Diagnostic Baselines in Long Term Aphasia

Pat A. Holtzapple
Veterans Administration Medical Center, Phoenix, Arizona

Marilyn Kerley
Arizona State University, Tempe, Arizona

Purpose. The purpose of this report is to examine the stability of diagnostic baseline patterns of long term aphasic patients upon entering a treatment program. After reviewing the serial test results of two aphasic patients ranging from 44-70 months postonset, I will describe the protocol which has grown out of these findings.

Background. Hersen and Barlow (1976) define baseline as the "... initial period of observation involving the repeated measurement of the natural frequency of occurrence of the target behaviors under study" (p.74). A baseline measurement typically involves the assessment of specified target behaviors (LaPointe, 1978). However, I am using the term diagnostic baseline to refer to the results of a wide range of tasks included on a standardized test battery. A diagnostic baseline would, by this definition, be determined by observing performance on the repeated administration of a test in the absence of treatment.

It is generally agreed that repeated measurement should continue until a stable pattern emerges, with a minimum of three observation points required to estimate a trend in the data (Davis, 1978; Hersen and Barlow, 1976). However, the number of observation points required to achieve a stable diagnostic baseline for a long term aphasic patient may range from three to infinity.

Porch (1969) has advocated repeated administration of the Porch Index of Communicative Ability, or PICA, to establish diagnostic baseline patterns before initiating a treatment program with the long term aphasic patient. Although few published accounts of pretreatment diagnostic baselines for long term aphasic patients were found by this author, evidence provided by Wertz (1977) and Broida (1977) suggests that stable diagnostic baselines can be achieved. Wertz reported a gain in the Overall percentile score from the 72%ile to the 75%ile when the PICA was administered twice, one week apart, to an aphasic patient at 8 years 9 months postonset. Broida found similar results, ranging from 0-3%ile units in the Overall percentile (OA%ile) score when the PICA was administered twice, two to four weeks apart, to six aphasic patients ranging from 15 to 74 months postonset.

Test-retest reliability. Studies based on group designs with adults and children of normal and mentally retarded populations have reported the occurrence of significant mean gains on intelligence test-retest scores (Anastasi, 1973; Kangas and Bradway, 1971; Matarazzo, Weins, Matarazzo, and Manaugh, 1973). Catron and Thompson (1979) suggest that the generally high correlations found on test-retest studies may mask the gains that typically occur on an IQ retest. They contend that "Correlations can mislead by implying similarity between test-retest scores when in fact essentially all scores gain on retest, and only the rank order of the scores is similar" (p.356).

There is evidence that suggests that aphasic patients perform with consistency when the overall scores of group data are investigated (Schuell, Jenkins and Jimenez-Pabon, 1964). However, group data do not account for the individual differences which are critical elements in planning treatment (Matarazzo, Matarazzo, Gallo and Weins). A respectable coefficient of reliability on test-retest data published in the test manual does not insure that any individual will replicate his/her performance on subsequent tests.

Gallaher (1979) administered the Token Test (DeRenzi and Vignolo, 1962) to 30 aphasic patients on three occasions over an eight day period and observed high test-retest reliability (r_s=.91 to .98). His sample was divided evenly into two groups: "recent," or one to three and one-half months postonset and "older," or four months to 23 years postonset. Although Gallaher concluded that time postonset did not noticeably affect reliability of performance, he did report that the sample with recent onset showed slightly higher reliability coefficients than those with less recent onset. Test-retest reliability measures were not performed on the Boston Diagnostic Aphasia Test (Goodglass and Kaplan, 1972), however, the authors of this test say:

> "Reliability in the sense of repeatability of results on retesting
> any patient, varies among aphasics to a degree rarely found in
> other types of patients....Once recovery has stabilized, however,
> the majority of aphasics will, on retest, repeat their original
> performance very closely" (p.12).

Unfortunately, a specific point in time in which aphasic patients become "stable" has not been established.

The test battery selected for this study was the Porch Index of Communicative Ability, or PICA. The PICA consists of 18 subtests which are categorized into three modalities: Gestural, Verbal and Graphic. The integrity of the PICA as a valid reflection of aphasic behavior on a limited sample of language tasks has been documented (1967). The PICA has been used in clinical research as a basis for recording and interpreting changes in the language status of aphasic patients following treatment intervention (Horner and LaPointe, 1979; Wertz, 1978).

The PICA test-retest coefficient of reliability of .99 reported by Porch (1967) was based on 40 aphasic patients with left hemisphere brain injury, only five of whom were beyond 1 year postonset. Twenty-nine of the 40 patients were within 6 months postonset and eight of these were within 1-4 weeks postonset. A slight improvement in the retest scores was reported by Porch; he attributed this improvement either to practice effect or spontaneous recovery. Porch found no strong relationship between the size of test-retest differences and the number of weeks postonset. However, he did observe that patients who were longer postonset demonstrated smaller test-retest differences. The presence or absence of treatment at the time of testing was not reported.

METHOD

Subjects. Two patients participated in this study. Patient A was a 46 year-old male with a ruptured cerebral aneurysm who was tested at 3 years 8 months postonset. Patient B was a 52 year-old male with left hemisphere cerebral vascular accident who was tested at 5 years 10 months postonset. Each patient reported that he had received some form of speech/language treatment in the past, but had never taken the PICA. Neither patient had received speech/language treatment for at least six months prior to this study.

Procedure. The PICA was administered three times to each patient at two week intervals. The patients received no feedback from the examiner regarding test results and no treatment was administered during the test-retest intervals. The patients were instructed to return in two weeks and were not told that the test would be readministered. All tests were recorded on videotape and scored by a second examiner.

RESULTS AND DISCUSSION

The diagnostic baseline patterns, consisting of PICA Overall, High and Low percentile scores for each test administration, are illustrated in Figure 1. The OA%ile score (the solid middle line) for Patients A and B increased by 6%ile and 7%ile units respectively between Tests 1 and 2. Between Tests 2 and 3, the OA%ile score of Patient A plateaued while the OA%ile score of Patient B continued to increase.
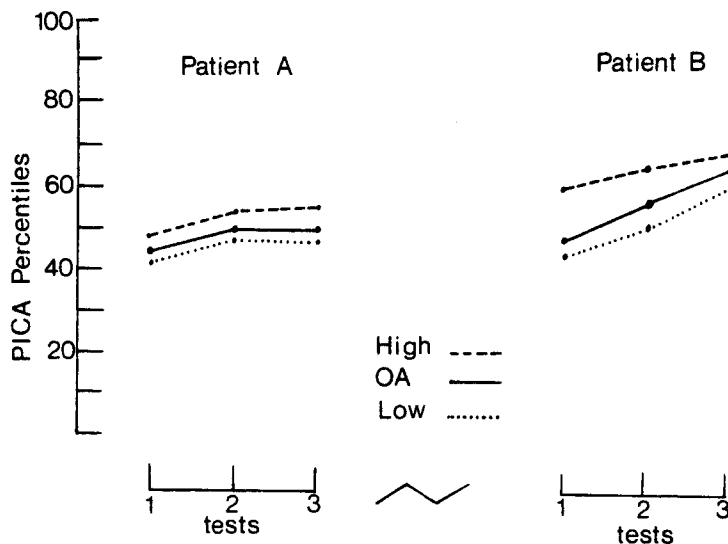

Figure 1. PICA diagnostic baseline scores

More specifically, the OA%ile score for Patient A increased from the 44%ile to the 50%ile for a total gain of 6%ile units. The OA%ile score for Patient B progressively increased from the 49 to the 56 to the 64%ile, which reflects a total increment of 15%ile units.

The Modality percentile scores are shown in Figure 2. It should be noted that while the OA%ile score for Patient A had plateaued between Tests 2 and 3, the Gestural modality score (solid line) had increased by 5%ile units. Both patients demonstrated the greatest variability across tests in the Gestural modality scores and the least variability in the Verbal modality scores.

Although additional testing was planned for both patients, some clinical realities prompted a change in strategy. We had noted to our satisfaction that test performance was not stable across testing sessions with these long term aphasic patients. However, both patients indicated
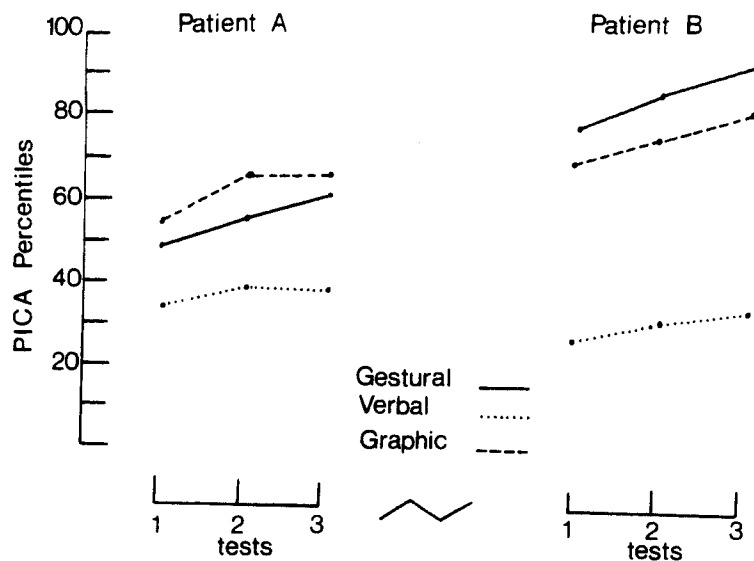
Figure 2. PICA modality percentile scores

that they were not interested in returning unless "therapy" would be initiated. Therefore, we chose to initiate treatment with these two patients and to modify our diagnostic approach with future patients.

Revised diagnostic procedure. Our experience with the two patients discussed revealed the need for some procedural changes. Thus, the pro-tocol for initiating treatment with aphasic patients beyond 1 year post-onset has been revised to include:

1. Observation of a stable diagnostic pattern, consisting of at least three consecutive test administrations at periodic intervals.

2. The test-retest interval will be consistent for a given patient, but may be scheduled for one week or two week intervals.

3. Both patient and family will be apprised of the test-retest schedule. They will be told that testing will be continued until three consecutive stable scores are achieved.

4. Both patient and family will be given immediate feedback on test results.

5. We will instruct the patient and family on the variables that may influence language behaviors. They will be told that we expect to see changes in the initial tests as the patient becomes accustomed to the clinician, the tasks, and the environment. In addition, these changes will reflect the patient's ability to manipulate his own communication and become an active agent in the therapeutic process. Both patient and family will be encouraged to maximize their communicative efforts at home. We will explain to them that the three stable scores indicate when they are ready to start a formal treatment program. At that point, a treatment program suitable for their performance will be employed. Thus, treatment intervention can be more appropriately reflected on subsequent test occasions.

CONCLUSION

Both patients exhibited a change in performance on repeated administrations of the PICA. What are the implications? Let us examine some of the factors that might influence these results.

1. Given that a test is highly reliable, can a score from a single administration be viewed as an accurate and stable index of an individual's functioning?

    No. Results of the previously cited body of research with measures of intelligence, and our own findings, indicate that an individual's level of performance can change over repeated test administration—even for the most reliable of measures.

2. What factors contribute to improved performance on serial testing?

    A partial list of potential confounding variables may include: time postonset, spontaneous recovery, severity of aphasia, motivation, coping behaviors, similarity between treatment procedures and test design, previous diagnostic and treatment experience, and test-taking skills.

3. What is the normal amount of variability that can be expected on repeated test administration?

    An increase of 6 to 12 points has been observed in the IQ tests of children on repeated testing (Haskins, 1978). Porch (1980) has suggested that a 10%ile gain on the PICA OA%ile score is clinically significant. He contends that a 3%ile fluctuation in the OA%ile score represents normal variability.

4. What can be done to reduce variability in performance across testing sessions?

    One thing that can be done is to attempt to minimize retest effects. For example, long term aphasic patients entering a treatment program may indeed perform on test-taking skills in a manner similar to the nonaphasic population previously noted. The effects of a standard pretest format or the presentation of demonstration items, as recommended by Aten and Lyon (1978), should be investigated to avoid unjust penalties.

    The Gestural modality percentile scores between Tests 1 and 3 for the two patients presented in this paper show a gain of 14%ile units for Patient A and 21%ile units for Patient B. Thus, a standardized pretest of tasks comparable to the Gestural modality subtests may minimize the test-retest effect.

5. Does it matter?

    Lack of stable performance across diagnostic sessions, even when the most reliable standardized tests are employed, is regularly demonstrated in the research literature. Factors which contribute to this lack of stability should continue to be scrutinized. In the meantime the clinician is faced with routine decisions involved in interpreting diagnostic results to determine candidacy for treatment, prognosis and treatment planning. Evaluation of diagnostic results can be interpreted with greater precision when patterns on baseline measures are available. Clinicians should not rely on the interpretation of a single test administration on left hemisphere aphasic patients in the "chronic zone."

REFERENCES

Aten, J.L. and Lyon, J. 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, 1978.

Anastasi, A. Psychological Testing. New York: MacMillan, 1973.

Broida, H. Language therapy effects in long term aphasia. Archives of Physical Medicine and Rehabilitation, 56, 248-253, 1977.

Catron, D. and Thompson, C. Test-retest gains in WAIS scores after four retest intervals. Journal of Clinical Psychology, 35, 352-357, 1979.

Davis, G.A. The clinical application of withdrawal, single-case research designs. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1978. Minneapolis, MN: BRK Publishers, 1978.

DeRenzi, E. and Vignolo, L.A. The Token Test: A sensitive test to detect receptive disturbances in aphasia. Brain, 85, 665-678, 1962.

Gallaher, A. Temporal reliability of aphasic performance on the Token Test. Brain and Language, 7, 34-41, 1979.

Goodglass, H. and Kaplan, E.F. Boston Diagnostic Aphasia Test. Philadelphia: Lea and Febiger, 1972.

Haskins, R., Ramey, C., Stedman, D., Blacher-Dixon, J. and Pierce, J. Effects of repeated assessment on standardized test performance by infants. American Journal of Mental Deficiency, 83, 233-239, 1978.

Hersen, J. and Barlow, D.H. Single-Case Experimental Design: Strategies for Studying Behavior Change. New York: Pergamon Press, 1976.

Horner, J. and LaPointe, L.L. Evaluation of learning potential of a severe aphasic adult through analysis of five performance variables. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1979. Minneapolis, MN: BRK Publishers, 1979.

Kangas, J. and Bradway, K. Intelligence at middle age: A thirty-eight year follow-up. Developmental Psychology, 5, 333-337, 1971.

LaPointe, L.L. Multiple baseline designs. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1978. Minneapolis, MN: BRK Publishers, 1978.

Matarazzo, R.G., Matarazzo, J.D., Gallo, Jr., A.E. and Wiens, A. IQ and neuropsychological changes following carotid endarterectomy. Journal of Clinical Neuropsychology, 1, 97-116, 1979.

Matarazzo, R.G., Wiens, A.N., Matarazzo, J.D., and Manaugh, T.S. Test-retest reliability of the WAIS in a normal population. Journal of Clinical Psychology, 29, 194-197, 1973.

Porch, B.E. Porch Index of Communicative Ability, Volume 1, Theory and Development. Palo Alto, California: Consulting Psychologists Press, 1967.

Porch, B.E. Workshop on Porch Index of Communicative Ability. Albuquerque, New Mexico, 1969.

Porch, B.E. Personal Communication, 1980.

Schuell, H., Jenkins, J. and Jimenez-Pabon, E. Aphasia in Adults. New York: Harper and Row, 1964.

Wertz, R. Appraisal and diagnosis of aphasia: Evaluating the effects of treatment. In M. Sullivan and M.S. Kommers (Eds.), Rationale for Adult Aphasia Therapy. University of Nebraska Medical Center, 1977.

Wertz, R.T., Collins, N.J., Weiss, D., Brookshire, R.H., Friden, T., Krutzke, J.F. and Pierce, J. Veterans Administration cooperative study on aphasia: Preliminary report on a comparison of individual and group treatment. Paper presented to the American Speech and Hearing Association, 1978.

DISCUSSION

Q: How long post onset was Patient B?
A: Six years.

Q: It seems that the best thing you can do is keep giving him the PICA over and over. You keep giving him the same test and he's jumping all those %iles. Just don't do anything else.
A: (laughter)

Q: The patients were obviously not thrilled with it. Did you happen to sample the loved ones to find out if the patients were doing better as a result of three test administrations?
A: The wife of Patient A came with him each time but was relatively unresponsive, Patient B had no family in town.
Comment: Why didn't you give her the PICA?

Q: Had either of these patients ever had a PICA before?
A: No. Both patients denied having seen the test items.

Q: I suspect there is a difference between patients who have had "25" PICA's before and the patient who has never taken the PICA.
A: Yes, I agree. And the same may apply for patients who have been in any active treatment program or who have had any type of standardized test. At some point, there is a "first test." If it occurs beyond 1 year postonset, I recommend obtaining a baseline.

Q: Did you measure the reliability of the examiners to account for possible drift or examiner bias?
A: Videotapes were scored by a second examiner and there was good agreement.

Q: What kind of changes have you observed with the second patient following treatment?
A: Patient B was at the 64 OA %ile on his third test in this series, and scored at the 73 OA %ile on his next test 1 month later, which follows the trend. One question we have to ask is "What is the purpose of testing?" We would have looked real good in treatment had we adminis-tered just one test.

Q: Do you think that the inability to find a stable baseline is a bad thing?
A: No. But when the baseline is not stable, I think there are some questions that need to be asked.

Q: I think the patient with the unstable baseline is trying to tell you something, that he's ready to go. The guy is showing an unstable baseline and is saying that if you test me and can plot a gradient through my baseline scores, especially if they are going up (and that this is what you would expect through the treatment phase) that my performance will follow this gradient. Maybe the guy with the ascending or fluctuating baseline might be a better treatment candidate than the one that is stable.

A: The Peak-Mean-Difference (PMD) scores decreased across tests for both patients, implying that the range of variability diminished. Patient A went from a PMD of 522 on Test 1, to 466 on Test 3, which is a loss of 56 points. Patient B went from 465 on Test 1 to 359 on Test 3, for a loss of 106 points.

Q: Are you also collecting measures of nonverbal intelligence to correlate stability across measures? It may not be necessary to give the patients 3 PICA's, which is a costly experience, and sometimes an alienating one.

A: No, we are not administering nonverbal intelligence tests. In response to the remainder of your question, if test scores fluctuate, yet are going to be interpreted as a reflection of the patient's changing language status over time, then it may not be so unkind to administer a test a second or third time, since the very fluctuation in performance should provide valid treatment information.

Q: I think it would be interesting to take this one step futther and teach the test to see if you can effect a greater slope in change by some purposeful intervention.

A: As I recall, Porch has reported teaching the PICA to a woman who had stabilized in treatment. Her test scores improved following a training period on the test. However, it appeared to be an artificial increment, because after a 2-week period with no training, her test performance returned to the pretraining level.

Q: Had either of these patients received therapy before?

A: Both patients reported that they had been in and out of treatment programs in the past. I was not able to determine the nature of the treatment.

Q: I think that for a question like this it would be particularly valuable to have a pretty large number of subjects. I think that what this does is point out the possibility of the problem, but not necessarily the existence of a general problem. These two people could have had fairly unusual circumstances, especially since the prior treatment was unknown. The psychological adjustment, being exposed to a new therapist, and possibility that someone can help you, may make a difference.

A: This _did_ happen with two patients. To apply this procedure to a large group of patients masks the very issue--that individuals may vary across testing. Further, this is usually done on larger groups for test-retest reliability measures, which are an index of stability of test scores, in which the individual's score is referenced against a group mean. In that situation, everyone could improve by 10 points and yield a reliability of 1.00. However, the information I want in treatment is, how does this individual perform? This is not limited to the PICA; it applies to any of our tests. When a long-term aphasic patient comes in, you may want to approach them in a slightly different way, that is, get a baseline.