Psychometric Aspects of Verbal Pragmatic Ratings

Ronald L. Bloom,* Lawrence H. Pick,†‡ Joan C. Borod,†‡§ Kashemi D. Rorie,§† Fani Andelman,** Loraine K. Obler,‡ Martin Sliwinski,‡‡ Alfonso L. Campbell,† James R. Tweedy,‡‡ and Joan Welkowitz¶¶

*Hofstra University; †Queens College of the City University of New York; ‡City University of New York Graduate School; §Mount Sinai School of Medicine; ¶Wake Forest University School of Medicine; ‡Howard University; **Sourasky Medical Center of Tel Aviv; ‡‡Albert Einstein College of Medicine; ‡‡‡Yale University School of Medicine; and ¶¶New York University

This study examined the psychometric aspects of a verbal pragmatic rating scale. The scale contained six pragmatic features (i.e., Conciseness, Lexical Selection, Quantity, Relevancy, Specificity, and Topic Maintenance) based on Grice’s cooperative principles. Fifteen right brain-damaged (RBD), 15 left brain-damaged (LBD), and 16 healthy normal control (NC) right-handed adult participants produced narratives while recollecting emotional and nonemotional experiences. Naive raters evaluated each pragmatic feature for appropriateness on a 5-point Likert scale. When reliability was examined, the overall internal consistency of the pragmatic scale was extremely high (α = .96). Factor analysis was conducted to examine the theoretical relations among the six pragmatic features. Three meaningful factors involving discourse content, conceptual unity, and parsimony were identified. Findings are discussed in light of Grice’s model and the construct validity of the scale.

INTRODUCTION

Ecologically valid evaluation and treatment methods require that the set of rules governing the use of language in context be considered along with

This research was supported, in part, by a Hofstra University Research and Development Grant to R. Bloom and by NIMH Grants Nos. MH42172 and MH44889 and Professional Staff Congress-CUNY Research Awards 662430 and 664078 to Queens College. Portions of this research were carried out as part of dissertation research for the Ph.D. by F. Andelman at Queens College and the Graduate School of CUNY and for the M.S. by K. Rorie at Howard University. A portion of this study was presented at the annual meeting of the American Speech-Language Hearing Association, Seattle, WA, November, 1996. A report on patient and control group differences on valence and emotionality measures of pragmatic appropriateness from the same database appears in Borod et al. (in press). We are grateful to Dr. B. Stemmer and the reviewers for their helpful comments.

Address correspondence and reprint requests to Ronald L. Bloom, Speech-Language-Hearing Sciences, Hofstra University, 106 Davidson Hall, Hempstead, NY 11549. E-mail: SPHRLB@hofstra.edu.

0093-934X/99 $30.00
Copyright © 1999 by Academic Press
All rights of reproduction in any form reserved.
a description of semantics and syntax. At present, analysis of discourse or language processing beyond the word and sentence levels is routinely incorporated into the assessment and treatment of language-based communication disorders. Pragmatics is a general approach to discourse analysis that focuses on the interaction among meaning, context, and communication (Schiffrin, 1994). In this view, successful communication of a spoken message is crucially linked to the listener’s interpretation of the speaker’s intention and the situational setting or context. However, the nature of pragmatics with its interacting components makes it particularly challenging to operationalize for research and clinical purposes.

Within recent years, several measures of pragmatics (e.g., Damico, 1992; Penn, 1988; Prutting & Kirchner, 1987) have been developed as guidelines for clinical observations. Gricean pragmatics has provided a theoretically meaningful framework for examining the discourse of individuals with communication disorders. Grice (1975) proposed that communication is governed by four principles of cooperation or rules between a listener and a speaker: Relevancy (i.e., the participant’s contribution should be applicable), Manner (i.e., the contribution should be direct and unambiguous), Quantity (i.e., the contribution should be as informative as required), and Quality (i.e., the contribution should be truthful and supported with evidence). Central to Grice’s theory is the view that listeners actively interpret what they hear as if it conforms to these principles, although participants do not strictly follow these rules in every communication situation.

The Pragmatic Protocol developed by Prutting and Kirchner (1987) applies Grice’s theory to describe language use in individuals with communication disorders. Consistent with Grice’s cooperative principle, Prutting and Kirchner note that pragmatic features are continuous throughout discourse and are derived from the listener’s perception of a speaker’s performance in a conversation. The Pragmatic Protocol has been widely used to identify the appropriateness of 30 general pragmatic abilities thought to be integral to communicative competence. The Pragmatic Protocol includes observation of three interacting communication modalities: verbal behaviors (e.g., message specificity, cohesion, topic selection, topic initiation, and topic maintenance), paralinguistic behaviors (e.g., prosody, vocal quality, and intelligibility), and nonverbal behaviors (e.g., facial expression, eye gaze, and proximity of the speaker to the listener). According to Prutting and Kirchner, these aspects of pragmatics are to be observed in a natural setting, with scoring focused on the success of the communicative interaction between the speaker and the listener rather than on the conventional linguistic expression of the message.

Prutting and Kirchner developed the Pragmatic Protocol to observe systematically how communication breaks down in different clinical populations. When used to examine patients with aphasia, it supported the view that aphasia was a deficit in linguistic retrieval, not in communicative competence. For example, in the face of linguistic deficits, individuals with fluent,
nonfluent, and mixed aphasia were generally found to be pragmatically appropriate. Specifically, 82% of the pragmatic behaviors of the left-brain-damaged (aphasic) group (LBDs) were judged to be pragmatically appropriate. However, the communication of the LBDs was characterized by lengthy pauses, a decrease in the amount of information conveyed, and inappropriate word selection. When Prutting and Kirchner examined the pragmatic abilities of right-brain-damaged individuals (RBDs), 84% of their pragmatic behaviors were judged to be appropriate. The RBDs displayed difficulty with topic maintenance, informativeness, eye gaze, and prosody.

In summary, Prutting and Kirchner’s findings indicated that brain-damaged patients show some strengths in pragmatic abilities, regardless of hemispheric side of damage (left vs right) or the site of lesion (anterior vs posterior). Additionally, they reported that some measurable degree of pragmatic impairment was apparent in both groups of brain-damaged patients. They further noted that RBDs’ difficulties were most noticed in the nonverbal domain, whereas the pragmatic difficulties of the LBDs appeared to be more verbal in nature.

Bloom, Borod, Obler, and Gerstman (1993) also applied Gricean principles to compare the profiles of verbal pragmatic behaviors that emerge in patients with damage to either the left or right hemisphere. This approach differed from that of Prutting and Kirchner because it focused solely on the verbal pragmatic aspects of language without contextual information or paralinguistic cues. In the Bloom et al. study (1993), pictorial stimuli were used to elicit discourse that contained emotional and nonemotional content. Verbatim transcripts were used to avoid cues provided by facial and vocal expression. Raters naive to the hypotheses of the investigation made judgments about the presence or absence of certain pragmatic features of discourse using a dichotomous scale. Bloom et al. found that overall RBDs and LBD aphasic individuals were pragmatically impaired relative to normal controls (NCs). Further, the authors noted a strong relation between semantic content and verbal pragmatic performance in these patients. Specifically, emotional content in discourse facilitated appropriate verbal pragmatic behavior in the LBDs but suppressed pragmatic appropriateness in the RBDs.

Using a different sample of 46 subjects, Borod et al. (in press) applied Grice’s principles to compare the discourse production of LBD, RBD, and neurologically healthy NC right-handed adults. In that study, subjects produced monologues while recollecting emotional and nonemotional experiences. These narratives that varied with respect to condition (i.e., emotionality) and valence (i.e., pleasantness) were then examined by naive raters. Rather than using previously developed dichotomous measures, this study utilized a 5-point Likert scale, from “1” (inappropriate) to “5” (appropriate), to analyze each narrative. Both RBDs and LBDs were rated as less pragmatically appropriate than NCs. RBDs were more impaired for positive emotional discourse, and LBDs were more impaired for negative emotional
discourse. Further, these findings demonstrated that emotional content facilitated the pragmatic performance of LBDs yet suppressed the pragmatic performance of RBDs, consistent with findings from the earlier study by Bloom et al. (1993).

As evinced, there are several reasons why Gricean pragmatics provides a particularly useful model for evaluating the communication of brain-damaged patients. First, Grice’s emphasis on the role of the listener in interpreting the message lends credence to the use of raters to judge the appropriateness of communication. Second, adoption of Grice’s model allows raters to make observations of continuous discourse that are independent from observations of linguistic form. Finally, Grice assumes that the features of pragmatics are distinguishable entities that operate simultaneously in discourse. A few studies have explored the psychometric properties of pragmatic measures (e.g., Ball, Davies, Duckworth, & Middelhurst, 1991; Hux, Sanger, Reid, & Maschka, 1997; Penn, 1988). Hux et al. (1997) caution that discourse rating procedures are limited by the substantial amount of time involved in data collection, the extensive training required, and uniformity in sample elicitation and collection. Because subjective ratings are used to determine pragmatic appropriateness, it is critical that these limitations be taken into account in order to establish the reliability and validity of these measures.

The purpose of the current study was to examine the psychometric aspects of a pragmatic rating scale and to evaluate the theoretical relation among the different verbal pragmatic features that comprise Grice’s model. In addition, the influence of emotional content on the psychometric aspects of the scale was examined in light of previous research suggesting that emotional information mediates pragmatic performance in brain-damaged patients (for review, see Borod, Bloom, & Haywood, 1998). Such an investigation is needed to understand the principles that underlie the construction of a pragmatic rating scale and to clarify its potential as a clinical measure.

Methods

Participants

Participants were 16 healthy normal control (NC), 15 left brain-damaged (LBD), and 15 right brain-damaged (RBD) right-handed adults. Inclusionary criteria for all participants were as follows: native speakers of English, no neurological disease (secondary for patient groups), no psychiatric history, and no history of substance abuse. The three groups (NCs, LBDs, and RBDs, respectively) did not differ significantly with respect to mean age (59.5 ± 10.5, 64.4 ± 12.4, and 68.9 ± 9.9 years), mean education (15.9 ± 2.7, 15.0 ± 2.8, and 13.1 ± 1.9 years), or gender (37, 31, and 44% female).

Unilateral lesions in the LBD and RBD groups resulted from a single-incident cerebrovascular accident (CVA). Within the LBD group, 13 subjects were aphasics, and the group included 5 nonfluent, 2 fluent, and 6 mixed (i.e., both expressive and receptive deficits) aphasias (Goodglass & Kaplan, 1983). Global or severe aphasics were excluded because of the discourse production requirements of the study.

Significant differences were not found between the two brain-damaged groups when comparing mean months poststroke onset (LBDs = 18.3 ± 31.2, RBDs = 23.0 ± 40.2), etiology,
or accompanying sensory and motor deficits. Lesion location was determined by a review of CT scans (available for 93% of the participants) and/or clinical neurological examination. These lesion locations were classified as anterior (i.e., frontal lobe) or posterior (i.e., temporal, parietal, and/or occipital lobe) and as cortical or subcortical (i.e., any nonneocortical site (e.g., basal ganglia)). The distribution of specific lesion sites was relatively similar for LBDs versus RBDs: cortical—anterior only (1 vs 1), posterior only (2 vs 4), and anterior plus posterior (2 vs 3); subcortical (5 vs 4); cortical and subcortical—posterior plus subcortical (2 vs 1) and anterior plus posterior plus subcortical (2 vs 1); and unknown (2 vs 2).

**Elicitation Procedure**

Participants were asked to produce narratives about recollected personal experiences that were emotional (happiness, sadness, and anger) and nonemotional ("characteristics of people'’; hair color, intelligence, and weakness). Both emotional and nonemotional stimuli were used in light of research demonstrating the differential influence of these two types of information content on the language performance of brain-damaged individuals. Nonemotional narratives were used to control for linguistic, cognitive, and motoric factors that can confound brain-damaged patients' performance on discourse production tasks (Borod, Bloom, & Haywood, 1998). Emotional and nonemotional categories, respectively, were relatively well matched for word frequency (M = 53.7 and 56.3) (Thorndike & Lorge, 1944) and abstractness/concreteness (M = 3.1 and 3.7) (Paivio, Yuille, & Madigan, 1968; Spreen & Schultz, 1966). Abstractness/concreteness was measured using a 7-point Likert-type scale, where ‘‘1’’ = highly abstract and ‘‘7’’ = highly concrete.

Participants were instructed to identify an event related to the category and to attempt to recreate and describe the experience in detail. For each of the six categories, participants were provided with a printed label and asked to speak for at least 1 min, but were encouraged to speak as long as was necessary to fully recount the experience. If participants were unable to produce a narrative, the trial was terminated and the examiner continued onto the next category. All narratives were audiotaped and were then transcribed verbatim for discourse analysis.

**Rating Procedure**

The written transcripts were used for rating purposes, eliminating any prosodic or facial cues. Each narrative was rated for appropriateness on six pragmatic features (Bloom et al., 1993; Prutting & Kirchner, 1987; Rorie, 1994): Conciseness, Lexical Selection, Quantity, Relevancy, Specificity, and Topic Maintenance. These particular features were selected because they are conveyed verbally (i.e., not through voice or face) and correspond well to Grice’s cooperative principles (Bloom, 1990). The rating scale developed for this study deemphasized contextual variables and examined verbal pragmatics alone. Such a fine grain analysis permits raters to be trained to recognize and distinguish among the six closely related verbal pragmatic features.

**Training sessions.** Three right-handed graduate students in communication sciences were employed as raters. All three spoke English as their native language and had no psychiatric or neurological histories. They were naive to the experimental hypotheses and to the participant characteristics. A three-part training session was used to familiarize the raters with examples of each pragmatic feature at the different levels of appropriateness. In session one, raters were exposed to each pragmatic feature and the 5-point Likert Scale for appropriateness [‘‘1’’ (inappropriate) to ‘‘5’’ (appropriate)]. The following definitions were provided for the pragmatic features: Conciseness—use of utterances that are informative as required, but without unnecessary details; Lexical Selection—use of items that fit the text with no evidence of word finding difficulties; Quantity—content of the utterance is complete; Relevancy—selection of topics that relate to the matter under consideration; Specificity—discourse conveys specific,
unambiguous information; and Topic Maintenance—use of utterances that share a theme with preceding utterances. Raters were shown sample narratives that represented each individual feature at every level of the 5-point scale. See Appendix for the scale that was presented to the raters for each of the pragmatic features. Raters were also instructed not to focus on the grammatical complexity of the language produced. During session two, raters conferred as a group about the appropriateness of each feature on a set of sample narratives that they rated. Any narrative which generated a one-point discrepancy between two raters was reviewed and discussed. During session three, raters independently judged a set of 24 narratives on each of the six pragmatic features. There was complete agreement or a 1-point difference among raters for 80.2% of the comparisons. This level of concordance among the raters was deemed sufficient to proceed with rating the experimental data.

Experimental ratings. Finally, the narratives produced during the experiment were rated. There were 288 potential narratives to be rated (6 narratives × 3 participant groups × 16 participants per group). However, data from two participants were excluded from the study: one aphasic LBD (who was not able to produce any of the narratives) and one RBD (who could only produce 1 narrative [happiness]). In addition, four subjects were unable to produce 1 narrative each: one NC (weakness), one NC (happiness), and two RBDs (weakness). Therefore, 272 narratives were rated and used in the statistical analyses to follow. The final set of narratives presented to the raters was completely randomized across individual participants, subject groups, pragmatic features, and emotional and nonemotional categories.

After the experimental ratings were made, complete and partial rater agreement were examined on a subset of the experimental data, which included the narratives of six LBDs, six RBDs, and six NCs, randomly selected. Overall, there was complete agreement or a one-point difference among the ratings for 80.5% of the 648 comparisons (6 narratives × 6 pragmatic features × 3 groups × 6 participants per group). There were no differences in these ratings as a function of narrative (79.2% for happiness, 82.2% for sadness, 80.4% for anger, 79.8% for intelligence, 80.3% for weakness, and 80.5% for hair color), pragmatic feature (78.8% for Conciseness, 80.6% for Lexical Selection, 81.5% for Quantity, 80.9% for Relevancy, 80.9% for Specificity, and 79.0% for Topic Maintenance), or participant group (82.4% for LBDs, 78.8% for RBDs, and 79.7% for NCs). These percentages reported above were high enough to combine raters’ scores in the statistical analyses to follow.

RESULTS

Internal Consistency

The first step was to examine the coherence of the six pragmatic features with respect to the underlying dimension of pragmatic competence. Cronbach’s alpha from the 46 participants’ data was computed to examine internal consistency among the pragmatic features (Nunnally, 1978). Values for individual narratives were quite high: happiness (.89), anger (.93), sadness (.94), intelligence (.91), weakness (.85), and hair color (.85). Three composite measures were created in order to examine ratings when isolating emotional content (E mean), nonemotional content (NE mean), and overall content [Total mean = (E + NE)/2]. Note that the “E mean” refers to the average value for happiness, anger, and sadness narratives and that the “NE mean” refers to the average value for intelligence, weakness, and hair color narratives. These three means were utilized in all analyses. When Cronbach’s alpha was computed for these composite measures, values for internal consistency were very high: E mean (.95), NE mean (.93), and Total mean (.96).
Alpha levels for each composite score were computed for each separate participant group to assess any between-group differences. Both brain-damaged groups separately evidenced alpha levels that were similar to the whole sample. The following alpha levels were obtained for the LBD and RBD groups, respectively: E mean (.95 and .96), NE mean (.93 and .94), and Total mean (.95 and .98). Interestingly, although the NC group also evidenced substantial internal consistency, alpha levels were somewhat lower compared to those for the individual brain-damaged groups: E mean (.85), NE mean (.78), and Total mean (.85).

Factor Analysis

Next, construct validity was examined on an exploratory basis. Even though the scale has strong internal consistency, we wanted to know the extent to which the individual features captured distinct aspects of pragmatic functioning. To explore the construct validity of the scale, a principal axis factor analysis with promax rotation was conducted (Tabachnick & Fidell, 1996). We constrained the results to four factors based on the number of principles in Grice’s theoretical conceptualization of pragmatics. Data from all participants were used to maximize power. Only those features loading at .50 or higher were considered. Rotation was used to improve interpretation of the factors and to further explore the factor structure of the scales.

Before rotation, not surprisingly (given the findings for internal consistency), the initial factor accounted for 83.7% of the variance and contained high loadings on each of the six pragmatic features (see Table 1). No other interpretable factors emerged prior to rotation. After rotation, there were three factors with pragmatic features loading at .50 or higher. These factors were interpreted. The first factor appeared to index “discourse content,” containing loadings for Lexical Selection (.90), Quantity (.84), and Specific-
ity (.71). The second factor indexed Conciseness (.71) and was labeled “parsimony.” The third factor appeared to index “conceptual unity” and contained loadings for Relevancy (.62) and Topic Maintenance (.53). The fourth factor was not interpreted because none of the loadings were above .50; loadings ranged from −.04 to .31, with .02 as the median.

Principal axis factor analyses were conducted separately for the NE and E means to ascertain the effects of emotional content on the pragmatic scale. Prior to rotation, a substantial “general factor” was identified in each case, which accounted for a significant portion of the variance, comparable to the finding for the Total mean: NE mean (74.2 %) and E mean (81.5 %). After promax rotation, the factor structures for the NE and E means were similar to the structure for the Total mean score. Following rotation for the NE scores, as with the Total scores, Factor I reflected “discourse content” (Lexical Selection = .93, Quantity = .81, Specificity = .79). Factor II, again indexed Conciseness (.69). Factor III, “conceptual unity,” had highest loadings for Topic Maintenance (.55) and Relevancy (.45). Similarly, when examining the factor structure for the E scores, Factor I included Quantity (.80), Lexical Selection (.79), Specificity (.64), and Topic Maintenance (.57). Factor II, again, indexed Conciseness (.74), and Factor III included Relevancy (.58) and Topic Maintenance (.34). For two of the variables in these analyses, factor loadings were less than .50. However, these were the variables with the next highest loadings on their respective factors.

Discussion

This article explored the psychometric aspects of a pragmatic rating scale based on Gricean pragmatics. Consonant with Grice’s view of pragmatics, the role of the listener as interpreter is assumed by trained raters. Further, ratings rely on observations of continuous discourse despite any irregularities in the structure of an individual’s words or sentences. Specifically, this study investigated Grice’s conceptualization that pragmatics may be construed as a distinctive set of rules or cooperating principles that operate simultaneously in discourse. In addition, internal consistency was determined, and a factor analysis was conducted to evaluate the psychometric characteristics of a 5-point verbal pragmatic rating scale.

To study the reliability of the scale, internal consistency was determined. When this analysis was conducted, findings revealed that the ratings of these verbal pragmatic features have substantial internal consistency. This was the case for each of the six narratives and for the Emotional, Nonemotional, and Total composite measures. Thus, these findings suggest that the pragmatic ratings are reliable measures, regardless of content. The six individual pragmatic features were highly correlated indicating that the overall measure has substantial integrity. Because the nonverbal aspects of pragmatics (e.g., facial, gestural, and prosodic cues) were excluded from this study, the level
of internal consistency for our scale may be higher than levels obtained when both verbal and nonverbal measures are used. Penn (1988) noted that extra-linguistic factors and social context substantially increase the complexity of pragmatic measures. In a similar vein, Hux et al. (1997) stated that assessments driven by pragmatically based observational data may be difficult to interpret. The approach used to study pragmatics in this article is different from approaches that examine functional communication where nonverbal information is an essential component of message transmission. Nevertheless, the current study deliberately isolated the verbal aspects of pragmatics to reduce the potential noise that occurs when examining more than one communication modality during a single observation.

High internal consistency was also achieved when individual participant groups were examined. It is interesting to note that the alpha coefficients were higher for the brain-damaged groups than for the normal group. These ratings may be more sensitive to breakdowns in pragmatics in individuals with communication disorders. Perhaps it is more difficult to detect discrepancies among the use of the pragmatic features in the discourse of the neurologically healthy participants because of the simultaneous way Gricean principles operate.

Borod et al. (in press) used this rating scale to examine differences in performance among patient and control groups. Overall, LBDs (M = 3.02 ± .57) and RBDs (M = 3.24 ± .58) were impaired relative to NCs (M = 3.64 ± .35). That the performance of the brain-damaged groups was differentiated from the performance of the NCs suggests that the scale possesses some degree of discriminant validity. Also noted were differences in the performance of the brain-damaged groups as a function of emotional content and valence. The significance of the Borod et al. (in press) findings is bolstered by the similar psychometric properties of the emotional and nonemotional composite scores revealed in the current study.

The factor analysis was conducted to explore the theoretical relations among the six pragmatic features and to provide some assessment of the scale’s construct validity. Consistent with the findings for internal consistency, the factor analysis conducted on composite scores revealed a substantial general factor before rotation. After rotation, the three factors that emerged indexed distinctive aspects of pragmatic performance. Interestingly, the three pragmatic features that relate most to Grice’s principle of Quantity (i.e., Quantity, Lexical Selection, and Specificity) loaded most highly on the initial factor and were differentiated from factors indexing Relevancy/Manner (i.e., Relevancy and Topic Maintenance) and Quality (i.e., Conciseness). On first inspection, the feature of Conciseness would seem to be related to Grice’s principle of Quantity. However, the description of Conciseness used here differs somewhat from the definition of this principle as it appears in the clinical literature (e.g., Damico, 1992; Hux et al., 1997). From our perspective, the feature of Conciseness is related to Grice’s principle of Quality.
(i.e., the contribution should be truthful and supported with evidence). The factor label “parsimony” for the Conciseness feature was derived to capture this relationship and underscore the importance of brevity in discourse. It appears as though Grice’s cooperative principles are represented among the three factors emerging from the factor analysis—Discourse Content, Conceptual Unity, and Parsimony. In sum, the findings for internal consistency suggest that the six pragmatic features measure the same underlying concept (i.e., verbal pragmatic competence), whereas the findings from the factor analysis suggest possible substructures.

Investigations of this nature reveal the complexity of psychometric issues in pragmatics. One issue that has received attention in the literature is rater reliability. Several aspects of rater reliability have been discussed. It is clear that raters require a structured training protocol in order to guarantee the soundness of the measures that will eventually be applied (Hux et al., 1997). In the current study, raters underwent an extensive training procedure that included the opportunity for conferencing and improvement to criterion. Two types of frequently used rating scales have been debated in the literature. Whereas Ball et al. (1991) reported higher interrater reliability with a dichotomous scale, Penn (1998) found higher reliability with a more extensive 5-point Likert scale. These discrepant findings suggest that the strength of a rating scale may be dependent on the methods used and the particular pragmatic features studied, as well as the range of the scale. When a more extensive scale is used, special issues in training raters arise. For example, Ball et al. (1991) point out that raters tend to choose scores that hover near the midpoint, leading to possible nonmeaningful differences. Research with more extended rating scales should be sure to train raters to appreciate the array of points, including use of both endpoints. The construction of the materials that are rated should also be considered. Many studies that examine pragmatics have used a limited number of samples which could reduce the statistical power of the investigation. An advantage of the current study was the use of multiple discourse samples from each participant which included control, as well as experimental, stimuli. Finally, the actual statistic used to assess reliability has bearing on the generalizability of the results (Cordes, 1994). Hux et al. (1997) discuss the relative strengths and weaknesses of various measures of reliability used to study pragmatics (e.g., Pearson product-moment correlations, interobserver agreement, Cohen’s kappa, and generalizability coefficients). They conclude that controlled research studies are needed to validate pragmatic measures before they are used for clinical purposes.

In terms of the theoretical implications of this study, future work should be conducted to provide a confirmatory model of the structure of pragmatics. Such investigations should use a larger sample size and increase the number of individual variables that could be specifically mapped onto Grice’s model. In addition, an increased sample size would permit an investigation of the
psychometric properties across different types of aphasia as well as in relationship to severity of language impairment (e.g., Penn, 1988). Overall, our findings seem to support Grice's theoretical assumption that there are several distinct pragmatic rules that operate simultaneously in discourse.

The features in our measure of pragmatics appear to provide a meaningful and valid index for identifying aspects of normal and disordered discourse. Pragmatic analysis based on Gricean principles has potential for describing aspects of the impaired language system and for illuminating how these verbal deficits influence social functioning. This scale has potential for clinical use and could be employed in conjunction with measures of paralinguistic (i.e., prosodic) and nonverbal (i.e., facial and gestural) aspects of pragmatics. Together, such an approach could provide a comprehensive analysis of pragmatic behavior.

APPENDIX

Verbal Pragmatic Rating Scale

CONCISENESS

Definition of pragmatic feature: Use of utterances that are informative as required, but not too informative. There are no unnecessary details.

Examples
a. Appropriate
   - Quantity of the information is sufficient to convey the point of the story.
   - Nothing unrelated is introduced.

b. Inappropriate
   - Unnecessary information and excessive details are provided.

Pragmatic scale

<table>
<thead>
<tr>
<th>Inappropriate</th>
<th>Appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Irrelevant or superfluous detail</td>
<td>Appropriate amount of detail</td>
</tr>
</tbody>
</table>

LEXICAL SELECTION

Definition of pragmatic feature: Use of lexical items that fit the text. There is no evidence of word finding difficulties; therefore there appears to be adequate access to words.

Examples:

a. Appropriate
   - Lexical choices convey information.
   - Specific referents are used.

b. Inappropriate
   - Lexical choices are nonspecific.
   - Words do not facilitate understanding of the text.

Pragmatic scale

<table>
<thead>
<tr>
<th>Inappropriate</th>
<th>Appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Inappropriate or limited selection of words</td>
<td>Appropriate and varied selection of words</td>
</tr>
</tbody>
</table>
### QUANTITY

Definition of pragmatic feature: Content of utterance is complete.

**Examples**
- **Appropriate**
  - Amount of information is sufficient to convey content.
- **Inappropriate**
  - An insufficient amount of information is given.
  - Content of the response is missing.

**Pragmatic scale**
- **Inappropriate**
  - 1
- **Appropriate**
  - 5

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Communication is incomplete</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Communication is complete</td>
</tr>
</tbody>
</table>

### RELEVANCY

Definition of pragmatic feature: Selection of topics that relate to the matter under consideration.

**Examples**
- **Appropriate**
  - Discourse is relevant and relates to the matter under consideration.
- **Inappropriate**
  - Remarks are unrelated and irrelevant given the nature of the task.

**Pragmatic scale**
- **Inappropriate**
  - 1
- **Appropriate**
  - 5

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Many irrelevant remarks included</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Remarks are relevant</td>
</tr>
</tbody>
</table>

### SPECIFICITY

Definition of pragmatic feature: Discourse conveys specific, unambiguous information.

**Examples**
- **Appropriate**
  - Specific information is provided and enhances the understanding of the topic.
- **Inappropriate**
  - Details are omitted.
  - Information is vague.

**Pragmatic scale**
- **Inappropriate**
  - 1
- **Appropriate**
  - 5

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information is vague</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Information is specific</td>
</tr>
</tbody>
</table>

### TOPIC MAINTENANCE

Definition of pragmatic feature: Use of utterances that share a theme with preceding utterances and add information to the prior communicative act.

**Examples**
- **Appropriate**
  - Utterances contribute to the topic.
  - Information or feedback is provided to move the topic forward.
- **Inappropriate**
  - New and unrelated topics are introduced.
  - Topics are not developed.

**Pragmatic scale**
- **Inappropriate**
  - 1
- **Appropriate**
  - 5

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Many unrelated topics introduced</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Topic was maintained and developed</td>
</tr>
</tbody>
</table>
REFERENCES


