Measurements Predictive of Generalization of Response Elaboration Training

Katherine E. Yedor, Caryn P. Conlon, and Kevin P. Kearns

As emphasis on functional outcomes of communicative intervention strategies has increased, speech-language pathologists have become more cognizant of the need to program generalization. Generalization promoting strategies, first conceptualized by Stokes and Baer (1977), have been incorporated into increasing numbers of aphasia treatment studies (Thompson & Byrne, 1984; Doyle, Goldstein, & Bourgeois, 1987; Doyle, Goldstein, Bourgeois, & Nakles, 1989; Doyle, Oleyar, & Goldstein, 1991). However, even when generalization is actively programmed, not all subjects achieve equivalent amounts of generalized responding. Conversely, "train and hope" studies, in which generalization across responses, settings, experimenters and/or time is measured but not actively pursued, *do* sometimes result in the occurrence of generalized responding. Therefore, training to promote generalization may not be equally effective or even necessary for all patients. Prognostic efforts may need to include a means of identifying patients who will not spontaneously generalize so that necessary levels of differential treatments can be employed.

"Loose training," rooted in divergent semantic intervention (Chapey, 1986), has been proven to effectively promote generalized language and communication improvements in aphasia (Gaddie, Kearns & Yedor, 1991; Kearns, 1985, 1986; Kearns & Potechin-Scher, 1988; Kearns & Yedor, 1993). Therefore, measuring inherent divergent language as reflected in patient response variability, relevance, and production of logical alternatives appears potentially useful for identifying intra-individual factors that predict generalization.

One such measure is "peak-mean difference" (PMD) (Porch, 1978) between the highest score and the mean for each subtest on the Porch

Index of Communicative Ability (PICA) (Porch, 1981). Although Porch devised PMD as a prognostic measure, research has not made useful predictions about recovery (Aten & Lyon, 1978; Porch & Callaghan, 1981; Wertz, Dronkers, & Hume, 1993). Because the PMD is based on variability in patient responses, it may be more useful in predicting generalization than recovery. That is, if response variability is associated with the ability to generalize training effects, individuals with high PMD scores might be expected to show more spontaneous generalization than those with low PMDs. Although PMD has been used to analyze performance on PICA subtests, the concept of calculating peak mean difference scores for patient responses produced in other situations appears valid.

We have used the "novel content" measure to determine how much content variety patients emitted during Response Elaboration Training (RET, [Gaddie, Kearns & Yedor, 1991]), a loose training approach designed to increase the amount of informational content produced by aphasic individuals. Thus, subjects who continuously produced the same rote response across sequential presentations of treatment stimuli received a low mean novel content score. However, subjects who produced a range of relevant logical alternatives over time demonstrated high mean novel content scores. Therefore, as novel content is also associated with response variability and the ability to generate relevant logical alternatives, analysis of mean novel content appears to be another potential means of measuring intrasubject factors predictive of generalization.

This study examined several pretreatment variables to determine if any were predictive of generalization in nine subjects enrolled in RET protocol (Kearns, 1985, 1986), which uses shaping, chaining, and modeling techniques to combine patients' spontaneous descriptions of action pictures and subsequent elaborations into longer, more contentive responses. RET uses a number of generalization-promoting techniques. It relies upon the individual patient's internal context to generate a range of acceptable responses rather than limiting the patient to a closed number of clinician pre-selected responses. Initial analysis of RET (Kearns & Potechin-Scher, 1988) reported a moderate degree of generalization across untrained stimuli—people and settings for most subjects. However, the amount of generalization was not equivalent across subjects, with some subjects showing only limited generalization (Kearns & Yedor, 1993).

We investigated whether the amount of generalization to untrained stimuli after participation in a RET protocol can be predicted by any of the following pretreatment measurements of response variability and ability to generate logical alternatives: PMD for PICA verbal subtests; PMD calculated for patient responses emitted during baseline probe sessions; and mean novel content produced during baseline sessions.

METHOD

Subjects

Records of nine subjects who had completed RET programs were reviewed. All patients suffered a single left hemispheric cerebrovascular accident (CVA) at least 5 months prior to enrollment in the protocol (range = 5 months to 12 years, $\bar{x}$ = 5.4 years). The subjects all had at least an eighth-grade education, were right-handed, and varied in age from 57 to 69 years ($\bar{x}$ = 65.1). Five subjects were classified as Broca's aphasia on the Western Aphasia Battery (WAB) (Kertesz, 1979) and had overall PICA severity levels ranging from 9.43 to 12.08. The four remaining subjects demonstrated Conduction and Anomic aphasia on the WAB. PICA overall severity levels were 10.86 and 10.4 for the two subjects with Conduction aphasia and 13.2 and 14.7 for the two anomic aphasic subjects.

Design

A post hoc analysis was conducted on data obtained from the nine subjects. Aspects of pretreatment performance hypothesized to be associated with response variety and ability to generate logical alternatives were examined for each patient via descriptive analysis. The relationship between these descriptive analyses and the amount of generalization was examined by computing Pearson product-moment correlation coefficients and then determining the proportion of variance accounted for by each variable.

Independent Variables

Qualitative factors used to predict extent of generalization included (a) analysis of PMD for PICA verbal subtests; (b) analysis of PMD for content words produced during baseline sessions; (c) calculation of mean novel content produced during baseline sessions. These measurements are defined as follows:

PICA Verbal PMD (PICA PMD). This is the mean difference between the highest score and mean score for each verbal subtest on the PICA.

Baseline Peak Mean Difference (BL PMD). A peak mean difference score was also computed using data obtained during baseline sessions. All words deemed closely related to the primary theme of the stimulus picture and conveying new information were tallied as "content words." For each subject, the difference was calculated between the highest number of content words and mean number of content words produced for each set of training stimuli during each baseline session. These scores were then averaged across all baseline sessions to generate the BL PMD, which reflects the fluctuation in amount of content produced by each subject at baseline.

Baseline Mean Novel Content (MNC). A novel content word was defined as a content word that was not produced in response to the same stimulus picture during a comparison baseline session. Mean novel content, therefore, referred to the average number of content words that were not produced during the previous baseline session in reference to the same stimulus items. It thus reflected the variety of patient responses.

Dependent Variable

The extent of generalization to untrained stimuli was the dependent variable for this study. Generalization was measured by determining the "Generalization Quotient" or GQ, the proportion of improvement that occurred for trained stimulus items and was also present for untrained sets. The extent of generalization was calculated this way for three different situations (generalization to the second training set and an untrained set when the criteria for the first set of trained items was met; and generalization to a set of untrained items when the criteria for the second set of training items was met). These three scores were averaged for each subject to form the GQ.

Reliability

Since all calculations were derived from content word scores, point-to-point interjudge reliability was assessed to identify content words by independently rescoring written transcripts of baseline and criteria probe sessions by a second clinician. Point-to-point reliability averaged 96.08% with a range of 95% to 98%.

RESULTS

Calculation of GQs revealed that even though all subjects were enrolled in an identical treatment protocol, there was considerable between-subject variability in the amount of generalization (see Table 1). On average, however, subjects demonstrated improvement for untrained items which was 33.6% of that for trained items. The range was considerable, however; one

TABLE 1. GENERALIZATION QUOTIENTS (GQ) CALCULATED FOR NINE SUBJECTS ENROLLED IN RESPONSE ELABORATION TRAINING PROGRAMS

Aphasia Type	*GQ*	*Degree of Generalization*
Broca's		
LF	-2.3	None
NS	8	Limited
RW	31	Moderate
VS	46	Moderate
JT	68	Significant
Anomic		
JU	2	Limited
JA	15	Limited
Conduction		
JJ	51	Moderate
RG	84	Significant

patient (L. F.) actually produced less content at the end of treatment for untrained stimulus items.

The relationship between the pretreatment *PICA* PMD score and GQ is shown in Figure 1 for non-fluent and fluent aphasic subjects. An inconsistent pattern of results was obtained. For non-fluent aphasic subjects only a slight negative relationship is present, while for fluent aphasic subjects, a much stronger positive relationship can be seen. Compared to previous studies examining the ability of PICA PMD to predict prognosis for improved language skills (Aten & Lyon, 1978; Porch & Callaghan, 1981; Wertz, Dronkers, & Hume, 1993), it appeared to reasonably predict generalization for three of four fluent aphasic subjects. Similarly, calculation of Pearson product-moment correlation coefficients for GQ and PICA PMD revealed correlations of −.32 and .78 for nonfluent and fluent aphasic subjects, respectively.

Scattergrams depicting the relationship between PMD for content words produced during baseline sessions and GQs are shown in Figure 2. Baseline peak mean difference scores appear related to generalization for both nonfluent and fluent subjects. In fact, ranking of subjects according to BL PMD responses equalled the ranking according to GQ for four of five nonfluent and three of four fluent subjects. Calculation of correlation coefficients revealed a positive relationship between these two factors. Specifically, for non-fluent subjects, the correlation coefficient was .64 while for fluent aphasic subjects, it was .80.

Figure 3 illustrates the relationship between novel content produced during baseline sessions and GQs for both nonfluent and fluent aphasic

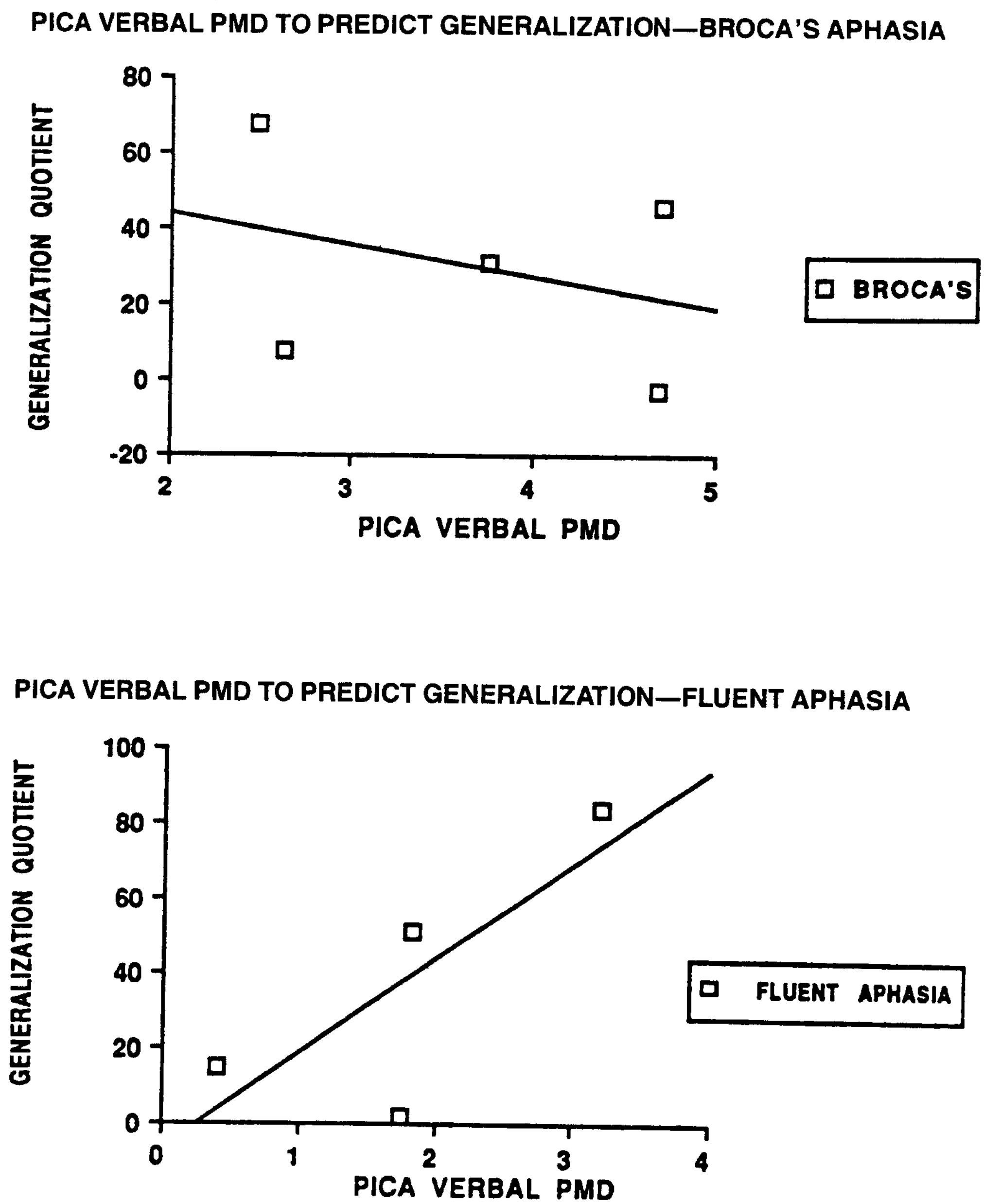

Figure 1. Relationship between PICA verbal peak, mean different scores, and generalization quotients for subjects with nonfluent (top) and fluent (bottom) aphasia.

groups. Two factors only demonstrate a slight relationship trend for non-fluent aphasic patients. However, for the fluent aphasic patients, the relationship appears to be more robust. Calculated correlation coefficients are only .24 for the non-fluent subjects but .76 for the individuals with fluent aphasia.

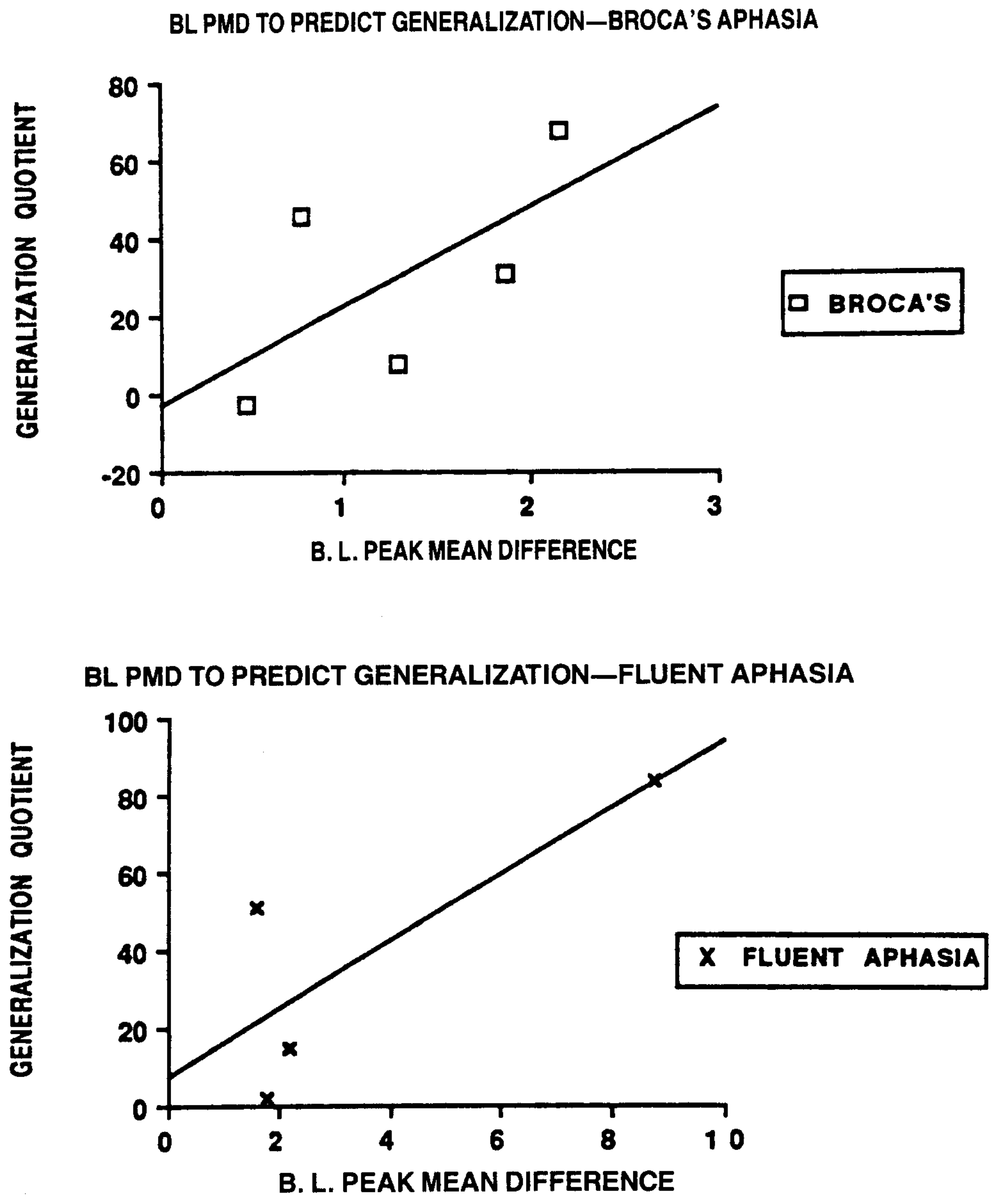

Figure 2. Relationship between peak mean different scores computed from baseline probe sessions and generalization quotients for subjects with non-fluent (top) and fluent (bottom) aphasia.

To summarize the correlation coefficients for the two groups, only baseline peak mean difference appears to be correlated with generalization quotient for the nonfluent group. However, all three measurements of response variability were reasonably correlated with generalization as measured by the GQ for the fluent subjects. Further analyses computed for each independent variable for the fluent group show that between 57%

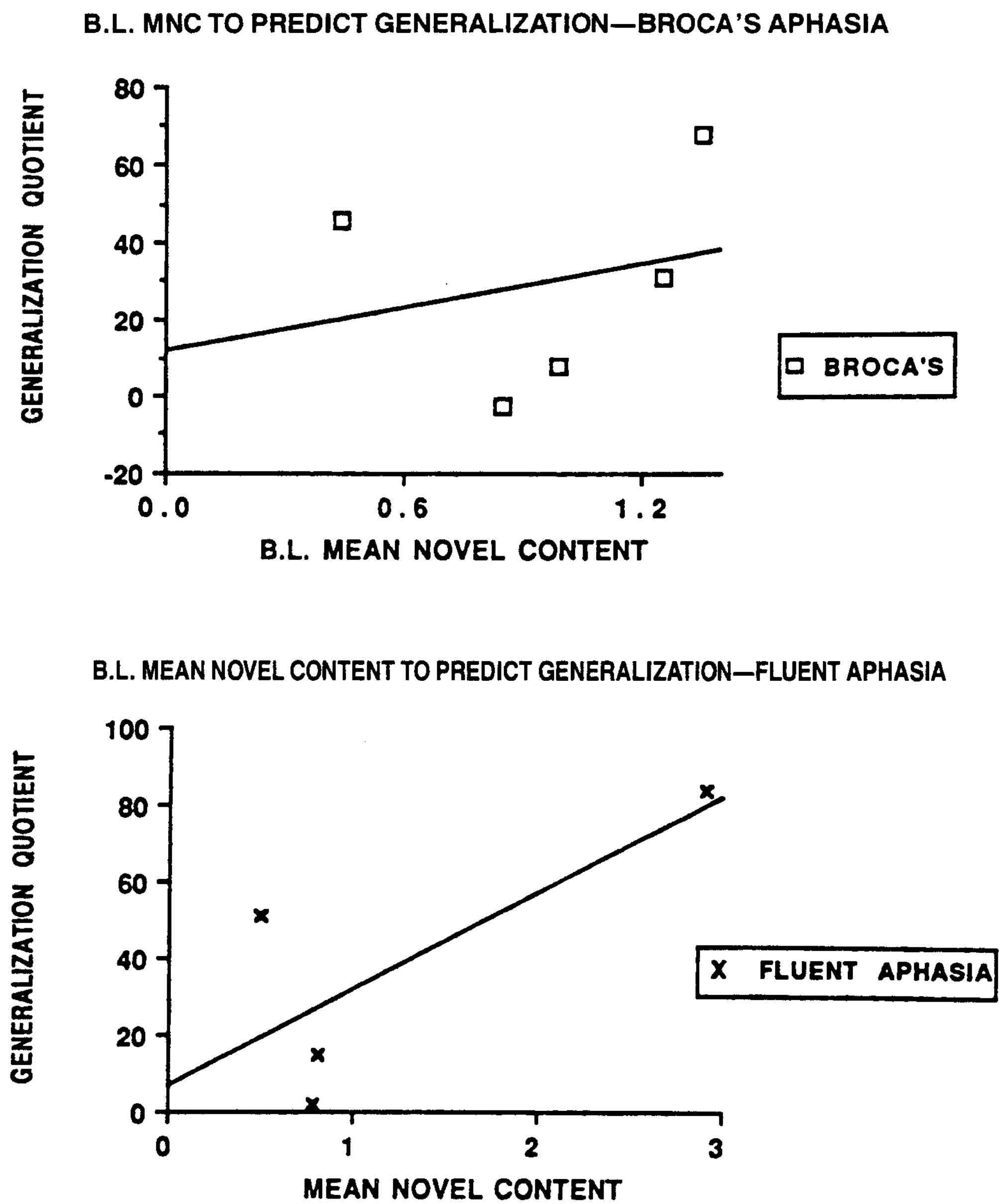

Figure 3. Relationship between mean novel content scores calculated from baseline probe sessions and generalization quotients for subjects with non-fluent (top) and fluent (bottom) aphasia.

and 64% of the variance for each subject can be attributed to the experimental variables, while 36% to 43% of the variance is unexplained. For non-fluent individuals, in the independent variable that was most strongly correlated to generalization (BL PMD), only 41% of the variance for each subject could be attributed to experimental factors.

CONCLUSIONS

Although replication with larger number of subjects is obviously needed, crucial measures of response variability appear to predict some extent of generalization. Our conclusions are preliminary, based on correlations for nine subjects divided into fluent and nonfluent groups.

Predicting the amount of generalization was most accurate for subjects with fluent aphasia. Specifically, while all three measurements of response variability examined were found to correlate with generalization for patients with anomic and conduction aphasia, for patients with Broca's aphasia only one measurement was of any value in predicting generalization. The reason for the proposed predictors' limited ability to correlate with actual generalization for this patient group may relate, in part, to the group's language characteristics: The number of content words produced by the nonfluent aphasic subjects during baseline sessions was extremely small (mean of 1.9 content words, with only one patient surpassing a mean of 1.7 content words). Given these limitations in ability to produce any content, it is not surprising that production of novel content or other measures of variability in baseline performance was limited prior to intervention.

In subjects with fluent aphasia, all three measures selected as indicative of patient response variability appear to have some value in predicting generalization. For the subjects with Broca's aphasia, however, only measurements of variability in the amount of content produced during baseline sessions were predictive of generalization. A significant amount of the variance in scores for both subject groups is unexplained, however, and it appears likely that factors in addition to divergent language abilities are responsible for the extent of generalization. For example, as some of the patients who produced greater amounts of content during baseline sessions had higher generalization quotients, it is possible that patients with only moderately impaired expressive language may be better able to generalize the results of our training. Similarly, for these patients with severely restricted verbal output, one important factor behind their ability to generalize the effects of our training may be their actual motor planning capabilities. To some extent, it seems possible that having the ability to program longer, more contentive utterances, at least occasionally, prior to intervention is a prerequisite for generalization of RET. This ability may be best reflected by the BL PMD score, which was in fact the only indicator that predicted generalization for both subject groups. As the "generalization quotient" was calculated based on increases in the *number* of relevant content words for trained and untrained stimulus items and BL PMD was the only variable that took into account the *number* of content words produced at baseline, this finding is not unexpected.

More surprising was the failure of PICA verbal PMD to predict generalization for nonfluent subjects, particularly given the strength of its predictions for individuals with fluent aphasia. It may be that PMD for PICA verbal subtests is less reliable than PMD calculated from baseline sessions. This may be in part due to the fact that the latter is derived from repeated measurements of patient performance over time. Measurements calculated from administration of the PICA at a single point in time may be more subject to external factors affecting performance, and observed patient variability may be due to extra- rather than intrasubject variables.

To increase the accuracy of mean novel content as a generalization predictor, particularly for persons with nonfluent aphasia, we are exploring whether baseline measurements should be collected in a manner that encourages production of novel responses. For example, for each stimulus item, after patients produced an initial response, they could be instructed to "tell me something different about this picture." Thus, the patient's ability to generate a variety of relevant alternatives under optimal conditions would be more clearly measured.

The efficacy of "loose training" procedures has previously been demonstrated (Thompson & Byrne, 1984; Doyle, Goldstein, Bourgeois, & Nakles, 1989). However, not all patients are equally adept at taking advantage of the myriad options offered by the loose training techniques of reducing clinician control over stimulus, response, and feedback. Despite our best attempts to present a wide variety of stimulus conditions and response options, some patients continue to respond in a highly structured manner with little or no response variety. Although the reasons behind these differences in responding have not yet been investigated, intrasubject factors such as differences in cognitive style may be important. Response variability and ability to generate logical alternatives are reasonable ways to begin measuring these differences. By developing a way to identify these individuals who will most likely not generalize prior to our interventions, we can modify our treatment protocols to further ensure that changes in language behaviors during therapy sessions translate into changes in communication behaviors during real-life situations.

ACKNOWLEDGMENTS

The writing of this manuscript was supported by funds from the Veterans Affairs Rehabilitation Research and Development Service. The authors extend their thanks to Megan Ballantine for her assistance with data analysis.

REFERENCES

Aten, J., & Lyon, J. (1978). Measures of PICA subtest variance: A preliminary assessment of their value as predictors of language recovery in aphasic patients.

In R. H. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings: 1978.* Minneapolis, MN: BRK Publishers.

Chapey, R. (1986). Cognitive intervention: Stimulation of cognition, memory, convergent thinking, divergent thinking, and evaluative thinking. In R. Chapey (Ed.), *Language intervention strategies in adult aphasia.* (2nd ed.). Baltimore, MD: Williams & Wilkins.

Doyle, P., Goldstein, H., & Bourgeois, M. (1987). Experimental analysis of syntax training in Broca's aphasia: A generalization and social validation study. *Journal of Speech & Hearing Disorders, 52,* 143–155.

Doyle, P., Goldstein, H., Bourgeois, M. S., & Nakles, K. (1989). Facilitating generalized requesting behavior in Broca's aphasia: An experimental analysis of a generalization treatment procedure. *Journal of Applied Behavior Analysis, 22,* 157–170.

Doyle, P., Oleyar, K. S., & Goldstein, H. (1991). Facilitating functional conversational skills in aphasia: An experimental analysis of a generalization training procedure. In T. E. Prescott (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 19, pp. 229–241). Austin, TX: PRO-ED.

Gaddie, A., Kearns, K. P., & Yedor, K. (1991). A qualitative analysis of response elaboration training effects. In T. E. Prescott (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 19, pp. 171–183). Austin, TX: PRO-ED.

Kearns, K. P. (1985). Response elaboration training of patient initiated utterances. In R. H. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 15, pp. 196–204). Minneapolis, MN: BRK Publishers.

Kearns, K. P. (1986). Systematic programming of a pragmatic approach to aphasia management. In R. C. Marshall (Ed.), *Case studies in aphasia rehabilitation.* Austin, TX: PRO-ED.

Kearns, K. P., & Potechin-Scher, G. (1988). The generalization of response elaboration training effects. In T. Prescott (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 18, pp. 223–246). San Diego CA: College-Hill Press.

Kearns, K. P., & Yedor, K. (1991). An alternating treatments comparison of loose training and a convergent treatment strategy. In T. E. Prescott (Ed.), *Clinical Aphasiology* (Vol. 20, pp. 223–238). Austin, TX: PRO-ED.

Kertesz, A. (1979). *Western Aphasia Battery.* New York: Grune & Stratton.

Porch, B. E. (1978). *Diagnosis of aphasia and brain dysfunction.* Albuquerque, NM: VA Satellite Presentation.

Porch, B. E. (1981). *Porch Index of Communicative Ability,* (Vol. 2). Palo Alto, CA: Consulting Psychologists Press.

Porch, B. E., & Callaghan, S. (1981). Making predictions about recovery: Is there HOAP? In R. H. Brookshire (Ed.), *Clinical Aphasiology Conference Proceedings* (Vol. 11, pp. 187–200). Minneapolis, MN: BRK Publishers.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349–367.

Thompson, C. K., & Byrne, M. E. (1984). Across setting generalization of social conventions in aphasia: An experimental analysis of "loose training." In R. H. Brookshire (Ed.), *Clinical Aphasiology* (Vol. 14, pp. 132–144). Minneapolis, MN: BRK Publishers.

Wertz, R. T., Dronkers, N. F., and Hume, J. (1993). PICA intrasubtest variability (PMD) and prognosis for improvement in aphasia. In M. Lemme (Ed.), *Clinical Aphasiology* (Vol. 21, pp. 207–211). Austin, TX: PRO-ED.