readers. It is a predictive
device in the sense that no actual participation by
readers is needed.

Review of Commonly Used Formulas

1939 - Lorge
1943 - Flesch
1948 - Dale-Chall
1953 - Spache
1961 - Fry original
1962 - Botel
1969 - SMOG McLaughlin
1977 - Fry college extended

Lorge, 1939

Irving Lorge is credited with building the first formula
that has practicality in that it is easy enough to be used. He
is more notably credited with breaking the ice to allow other
designers to aim toward simplifying the labourious process of
determining readability. The original Lorge formula determined
grade placement by counting: the average sentence length in
words; the number of prepositional phrases per every 100 words;
and the number of different "hard" words not included on the
Dale list of 769 words. He calculated his formula thus:

\[ X_1 = 0.07X_2 + 0.1301X_3 + 0.1073X_4 + 1.6126 \]

In 1948, after discovering an error, he recalculated it:

\[ X_1 = 0.06X_2 + 0.10X_3 + 0.10X_4 + 1.99 \]

---

where:

\[ X_1 = \text{grade level} \]
\[ X_2 = \text{average sentence length in words} \]
\[ X_3 = \text{number of prepositional phrases per 100 words} \]
\[ X_4 = \text{number of different hard words not on the Dale list of 769 words}. \]

In order to determine grade level, Lorge used as his criterion the ability to correctly answer 75% of the test questions of the McCall-Crabbs STANDARD TEST LESSONS IN READING which had been published in 1925. The prepositional phrase count was soon found out to be unreliable, not statistically, but actually. For example, if a passage were given to five different persons to count the prepositional phrases, the likelihood is high that it would yield five different counts. What actually constitutes a prepositional phrase and what does not caused later researchers to seek a formula that could eliminate the prepositional phrase variable. Also some revision work was done to attempt to eliminate the Dale word list count by using a count of the letter length of words. Lorge's contribution as a precursor is undeniable.9

Flesch, 1943

Rudolf Flesch objected to the use of the Dale list of 769 words in that he determined that it did not differentiate between the higher levels of difficulty. He devised his formula

---

using affixed morphemes instead. He based his concept on that used by Lorge in that he used as his criterion the McCall-Crabbs STANDARD TEST LESSONS IN READING, with his formula so constructed to also predict the average grade level of a student who answered correctly 75% of the test questions. Its multiple correlation coefficient was \( R = .74 \). The specific purpose which Flesch had in mind in eliminating the word list was an attempt to measure abstractness. Another major way in which Flesch differed from Lorge is that he did not design his formula to be read in a grade level but rather into factors from 0 to 100, "with 0 meaning practically unreadable and 100 meaning easy for any literate person." Since these factors were not read as grade levels, Flesch decided to develop his readability on two different scales. One he called the reading ease scale: the other he called the human interest scale. A human interest scale factor of 0 indicated no human interest: a scale factor of 100 indicated the passage was full of human interest. In 1948, Flesch revised his own formula because he learned that the count of affixes was too time consuming. His reading ease formula:

\[
\text{reading ease} = 206.835 + .846w_l + 1.015s_l
\]

where:

- \( w_l \) = number of syllables per 100 words
- \( s_l \) = average number of words per sentence

His human interest formula:

human interest = 3.635 pw + .314 ps

where:

pw = number of personal words per 100 words

ps = number of personal sentences per 100 sentences.

The human interest formula was widely used by newspapers, advertising, government publications, and many academic institutions used it in the curriculum for their journalism classes. It was the reading ease formula, however, which attracted the most attention. The simplifications he designed were such an improvement over the tedious process of checking against a word list which Lorge had proposed that it became usable to teachers in selecting classroom materials even though it was still time consuming to count and complex to compute. "Although the average time needed to do a readability is considerably faster than Lorge's, it is still too long for practical use."11

Dale-Chall, 1948

Edgar Dale and Jeanne Chall were commissioned to perform a readability service on the educational materials published by the National Tuberculosis Association in order to rewrite them to be able to be understood by the average adult. They selected the original Flesch formula although they were concerned when they found it arbitrary in the sense that two people who made a count usually came out with a different number of affixes.

---

Accuracy could only be counted upon if a dictionary were consulted, and that was too time consuming. Secondly, they found the count of personal references not to be a reliable index of difficulty. They undertook to find a more efficient means of predicting readability.

They also based their sample criterion on the McCall-Crabbs STANDARD TEST LESSONS IN READING. Lorge contributed his data sheets on these samplings to them, and Flesch contributed his counts of affixed morphemes and personal references. Their theory was that Lorge's idea of a word list was a better predictor, but that his word list had been too small. They used the Dale list of 3000 words instead. They did intercorrelation studies and after making several combinations of factors decided that the most efficient empirical formula was:

\[ X_{c50} = 0.1579X_1 + 0.0496X_2 + 3.6365 \]

where:

- \( X_{c50} \) = reading score of a pupil who could answer 50% of the test questions
- \( X_1 \) = average number of words outside Dale list of 3000 words
- \( X_2 \) = average sentence length
- 3.6365 = constant

Dale-Chall did agree with Lorge and Flesch in thinking that the measure of vocabulary load is the most important factor in reading difficulty. They next conducted numerous studies comparing the formula with: (1) the readability levels made by experts' judgments, (2) other readability scores based on
other formulas, and (3) actual comprehension scores of readers. As a result of their studies, they developed an estimated corrected grade level to within two years from 4th grade to 12th grade, as well as a formula score converted to "college" and an additional score range to "college graduate." Their new formula of two factors (average sentence length and percentage of unfamiliar words outside the 3000 word Dale list) contributed extensively to scores of researchers testing, modifying and validating it in ensuing years.

We must remember at all times that a formula is a statistical device. It does not mean that all long sentences are hard to understand. There are some very short sentences that may be harder to comprehend than longer ones. The same holds true for the use of words. Sometimes familiar words are used in a symbolic or metaphoric sense. Readability formulas are not sensitive to subtle variations in meaning. We do not claim that the formula developed here is definitive. The nature of the multiple-correlation coefficient makes this point rather obvious.\(^\text{12}\)

Spache, 1953

An important contribution to readability formulas came from the educator, author, reading expert, George Spache. Spache validated his formula against levels of classroom use for 152 books in grades one to three and found a multiple correlation coefficient of .818. He also found a correlation coefficient of .95 between his formula scores and grade levels.

of primary books. His formula is:

\[ \text{grade level} = 0.141X_1 + 0.086X_2 + 0.839 \]

where:

\[ X_1 = \text{average sentence length in words} \]
\[ X_2 = \text{number of words outside the Dale list of 769 words.} \]

His formula was designed for primary grades 1 to 3 only.\(^{13}\)

**Fry 1961 and 1977**

Edward Fry, currently a professor at Rutgers University, was on a Fulbright lectureship at Makerere College in Uganda, Africa, at the same time that a group of African teachers on a UNESCO training project were teaching English as a second language. They asked him to help them in textbook selection. Although readability formulas had been around for quite awhile by this time, they required too much time and the statistics were not always that accurate.

> I admit that statistics are an important research tool, but certainly not the only one . . . there is nothing wrong with trying out something. Quite a few things were invented, developed, or improved by trial and error.\(^{14}\)

This is exactly what Fry set out to do, not as a research project with a "proper" statistical design, but to realistically aid teachers to select material on a proper difficulty level.

---


This is the procedure that Fry followed. He plotted a large number of passages from the OXFORD ENGLISH READERS, which have a wide range of difficulty, on to a graph. In order to plot them, he used two linguistic variables: morphology and syntax within a contained space measured in length of semantic units. Morphology is measured in syllables: syntax is measured by the sentence frequency within a 100 word passage: the contained length of semantic units is predetermined to be 100 words. When the sample passages had been counted, he drew a curved line down the graph which represents the smoothed mean of the plots of sample passages.

If you plot a large number of passages with a wide range, they will tend to fall somewhere near the line. In short, it is an 'eye ball' job. However . . . higher mathematics tell me that 'smoothing a curve' in this manner is just about as accurate as doing it by complicated formula. 15

Grade levels were not assigned to the graph at the time it was developed and Fry included the graph in the appendix of the book he was publishing in 1963, Teaching Faster Reading. The following year, in 1964, he published it in a British journal. For years nobody ever used the graph, possibly because American educators tend not to read British journals. In 1968, Fry decided to give his graph some "Americanization". With America's all consuming concern with grade levels, that

meant that grade level designations would have to be added to it.

Grade level designations were determined by simply plotting lots of books which publishers said were third grade readers, fifth grade readers, etc. I then looked for clusters and 'smoothed the curve.' After some use and correlational studies, the grade level areas were adjusted. The adjustments that he made followed correlation studies with Spache, Dale-Chall, Flesch and Lorge.

By 1969, American educators were using the graph in teacher training classes as well as in textbooks.

The readability graph's contribution seems to be in simplicity of use without sacrificing much, if any, accuracy, and its wide and continuous range from grade one up through college. That it was not copyrighted and could be reproduced on one sheet of paper might have helped also.

Fry's graph spawned scores of research studies for validity and reliability. Fortunately time and other research studies have continued to show the efficacy of the two inputs of the graph: morphemes measured in syllables and syntax measured in sentence length found within 100 semantic units. The Fry graph correlates .95 with Flesch, .85 with Dale-Chall, and

---


.85 with Carver's Rauding Technique. In 1969 Magginnis extended the graph downward into preprimer levels to enable it to be used with those required shorter passages. Although some researchers attempted to improve the accuracy of the graph by adding sets of vocabulary words to be consulted, a la Dale-Chall, it was learned that this did not improve accuracy and certainly complicated the graph's use. In 1977, Fry extended the graph to 17th grade level in order to be used specifically in college level material.

It is known that vocabulary continues to increase throughout the college years. I therefore am proposing this extension as a relative difficulty differentiation rather than a normed score.

Botel, 1962

Morton Botel, a professor in the Graduate School at the University of Pennsylvania, decided to devise a method of predicting readability that was not a graph such as Fry's, nor was it a regression equation such as Flesch, Dale-Chall, Spache and Lorge used. Botel termed his concept, "a readability technique" and published it in 1962 under the title: Botel Predicting Readability Levels: A Simple Technique for Establishing Reading Levels of Books. Botel's method is to predict reading levels from the median difficulty of samples

---


level beyond 4th grade. He designed two tally sheets, one for primary and one for 4th grade and above. A readability count was done by putting tally marks into whichever block on the worksheet each word would be determined as, by checking it against this graded vocabulary list. In this way Botel extrapolated levels up to grade twelve. Botel validated his method, which has ended up being used as a general formula, by comparing the vocabulary with that used in various reading materials from elementary textbooks, junior high books, senior high books, Time, Reader's Digest and the New York Times.

Smog, 1969

G. Harry McLaughlin, a psycholinguist, was a professor in the School of Journalism at Syracuse University when he decided to develop a readability formula that was better than Fry's in that it is quicker, and better than Lorge's, Flesch's, Dale-Chall's or Spache's because it did not rely upon a regression analysis.

Regression analysis can find the best formula only if the investigator happens to have chosen the best general form for the equation. What previous investigators have generally overlooked is the fact that semantic and syntactic difficulty interact. A slight difference in word or sentence length between two passages does not indicate the same degree of difference in difficulty for hard passages, as it does for easy passages. Therefore, a readability formula should not be of the usual form.

The usual form is to use a constant and add to it whatever variables the designer assessed as proper, such as: a constant plus word length, plus sentence length, or \( a + b + c \). McLaughlin said because of the interaction of semantics and syntactic difficulty, the constant should be added to the square root of the polysyllabic count, or \( a + b \). McLaughlin found that thirty sentences are needed for the criterion of readability using his formula. Other formulas require samples of 100 words, at least three such samples. Thirty sentences typically cover 600 words which increases the reliability. In order to improve the sampling, he directed us to select the thirty sentences in clusters of ten from three different places in the book being tested: next, count the polysyllabic words within these thirty sentences, and convert that into some meaningful number. McLaughlin validated his formula against the McCall-Crabbs STANDARD TEST LESSONS. One major difference in his validation is that he used scores of 100% comprehension as opposed to the 50% or 75% used by earlier formula makers as his indicator of difficulty. He next determined a regression equation relating the polysyllable count on each lesson in McCall-Crabbs to the mean grade score of students who scored the 100% comprehension. After considerable statistical computation using an IBM 360/50 computer, he found that "3" computed as the constant. His formula reads:

\[
\text{SMOG grade} = 3 + \text{square root of polysyllabic count}
\]

McLaughlin named his formula SMOG in deference to Robert Gunning who first came up with the idea of using polysyllabic
words as a measure of semantic difficulty. Gunning, in an effort to give credit to Flesch, used the first initials of their last names, F and G, and named his FOG. McLaughlin wanted to give tribute to Gunning while also giving credit to his London birthplace. With tongue in cheek he named it the SMOG grading formula.

Discussion of Readability

The academic pursuit which motivated the readability predictor originators is evidenced by the revising, updating, willing corrections of errors and recalculation. Aided by each other they worked with a concept that borders on the mystical (language and the processing of meaning from the printed word.) The fact that they shaped this into a statistically meaningful predictability is a contribution to academia.

The researchers agree that every piece of textbook material cannot be given the precision, time and effort of large samplings of written tests of comprehension whose score must be evidenced as to reliability and validity in order to determine readability. A book-by-book field testing is not a practical solution in determining the grade levels which our mass educational system demands to coincide with the grade levels of our masses in public education. Formula makers all realize that an acceptance of predictive readability formulas devised statistically to predict "comprehension" is the only pragmatic answer. Other researchers who have published in this
area also realize, and accept that formulas are not perfect predictors. Variations among the formulas include: different sampling lengths, different sampling intervals, different linguistic variables computed within the samples, and different methods of computation. The formula builders also realize that assessing passages, elementary books, or articles is not as complex or unwieldy as determining college level textbooks for readability.

The fact is that readability formulas are just statistical devices used because we can weigh and measure with them, but that "If the learning of language remained natural and relatively easy, what caused the learning of reading which requires no new, unsolvable demands from the mind-brain-body system to produce so many problems that it is of national concern? How did we come to 'see' and treat the reading process as only a sequential, linear activity?"

Researchers agree that the basis for these restrictive models of reading are the disciplines that reading theorists chose to adopt, out of necessity: physics and medicine. It was these sciences that were held in highest esteem in our society. To achieve credibility and respectability for any discipline, a theorist need only apply the methods, techniques, observations, collections and quantifications of physics in conjunction with the diagnostic-prescriptive methods of the

---

medical assemblage. Because theorists created this more scientific, more analytical, more 'objective', left hemispheric theory of the reading process, we must not develop a concomitant tunnel vision. We must not be blinded by the spectacular successes of these mechanistic models. We must see and accept the inadequacies and the limitations of these formulas as are equally visible in the sciences from which they grew. The builders of readability formulas frequently caution us that we must not lose the creative process that made language learning the holistic function which caused Bloom to taxonomize learning into both a cognitive and affective domain. Within his hierarchy of learning in his discussion of comprehension, Bloom cautions us to be aware that "comprehension" is beyond simple recall. "The reader must, if he is to make full use of a communication, be able to extend it beyond the limits set by the writer as well as to apply some of the ideas of the communication to situations and problems not included explicitly in the communication."23 We must remember that college textbook reading requires all six of the members of Bloom's taxonomy: knowledge - the recall of specifics; comprehension - the understanding of an idea being communicated; application - the use of abstractions to concrete situations; analysis - the ability to see the relationships

between ideas; synthesis - the combining of ideas to rearrange them into a new pattern; evaluation - the quantitative and qualitative judgments about the communication.

A readability formula cannot accurately measure these components. Although reading has been reduced to this simplistic, meticulous, prescriptive approach, we are becoming aware that the "principles of reading are not found in phonetic and structural analysis, but instead, within each student in the form of creativity. The reading of print is not a passive experience but an active one."24 A formula is a statistical device to assess the relative readability of written material, using an equation, a word list or a graph, to predict a grade level.

Choice of Variables

A review of the general formulas available provides further suggestions . . . there is little to be gained from choosing a highly complex formula. A simple 2-variable formula should be sufficient, especially if one of the variables is a semantic variable. Beyond these 2 variables, further additions add relatively little predictive validity . . .

Of these two linguistic variables, syllabication as a measure of word difficulty in opposition to a word list for college level readability, is discussed throughout this chapter. Sentence length, the other variable as a determinant of syntactic complexity, demands discussion.

Most longer sentences were not simply independent clauses connected with coordinate conjunctions. Most of those included noun modifiers, dependent clauses, nominalized verbs, deletions in coordinate clauses, appositives and clauses used as subjects. Most sentences that are long are syntactically complex.26

In a more recent study sentence length is further substantiated as a measure of semantic complexity.

Sentence length is perhaps the easiest, most apparent measure of syntactic complexity . . . long sentences contribute to the complexity of the reading material. The advent of the transformational-generative grammar movement in the late 1950's


26Susan Mandel Glazer, "Is Sentence Length a Valid Measure of Difficulty in Readability Formulas?", The Reading Teacher, XXVII (February, 1974), 467.
suggests that readability researchers, with their concern for sentence length, had been on the right track all along.\textsuperscript{27}

\textbf{Frequency of Samplings}

Recently Fitzgerald from the LaFayette Reading Academy investigated reliability based upon the recommended sample size of three. Since statistic standard error decreases as sample size increases, Fitzgerald suggested that a sample size of three is inadequate because the standard error of samples is frequently very high. She said that a sample size should be sufficiently large to approximate population means. She further warned us to "proceed with caution until further research on reliability of sample size has been conducted."\textsuperscript{28} In response to Fitzgerald's concern, Fry found that additional samples simply clustered around the mean and constituted nothing but additional unnecessary work.

When . . . programmed a computer to continuously sample every hundred words for a 20,000 word passage, . . . found that the readability scores tended to follow a normal distribution curve.\textsuperscript{29}

This study will use three samplings as has been directed by all of the formula builders.


\textsuperscript{29}Edward Fry "Fry's Readability Graph Clarification, Validity, and Extension to Level 17," \textit{Journal of Reading}, XXI (December, 1977), 246.
Rationale For Choice of Method:

Conducting a readability analysis of textbooks, regardless of the formula used, is tedious, time-consuming business.

Readability formulas can be grouped in four categories: a regression equation based upon multiple correlations; a regression equation based upon polysyllabic words; entirely dependent upon vocabulary; or plotted on a graph.

Within the first category, Lorge, Flesch and Dale-Chall all used a regression equation. They contributed to and built upon each other.

1) Lorge developed an extremely cumbersome formula, although it was seen as a simplification in 1939. He included such unwieldy variables as the number of prepositional phrases and the Dale list of 769 words. Even when Barker and Stokes changed the use of a word list to the use of word length, the formula was still dependent upon the unreliable factor of prepositional phrases. The Lorge formula is used very little currently, and would not be suitable with only one person serving as the counter as there would be no check for reliability in prepositional phrase count.

---

2) Flesch developed his formula for the express purpose of improving upon the Lorge formula by abolishing the prepositional phrase count. The Flesch formula human interest score using personal words was not designed for college textbook predictability as suitably as it was designed for newspaper and periodical predictability. For example, an anatomy textbook of extreme complexity may measure 0 on a human interest score. However, the reading ease score which uses two variables, word length in syllables and sentence length, is an excellent choice linguistically. A criticism made of the Flesch formula is that he used a his criterion the STANDARD TEST LESSONS IN READING.

The grade level of children answering test questions is not the best criterion for general readability, but . . . such data . . . (is all that is) . . . available today.

The complexity of his numerical computations manually make his formula an unsuitable choice for large numbers of books such as are designed in this study. However, the recent computerization of the Flesch formula makes it a viable choice.

3) Because Dale-Chall thought that a word list was a good method of predictability, they increased the word list to 3000 words. Word lists are relied upon as a measure of the likelihood of any student knowing any given word.

---

A function of the vocabulary used is certainly the most significant predictor of reading difficulty; however, word frequency, such as a word list, is more suitable for elementary reading. It is not a good criterion for college level reading. A college student is no longer a "sight reader."

The results of this pilot study failed to explain why some graduate students who have highly developed reading abilities have reported to me that they have difficulty comprehending the instructions (for form 1040). The answer to this apparent paradox may lie in the fact that the Dale-Chall readability formula is not definitive. Materials that contain subtle variations in meaning, vague and ambiguous words and phrases, and so forth may be more difficult than predicted by the Dale-Chall formula. 32

The Dale-Chall formula, with its dependence upon a word list, will not be suitable for complex nursing textbooks which contain specialized vocabularies.

4) Botel developed a technique whose only variable is vocabulary. Neither the semantic nor the polysyllabic value of the word is taken into consideration.

Studies have consistently reported that higher frequency words are recognized faster than lower frequency words . . . it seems sensible to assume that high frequency words do require less processing time. It is impossible, however, to assign a reason for that finding. Further, word recognition latency studies and visual duration threshold studies do not provide any information about a subject's understanding of the meaning of the word. 33

32 Fred Pyrczak, "Readability of Instructions for Form 1040," Journal of Reading, XX (November, 1976), 123.

A context which used "mixture and a context which used "amalgamation" would both tally as above third grade, since neither word is on his list of common words. The method of extrapolation used gives little consideration to semantics because it was expressly designed for primary grades where children do sight-read. Sight-reading means that when a child encounters a word to which he has not been previously exposed, he will not be able to "comprehend" it. He may be able to phonetically "sound it out", but still be unable to decipher any understanding of its semantic sense. When a college student encounters an unknown word, he will rely upon contextual clues to decipher its meaning; he will rely upon structural analysis to determine its probable meaning; he will rely upon his vast storehouse of experiences to form an association to give enough value to convey meaning. The Botel technique, based entirely upon a word list, will not be suitable for this study.

5) Spache designed his readability formula specifically for primary levels, and it is well respected in that area. He has reported a correlation of .95 between formula scores and grade level of primary books.\(^{34}\)

The Spache formula was designed to be checked against a word list which had been expanded several times: in 1956, in 1966, and again in 1974. It still is word list dependent and primary level specific.

None of the three leading formulas . . . the Flesch, the Lorge, or the Dale-Chall . . . is applicable to materials written for individuals reading on levels below Grade IV. The Spache formula is not suitable for college textbook material. It is most appropriate for primary grades and has contributed greatly to the concept of readability.

6) McLaughlin's SMOG index is a regression equation based upon polysyllabic count. His purpose was to develop a formula that was easier to calculate because it used one less constant than traditional formulas. It further eliminated the chore of multiplication completely. He agreed with Fry and Flesch in thinking that word length and sentence length are the two most reliable variables.

Fortunately there is no need to follow Flesch's system of counting every syllable in a passage in order to obtain a valid measure of its semantic difficulty. I have found a law relating the number of syllables in a passage to the percentage of polysyllabic words, defined as words of three or more syllables.

Additionally, McLaughlin eliminated the constant multiplier by making it equal to unity through the simple device of picking a suitable arbitrary number of sentences to be counted. The criticisms of the SMOG formula are: a) McLaughlin, like Lorge, Flesch and Dale-Chall validated grade levels with the STANDARD TEST LESSONS IN READING.

---


It has serious difficulties. Although I have proposed a procedure for obtaining a more valid criterion . . . the TEST LESSONS still provide the best criterion we have.37

b) McLaughlin relies upon the law relating the number of syllables to the percentage of polysyllabic words, yet he does not divulge the law nor its source. c) The standard error of the predictions the SMOG formula can give is about 1.5 grades which means it is less accurate than the predictions given by Flesch in his regression formula or the Fry graph. The SMOG formula will not be used in this study.

7) Fry is aware that the regression formulas relied upon arbitrary conversion tables or charts in order to yield a grade level. His grade levels are determined by plotting on a graph the actual computations of real textbooks which publishers market to educational institutions with a published grade level. On the other hand, this is one of the criticisms leveled at Fry.

Fry based his reliability upon the belief that texts on whose readability he built his graph do in fact have the readability which the publishers say they do.38

Fry agrees that a problem does exist.

Hence the problem of validity is compounded by trying to determine grade levels when grade levels won't stand still.39


Fry believes that the computation of grade levels by graph form is superior to validation with the STANDARD TEST LESSONS IN READING. The problem that exists is not in the form of graph plotting but in the form of grade level itself.

I openly confess to not having any data about the difference between thirteenth through sixteenth grade material. I do hope someone will gather some for validation. . . . Part of the difficulty in determining college norms is that college populations have wide divergences in academic qualifications of students. College reading ability also tends to become very subject specific. That means that what may be normal reading for a physics student could be more difficult for a philosophy student and vice versa. These variables are all in addition to the readability principle that high motivation overcomes high readability level, but low motivation demands a low readability level.

If the Fry formula were computerized it would be a viable choice.

Computerization

The computerized version of the Flesch index to produce a grade level was developed by General Motors Corporation and is made available through their public relation staff as a tool to improve communication. This solution to computerization was titled the Simple Test Approach for Readability and is known by its acronym, STAR. STAR is generously shared with institutions who request it in the form of a booklet with computer programs for two languages, both FORTRAN IV and BASIC.

---

40 Edward Fry, "Fry's Readability Graph Clarifications, Validity, and Extension to Level 17," Journal of Reading, XXI (December, 1977), 251.
The adaptation of the STAR program to the San Bernardino Valley College Digital Equipment PDP 11/45 was done under the direction of Mr. Henry James, head of the Computer Science Department. His adaptation not only included using the FORTRAN IV language, but he altered the output to go to line printing instead of a terminal.

The STAR program, based upon the Flesch formula using syllables and sentence length as the variables with a three sample sampling, will be used in this study.
RESULTS OF COMPUTERIZED READABILITY ASSESSMENT

There are nine textbooks used in this study. Additional paperbacks and manuals are used in the R.N. program, but they are subject to change in accordance with the times and in keeping with current educational needs. These nine textbooks remain the mainstay of the program. Even though this is a four semester program, it must not be assumed that two or three books are used per semester with ascending orders of difficulty. These nine textbooks bear assigned readings throughout the entire four semester program.

RESULTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>GRADE LEVEL</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>author</td>
<td>grade level</td>
</tr>
<tr>
<td></td>
<td>publisher</td>
<td>years, months</td>
</tr>
<tr>
<td></td>
<td>copyright date</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>years, months</td>
</tr>
<tr>
<td>MEDICAL SURGICAL NURSING</td>
<td>Brunner &amp; Suddarth</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>Lippincott: Philadelphia</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>1980, 4th Edition</td>
<td>15.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14.6</td>
</tr>
<tr>
<td>CLINICAL PRACTICE OF MEDICAL-SURGICAL NURSING</td>
<td>Beyers &amp; Dudas</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>Little, Brown: Boston</td>
<td>15.3</td>
</tr>
<tr>
<td></td>
<td>1977</td>
<td>15.1</td>
</tr>
<tr>
<td>OBSTETRIC NURSING</td>
<td>Olds, London, Ladewig &amp; Davidson</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>Addison Wesley: Reading</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>1980</td>
<td>16.0</td>
</tr>
</tbody>
</table>

(see appendix 1 for sample printout)
<table>
<thead>
<tr>
<th>TITLE</th>
<th>GRADE LEVEL</th>
<th>AVERAGE GRADE LEVEL OF ENTIRE TEXTBOOK LOAD:</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUIDE TO DIAGNOSTIC PROCEDURES</td>
<td>18.5</td>
<td>15.5</td>
</tr>
<tr>
<td>French</td>
<td>15.5</td>
<td>17.5</td>
</tr>
<tr>
<td>McGraw-Hill: New York</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>BASIC NURSING: A PSYCHOPHYSIOLOGIC APPROACH</td>
<td>13.2</td>
<td>13.7</td>
</tr>
<tr>
<td>Sorensen and Luckmann</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>W. B. Saunders: Philadelphia</td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>NURSING CARE OF INFANTS AND CHILDREN</td>
<td>9.8</td>
<td>13.4</td>
</tr>
<tr>
<td>Whaley and Wong</td>
<td>15.2</td>
<td></td>
</tr>
<tr>
<td>C. V. Mosby: St. Louis</td>
<td>15.1</td>
<td></td>
</tr>
<tr>
<td>PATHO-PHYSIOLOGY: CLINICAL CONCEPTS OF DISEASE</td>
<td>20.2</td>
<td>16.5</td>
</tr>
<tr>
<td>PROCESS</td>
<td>13.8</td>
<td></td>
</tr>
<tr>
<td>Anderson-Price &amp; McCarty-Wilson</td>
<td>15.4</td>
<td></td>
</tr>
<tr>
<td>PHARMACOLOGY AND DRUG THERAPY IN NURSING</td>
<td>19.6</td>
<td>17.8</td>
</tr>
<tr>
<td>Rodman &amp; Smith</td>
<td>16.0</td>
<td></td>
</tr>
<tr>
<td>Lippincott: Philadelphia</td>
<td>17.6</td>
<td></td>
</tr>
<tr>
<td>1979 2nd edition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSYCHIATRIC NURSING</td>
<td>19.6</td>
<td>14.8</td>
</tr>
<tr>
<td>Wilson &amp; Kneisl</td>
<td>12.8</td>
<td></td>
</tr>
<tr>
<td>Addison-Wesley: Reading</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>1973</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVERAGE GRADE LEVEL OF ENTIRE TEXTBOOK LOAD:</td>
<td></td>
<td>15.5</td>
</tr>
</tbody>
</table>
SUMMARY

Discussion

Pyrczak informs us that one problem with measuring reading ability via grade level is that a student who can create a 12.0 grade level equivalency is not necessarily capable of doing 12th grade work. In the case of the Nelson/Denny test it contains material which spans from 6.0 grade to 15.3 grade. The requirement for acceptance into the registered nursing program is established by the college board at 12.0 effective in the spring semester, 1982. We are faced with the recognition that this does not assure us that the student, in reality, will be able to function at 12.0 grade reading level.

We are aware that no readability formula can accurately measure the grade level of textbooks because of an inability to scientifically weigh abstractness, specific vocabulary and numerous other factors.

Yet even with these two above points in mind, we are faced with the recognition that students entering the registered nursing program are going to be required to score at least 12.0 grade level on the Nelson/Denny or an equivalent reading test. We are further faced with the recognition that this readability assessment discovered that the entire range
of textbook difficulty spans from 9.8 grade level to 20.2 grade level with the average difficulty of the core textbooks in the program at 15.5 grade level.

Recommendations

Some method of assistance is quite likely to be needed by spring semester, 1982. The recommendation based upon this study is that the Reading Department at San Bernardino Valley College write a proposal for a new course to be qualified by the curriculum committee in the fall of 1981 in order to be ready for the spring semester.

The logistics of the class will include these factors. The course will be listed within the current numbering system at the non-transfer level as it will be designed to be taken concomitantly with a transfer level class. A prerequisite for registration into this proposed reading class will be registration in a transfer class in another discipline which requires a standard textbook reading load. The class need not be limited only to nursing students. The class will be an ADA-generating class of three hours per week with one hour of lab time to be arranged.

The objectives of the class will include the following: to enable the student to perceive the organization of the particular textbook being used; to be on time with assigned readings thereby avoiding anxiety; to work within the confines of each specific vocabulary; to utilize appropriate study skills such as outlining and notetaking with each particular
textbook; to learn and utilize some technique in approaching textbook reading such as the PQ3R method; and, to have an instructor model "good" reading methods and habits to be assimilated by the student.

Certain qualifications will be necessary to assure maximum gain: there will be a limitation of one textbook per student to be used throughout the semester; class size will be limited to 25 or less in keeping with the individual, remedial needs which may be encountered; the grades will be based on a credit/no-credit basis; the instructor will work in conjunction with the counseling staff for students with severe emotional problems and for students who find themselves in "way over their heads", and lastly, a committee will be established by the Reading Department to gain information from the numerous other colleges who do already offer such a course.
BIBLIOGRAPHY


Bradly, John, "Using Readability to Improve the Content Validity of Information Placement Tests," Reading Improvement, XIII (Fall 1976), 182-187.

Britton, G. & Lumpkin, M., "Computerized Readability Verification of Textbook Reading Levels," Reading Improvement, XIV (Fall 1977), 193-199.


Dale, Edgar, "Develop Critical Reading," Reading Improvement, XIII (Spring 1976), 30-33.


Glazer, Susan Mandel, "Is Sentence Length a Valid Measure of Difficulty in Readability Formulas?", The Reading Teacher, XXVII (February, 1974), 464-468.


Hancock, Debora, "Identifying the Reading Needs of the College Student," Claremont Reading Conference, XLII (1978), 210-18.


APPENDIX 1

Sample Printout
THE MOST COMMON ABNORMAL PHYSICAL FINDING IN NEONATES IS JAUNDICE (ICTERUS). JAUNDICE DEVELOPS FROM DEPOSIT OF THE YELLOW PIGMENT, BILIRUBIN, IN TISSUES. UNCONJUGATED (INDIRECT) BILIRUBIN IS A BREAKDOWN PRODUCT DERIVED FROM HEMOGLOBIN THAT IS RELEASED FROM LYSED RED BLOOD CELLS AND HEME pigMENTS FOUND IN CELL ELEMENTS (NONERYTHROCYTE BILIRUBIN). FETAL UNCONJUGATED BILIRUBIN IS NORMALLY CLEARED BY THE PLACENTA IN UTERO, SO TOTAL BILIRUBIN AT BIRTH IS USUALLY LESS THAN 3 MG/100 ML UNLESS AN ABNORMAL HEMOLYTIC PROCESS HAS BEEN PRESENT. POSTNATALLY, THE INFANT MUST CONJUGATE BILIRUBIN IN HIS LIVER, PRODUCING A RISE IN SERUM BILIRUBIN IN THE FIRST FEW DAYS OF LIFE.