A System for Scoring Proverb Interpretations Provided by Non–Brain-Damaged Adults and Aphasic Adults

Shelley B. Brundage and Robert H. Brookshire

Abstract thinking, which includes the ability to reason, generate inferences, and solve problems, is often impaired after brain damage. Proverb interpretation often is used to assess the effects of brain damage on abstract thinking (Erickson & Binder, 1986; Lezak, 1983). The *Random House College Dictionary* defines proverbs as "short, popular sayings that express . . . some commonplace truth or useful thought . . . requiring interpretation" (Stein, 1980). Proverb interpretation tasks are found in psychological tests (e.g., *The Wechsler Adult Intelligence Scale–Revised*, Wechsler, 1976) and speech and language batteries (e.g., Burns, Halper, & Mogil, 1985).

Most of the information on proverb interpretation comes from studies of schizophrenic adults. Speech–language pathologists use proverb interpretation informally during assessment, but their choice of proverbs and their methods for analyzing responses usually are nonstandardized and arbitrary. Few studies have been done of the reliability of scoring proverb interpretations of either brain-damaged or non–brain-damaged adults. However, the scoring of proverb interpretations appears to be problematic. Some investigators have found acceptable levels of interrater reliability among psychiatric professionals in scoring the proverb interpretations of schizophrenic, depressive, manic, and normal subjects (Reich, 1981), whereas others have reported poor interrater reliability (Andreasen, 1977; Burgos, 1986).

Investigators disagree about what to measure when assessing proverb interpretations. Some try to place responses on an abstract–concrete continuum (Gorham, 1956; Reich, 1981), whereas others suggest that

literalness should be the variable of interest (Hertler, Chapman, & Chapman, 1978). This disagreement about what to measure may arise from failure to address proverb-related and subject-related variables that may affect proverb interpretations produced by either brain-damaged or non–brain-damaged adults. Proverbs differ in familiarity, abstractness, syntactic complexity, and number of words (Cunningham, Ridley, & Campbell, 1987; Nippold, Martin, & Erskine, 1988), and these variables may influence how easy or difficult a proverb is to interpret. A subject's characteristics, such as age and education (Penn, Jacob, & Brown, 1988), and psychological state at the time of testing (Sander & Greenberg, 1968) also are thought to influence proverb interpretation.

The purpose of this paper is to describe a system for scoring proverb interpretations, and to compare the proverb interpretation performance of non–brain-damaged adults with that of aphasic adults, using this scoring system.

METHOD

Subjects

Two groups of subjects participated in the study. The non–brain-damaged (NBD) group consisted of 10 subjects (3 women and 7 men) aged 50 to 73 years ($M = 64.9$, $SD = 7.76$). All were independently living, Minnesota residents, with no reported history of neurologic deficit. Their scores on the *Mini-Mental State* examination (Folstein, Folstein, & McHugh, 1975) ranged from 27 to 30 points, and were within the normal range. The mean number of years of education for this group was 11.8 (range = 8 to 16). A demographically based estimate of premorbid IQ was obtained for each subject using the method of Barona, Reynolds, and Chastain (1984). The estimated IQs of the NBD group ranged from 91 to 119.

The aphasic (APH) group consisted of 10 subjects (3 women and 7 men) aged 59 to 79 years ($M = 69.1$, $SD = 9.28$). All had sustained a single, left hemisphere, thromboembolic cerebrovascular accident, and all were recruited from speech clinics at two Minneapolis hospitals. Time postonset ranged from 2 to 64 months. Three of the aphasic subjects exhibited mild to moderate fluent-mixed aphasia, characterized by fluent speech and the presence of verbal and literal paraphasias. Four subjects exhibited mild anomic aphasia. The remaining 3 subjects exhibited mild to moderate nonfluent aphasia. The mean number of

years of education for the aphasic group was 12.3 years (range = 10 to 16). The estimated IQs of the aphasic group ranged from 92 to 119.

Stimulus Selection

Twenty-four proverbs, with differing levels of familiarity, abstractness, and syntactic complexity, were used in this study (seeAppendix A for examples). To determine the level of familiarity and abstractness for each proverb, 10 speech–language pathologists rated 100 proverbs on these two dimensions. First, they rated each proverb's level of abstractness on a 5-point scale. Then they rated each proverb's familiarity as either high or low, depending on whether they had heard it before. To judge the level of syntactic complexity, rules were written based in part on the work of Schulte and Brandt (1989). Proverbs were considered high in syntactic complexity if they contained negatives, comparatives, temporal markers, or conjunctions. The investigator used the rules to divide the proverbs into high and low syntactic complexity groups. Another speech pathologist independently confirmed these judgments using the syntax rules. The proverbs used in this study represent those at the extremes of the abstractness scale, and those with the highest proportions of "familiar" and "unfamiliar" ratings.

Procedure

Subjects were tested in a quiet room. The investigator presented a card with a proverb printed on it, said the proverb aloud, and then asked the subject to "tell me what it means." After the subject interpreted each proverb, he or she was asked whether it was familiar. All proverb interpretations were audiotaped and transcribed orthographically.

Scoring System

A categorical scoring system, similar to Nippold and Martin's (1989) system for scoring idiom interpretations, was created to assess the quality of each interpretation. Rules were written to allow independent judges to place proverb interpretations into one of five categories. A score of *unrelated* was given to responses that had no apparent connection with the proverb. *Rejection* was scored when a subject refused to attempt an interpretation. Responses were scored as *related* if they were based on the words in the proverb, but ignored the proverb's

abstract meaning, or if they were verbatim repetitions or general comments about proverbs. A score of *abstract/inadequate* was given to interpretations that were abstract, but were inappropriate interpretations of the proverb's abstract meaning. A score of *abstract/adequate* was given to interpretations that conveyed the proverb's abstract meaning (see Appendix B for examples).

Three judges independently scored the transcripts of 6 subjects (4 NBD and 2 APH). Point-to-point percentage agreement ranged from 80% to 90% across all subjects.

RESULTS

For the purposes of this paper, abstract/adequate scores were labeled *adequate,* and abstract/inadequate scores were labeled *inadequate.* The related, unrelated, and rejection categories were combined into a single *error* category. To produce a difficulty rating for each proverb, each subject's performance was evaluated using a scoring system in which these scoring categories were given numerical values of 2, 1, and 0, respectively. This yielded a total possible score of 20 points per group for each proverb. Labovitz (1970) showed that treating ordinal measures as interval ones has negligible negative effects on common statistical procedures.

Adequacy of Responses

The mean adequacy score for the NBD group across all proverbs was 14.95 (SD = 4.38), and the mean adequacy score for the APH group was 8.70 (SD = 4.71). A t-test comparing the scores of the NBD and APH groups was significant (t = −9.46, p = .00), suggesting that the NBD group produced more adequate responses than the APH group.

To determine if the order of proverb difficulty was similar for the NBD and the APH groups, the proverbs were arranged so that the NBD group's mean scores for individual proverbs were in descending order. Then the performance of the APH group was compared with that of the NBD group (Figure 1). The APH group received lower average scores on each proverb than the NBD group. The APH responses were considerably more variable than the NBD responses. However, the slope of the lines for the two groups was similar, and a Pearson correlation coefficient calculated between the NBD and the APH subjects' average scores on the 24 proverbs yielded r = .76, suggesting that proverbs that were difficult for the NBD group were also likely to be difficult
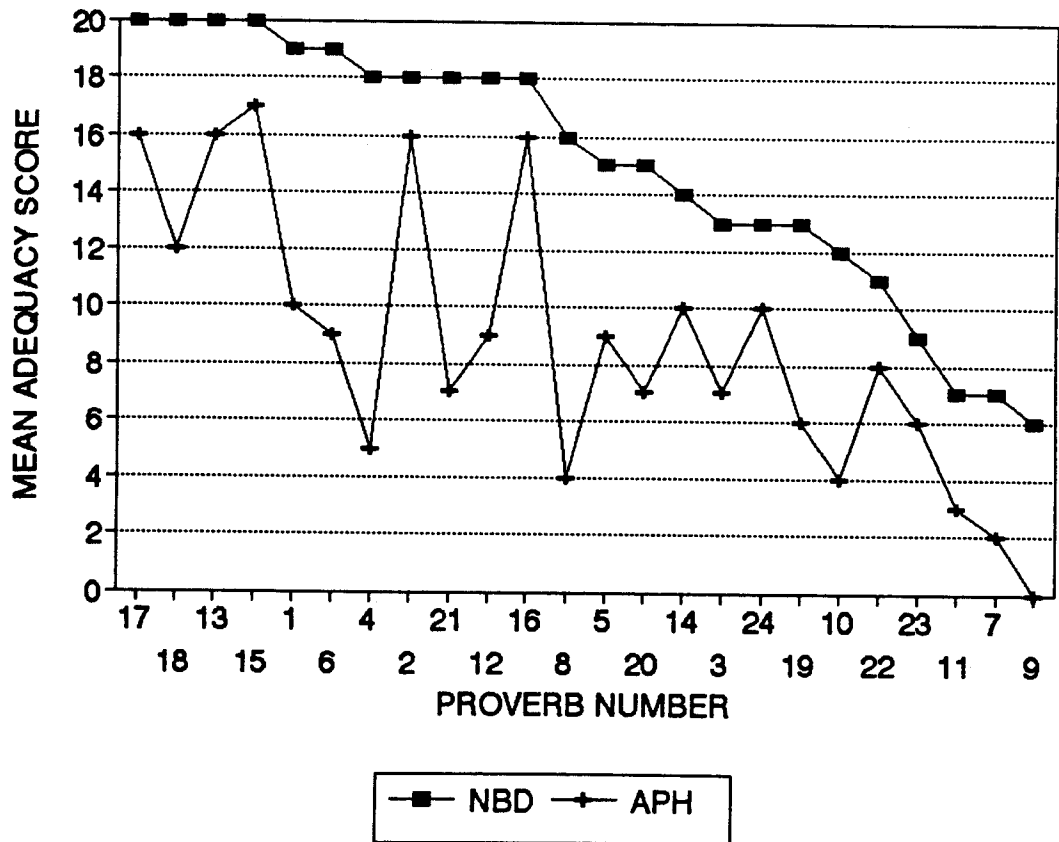
Figure 1. Proverb adequacy scores of the non–brain-damaged (NBD) and aphasic (APH) groups, arranged in descending order of adequacy for NBD subjects.

for the APH group. However, although this correlation is statistically significant ($p < .05$), it accounts for only about 58% of the variance between the two groups.

Effects of Familiarity

Each group's mean scores for proverbs with high levels of familiarity and those with low levels of familiarity are given in Figure 2. We calculated t-tests to determine if familiarity affected the proverb interpretation scores of each group. The differences between familiar and unfamiliar proverbs were significant for both groups (NBD: $t = 4.30$, $p = .001$; APH: $t = 4.03$, $p = .002$). For both groups, familiar proverbs were easier to interpret than unfamiliar proverbs (Figure 2).
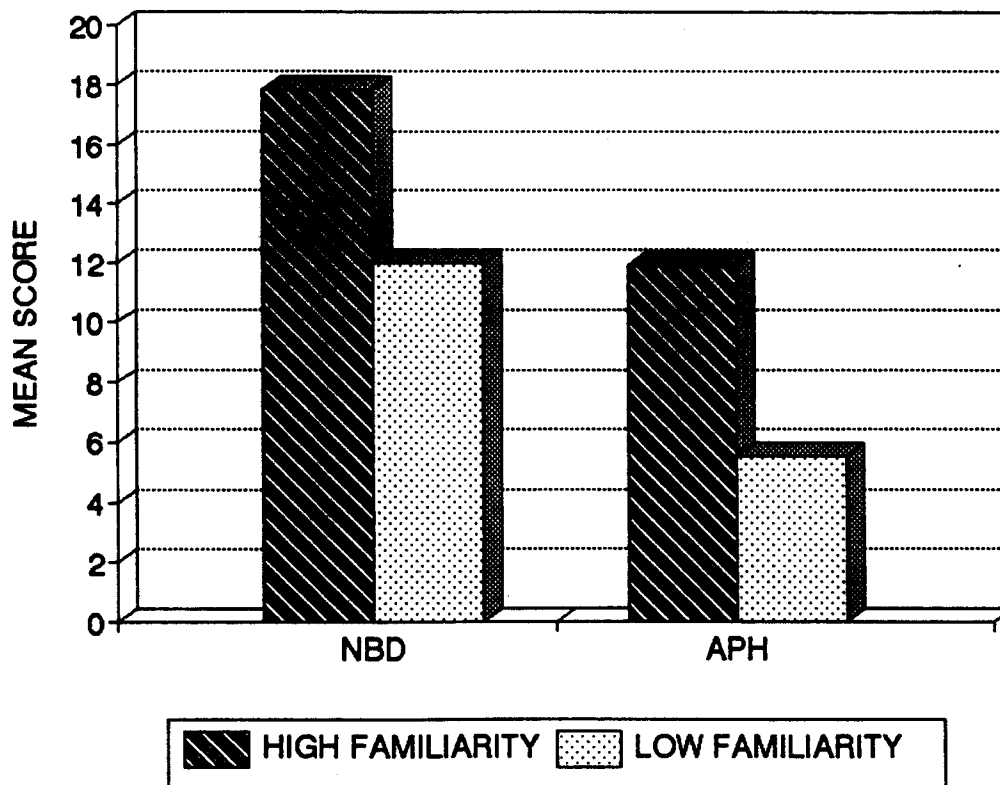
Figure 2. Effects of proverb familiarity on the adequacy scores of the non–brain-damaged (NBD) and aphasic (APH) groups.

Effects of Abstractness

Each group's mean scores for proverbs with high levels of abstractness and those with low levels of abstractness are given in Figure 3. A t-test comparing scores on proverbs with high and low levels of abstractness was not significant for the NBD group ($t = -1.10, p = .29$), but an equivalent t-test for the APH group yielded a significant result ($t = -2.53, p = .02$). These results suggest that the APH group performed significantly better on proverbs with low levels of abstractness, but that the NBD group was not significantly affected by the abstractness of the proverbs (Figure 3).

Effects of Syntactic Complexity

Each group's mean scores for proverbs with high levels of syntactic complexity and those with low levels of syntactic complexity are given
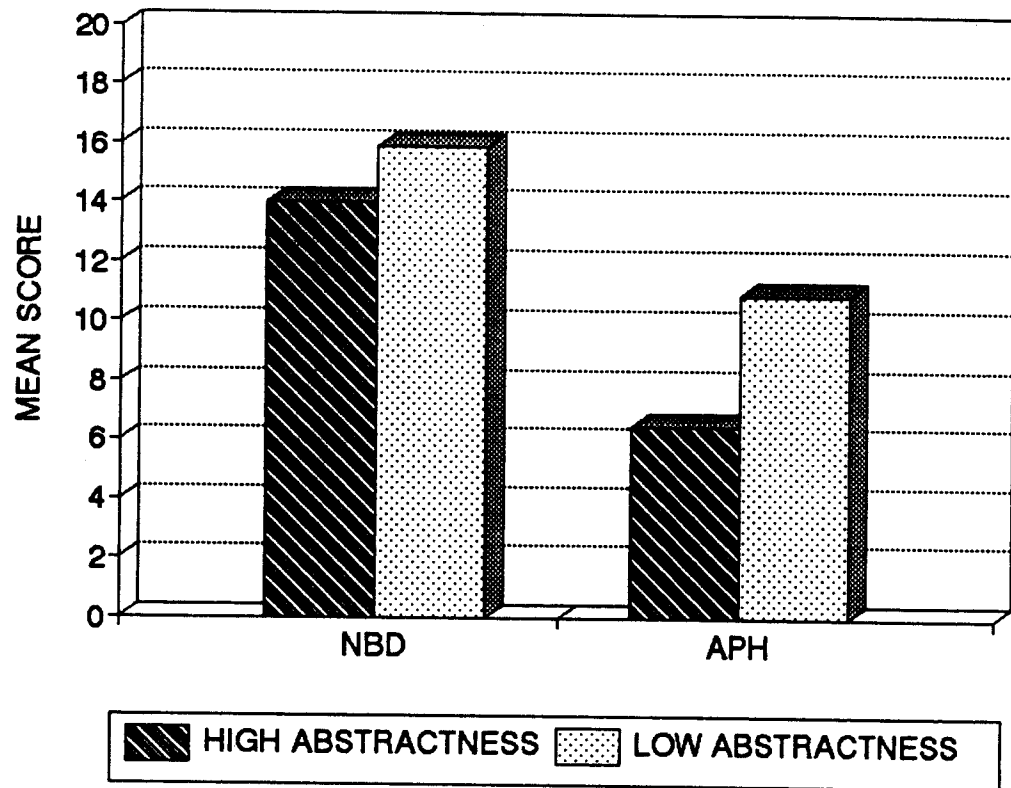
Figure 3. Effects of proverb abstractness on the adequacy scores of the non–brain-damaged (NBD) and aphasic (APH) groups.

in Figure 4. Dependent-measure t-tests comparing differences between high and low levels of syntactic complexity were calculated for each group (NBD and APH). Neither comparison yielded a significant t value (NBD: $t = -.05$, $p = .92$; APH: $t = -.20$, $p = .83$). The levels of syntactic complexity studied did not significantly affect the adequacy of proverb interpretations for either the NBD or the APH group.

DISCUSSION

Group Effects

The non–brain-damaged and aphasic subjects in this study responded similarly to the proverb interpretation task. Although the aphasic group's scores were always lower and more variable, the effects of
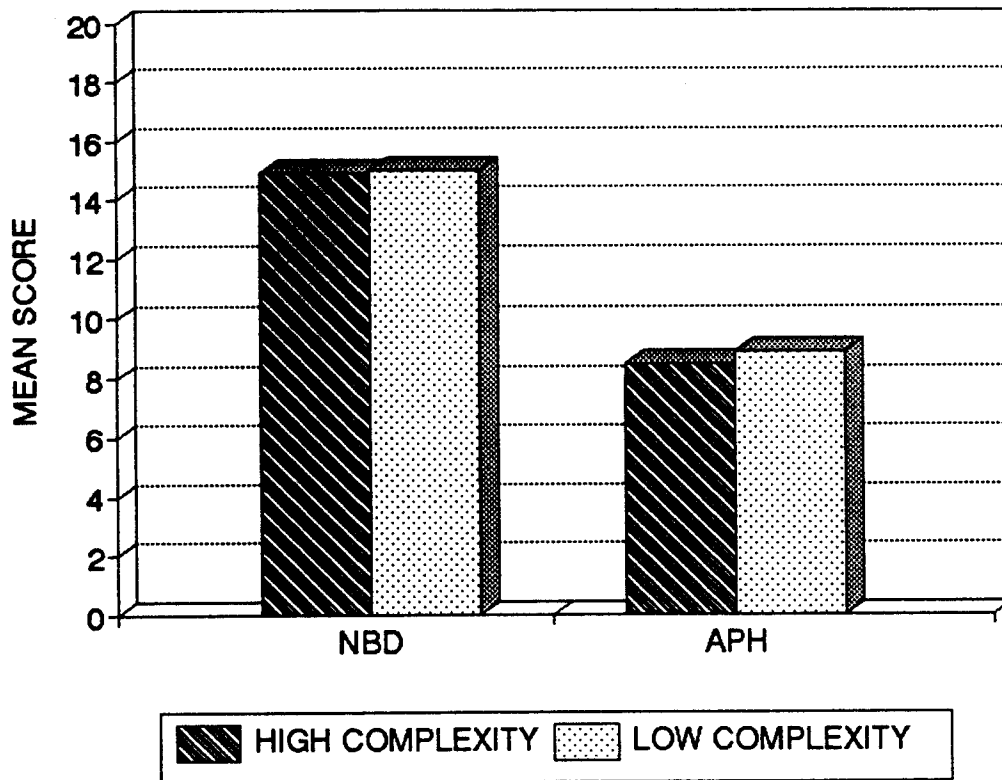
Figure 4. Effects of syntactic complexity on the adequacy scores of the non–brain-damaged (NBD) and aphasic (APH) groups.

familiarity and syntactic complexity on their performance were similar to those for the non–brain-damaged (NBD) subjects. When the NBD group was affected, so was the aphasic group, and in the same direction; when the NBD group was not affected, neither was the aphasic group. This suggests that both groups may be using the same processes to decode proverbs and produce proverb interpretations, and that proverbs normed on control subjects should be valid for use with aphasic subjects.

Of the three conditions studied, familiarity had by far the largest effect on proverb interpretation for both groups, with familiar proverbs being easier to interpret than unfamiliar ones. The abstractness of proverbs affected the aphasic group more than the NBD group. Abstract proverbs had a pronounced negative effect on the aphasic group's scores, whereas abstractness did not affect the NBD group's scores. Syntactic complexity, at least at the levels considered in this study, did not affect the proverb interpretations of either group.

The results of this study suggest that one can ignore syntactic complexity when selecting proverbs for use in assessing or treating aphasic

adults, as long as unusually complex structures, such as embedded clauses, are avoided. However, clinicians should be cognizant of a proverb's familiarity and level of abstractness when choosing proverbs for use with aphasic adults. When proverbs were both unfamiliar and abstract, the performance of even the mildly impaired aphasic subjects in this study was near chance levels. Consequently, one may wish to use only familiar proverbs and manipulate their abstractness to assess the abstraction abilities of aphasic adults.

Methodological Considerations

The scoring system used in this study appears to be a clinically useful tool that is relatively easy to learn to use. However, there is a potential problem with the *related* category. Because this category contains both verbatim repetitions and literal responses, it is difficult to describe what the subjects were doing when they failed to provide an abstract interpretation. Further refinements of the scoring system may be needed to characterize a subject's change over time, from merely repeating a proverb to attempting an interpretation, albeit a literal one.

Another potential methodologic problem is that the ratings of proverb familiarity and abstractness were made by speech–language pathologists, and not by age- and education-matched peers of the NBD and aphasic subjects. The familiarity and abstractness ratings of these three groups might differ. We did not ask our subjects to rate the abstractness of each proverb, but we did ask them if they had heard each one before. The percent agreement between the familiarity ratings of the NBD group and the speech–language pathologists was 80% or above on 21 of the 24 proverbs (88%). The percent agreement between the aphasic group and the speech–language pathologists was lower, but this is not surprising due to the language impairments of the aphasic subjects.

Finally, there is the issue of what a proverb interpretation task measures. Proverb interpretation is a highly complex task, involving many *hypothesized cognitive processes*. The task's complexity may preclude knowing exactly what happens inside the subject's head when attempting to interpret a proverb. Failure to provide an adequate interpretation may occur for any number of different reasons—for example, the subject may fail to recognize the proverb, may fail to decipher the proverb even though it is recognized, or may have speech production difficulties that interfere with the interpretation of an accurately recognized and deciphered proverb. Conversely, providing an adequate proverb interpretation does not in itself prove that the subject had heard the proverb before. Subjects in this study provided adequate inter-

pretations of proverbs that were not familiar to them; they also provided inadequate interpretations of proverbs that they *had* heard before.

ACKNOWLEDGMENTS

REFERENCES

Andreasen, N. (1977). Reliability and validity of proverb interpretation to assess mental status. *Comprehensive Psychiatry, 18*, 465–472.

Barona, A., Reynolds, C., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52*, 885–887.

Burgos, L. (1986). Inter-rater reliability and the use of proverb interpretation in the detection of disordered thinking. *Dissertation Abstracts International, 46*(7), 2453B.

Burns, M., Halper, A., & Mogil, S. (1985). *RIC evaluation of communication problems in right hemisphere dysfunction*. Rockville, MD: Aspen Systems.

Cunningham, D., Ridley, S., & Campbell, A. (1987). Relationship between proverb familiarity and proverb interpretation: Implications for clinical practice. *Psychological Reports, 60*, 895–898.

Erickson, R., & Binder, L. (1986). Cognitive deficits among functionally psychotic patients: A rehabilitative perspective. *Journal of Clinical and Experimental Neuropsychology, 8*, 257–274.

Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Gorham, D. (1956). A proverbs test for clinical and experimental use. *Psychological Reports, 2*, 1–12.

Hertler, C., Chapman, L., & Chapman, J. (1978). A scoring manual for literalness in proverb interpretation. *Journal of Consulting and Clinical Psychology, 46*, 551–555.

Labovitz, S. (1970). The assignment of numbers to rank-order categories. *American Sociological Review*, pp. 515–524.

Lezak, M. (1983). *Neuropsychological assessment*. New York: Oxford University Press.

Nippold, M., & Martin, S. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech and Hearing Research, 32*, 59–66.

Nippold, M., Martin, S., & Erskine, B. (1988). Proverb comprehension in context: A developmental study with children and adolescents. *Journal of Speech and Hearing Research, 31,* 19–28.

Penn, N., Jacob, T., & Brown, M. (1988). Familiarity with proverbs and performance of a black population on Gorham's Proverbs Test. *Perceptual and Motor Skills, 66,* 847–854.

Reich, J. (1981). Proverbs and the modern mental status exam. *Comprehensive Psychiatry, 22,* 528–531.

Sander, F., & Greenberg, H. (1968). A proverbial excursion: On the hazards of administering the proverbs to test the capacity to abstract. *Psychiatric Quarterly, 42,* 696–697.

Schulte, E., & Brandt, S. (1989). Auditory verbal comprehension impairment. In C. Code (Ed.), *The characteristics of aphasia.* London: Taylor and Francis.

Stein, J. (1980). *Random House college dictionary.* New York: Random House.

Wechsler, D. (1976). *Wechsler Adult Intelligence Scale–Revised.* San Antonio, TX: Psychological Corporation.

APPENDIX A: EXAMPLES OF PROVERBS WITH DIFFERING LEVELS OF FAMILIARITY, ABSTRACTNESS, AND SYNTACTIC COMPLEXITY

Proverb	Familiarity	Abstractness	Syntactic Complexity
Blood is thicker than water.	High	High	High
The squeaky wheel gets the oil.	High	High	Low
Two heads are better than one.	High	Low	High
The early bird catches the worm.	High	Low	Low
The hot coal burns, the cold one blackens.	Low	High	High
The bread of strangers can be very hard.	Low	High	Low
The rich never lack relatives.	Low	Low	High
Wild colts make good horses.	Low	Low	Low

APPENDIX B:
SCORING EXAMPLES FOR THE PROVERB, "BLOOD IS THICKER THAN WATER"

Scoring Category	Response
Unrelated	I had a toolbox once.
Related	Blood has to be thicker than water. It's water plus blood cells.
Abstract/Inadequate	That means that man is part of the animal kingdom.
Abstract/Adequate	People feel stronger attachments to their families than they do to people who are not related to them.