

Reliability Studies Related to the Personal Life Quality Protocol and Component Scales

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

Conroy, J. (1995, January, Revised December). *Reliability of the Personal Life Quality Protocol. Report Number 7 of the 5 Year Coffelt Quality Tracking Project*. Submitted to the California Department of Developmental Services and California Protection & Advocacy, Inc. Ardmore, PA: The Center for Outcome Analysis.

Executive Summary: This study of the reliability properties of the Personal Life Quality Protocol (PLQP) has investigated test-retest, interrater, and internal consistency for many of the most important outcome indicators in the package. The results have shown that basic demographic information and simple quality items are being collected accurately. Furthermore, most of the major indicators and scales display extremely good reliability characteristics. The scales of adaptive behavior, challenging behavior, and choice-making are particularly strong.

The way the study was designed produced very conservative estimates of reliability, because test-retest and interrater aspects of measurement error were combined. However, it was possible to separate out the test-retest from the interrater aspects to some degree, following the advice of Devlin (1989). This approach led to three indicators for each important scale:

- the raw correlation, in which test-retest and interrater sources of error were combined,
- the pure test-retest correlation (where respondents at Time-1 and Time-2 were identical), and
- the pure interrater correlation (calculated by a formula which presumes that any error not due to instability over time must be due to lack of agreement across respondents).

Table 5 summarizes the results of these analyses.

**Table 5
Summary of Reliability Findings**

Dimension	Raw Correlation (Confounded)	Same Respondent (Test-Retest)	Corrected (Inter- Rater)
Adaptive Behavior	0.973	0.996	0.977
Challenging Behavior	0.866	0.999	0.867
Choice-Making	0.859	0.983	0.876
Reported Progress on Goals	0.620	0.668	0.952
Day Program Hours	0.696	0.932	0.764
Earnings	0.668	0.999	0.669
Integration Scale	0.440	0.446	0.994
Quality of Life Then	0.765	0.930	0.835
Quality of Life Now	0.757	0.963	0.794

The two columns to the right represent the 'pure' estimates of test-retest and interrater reliability. The results are generally very high, indicating acceptable reliability of most of the measures.

In addition to the scales represented in Table 5, data on developmentally oriented services rendered appear to be reliable across time and Visitors.

There are two problems, and both are in the test-retest area. The Reported Progress on Goals does not seem to be as stable as other measures over time (test-retest .668), although it is apparently strong on the interrater measure. The second problem is with the Integrative Activities scale, which displays exactly the same problem. Further work with these scales in

community settings will be needed. Greater variety in type of class members, types of lifestyles, and types of respondents will be necessary to adequately test these two scales and ascertain the causes of any psychometric weakness.

In summary, this study has supported the inference that the Coffelt project data are generally being collected accurately, objectively, and reliably.

Conroy, J. (1980). *Reliability of the Behavior Development Survey* (Technical Report 80-1-1). Philadelphia: Temple University Developmental Disabilities Center.

Found reliability of the behavior scales to be above .80, with adaptive behavior even higher.

Conroy, J., Efthimiou, J., & Lemanowicz, J. (1981). *Reliability of the Behavior Development Survey: Maladaptive behavior section* (Pennhurst Study Brief Report No. 11). Philadelphia: Temple University Developmental Disabilities Center.

Reexamined the reliability properties of the maladaptive behavior section of the BDS, and found acceptable interrater reliabilities and considerably higher test-retest scores.

Devlin, S. (1989). *Reliability assessment of the instruments used to monitor the Pennhurst class members*. Philadelphia: Temple University Developmental Disabilities Center.

Abstract: The goal of this evaluation was to determine the internal consistency, test-retest and inter-rater reliability of the five instruments (BDS Adaptive, BDS Maladaptive, NORM, PQ, GHMS and LS scales) used by Temple University's Developmental Disabilities Center to monitor the progress of the Pennhurst Plaintiff Class members. Twenty-nine class members, who were living in community living arrangements were randomly selected to serve as the subjects for this study. The data suggests that the majority of these instruments provide a reliable means of monitoring the progress individuals with developmental disabilities. Recommendations are made for improving the reliability of the scales through more structured training of the data collectors.

The purpose of the present study was to assess the test-retest reliability, inter-rater reliability and internal consistency of the instruments used by Temple University's Developmental Disabilities Center for the past 11 years. In 1978 Judge Raymond J. Broderick, who was appointed Special Master in the Pennhurst case ordered that data be gathered on the status of every individual living in Pennhurst, a state institution for adults with developmental disabilities. This information was then used to plan for the development of community residences for the Pennhurst residents, following the District Court decision to close Pennhurst. Since 1978 the instruments have been used as a means for monitoring the status of the former residents of Pennhurst who are now living in a variety of community residential programs throughout Pennsylvania.

Fullerton, A. Douglass, M. & Dodder, R. (1996). *A systematic study examining the reliability of quality assurance measures*. Report of the Oklahoma State University Quality Assurance Project. Stillwater, OK.

In a nested design across settings and types of people, reliability of the COA adaptation of instruments for Oklahoma was investigated. Reliability on all scales was found to be acceptable, although some items in the health section were not stable over time. Reliability varied significantly from one year to the next, but in general, the levels of reliability were high and the authors concluded that the methodology was worthy of continuation.

Fullerton, A. Douglass, M. & Dodder, R. (1999). A reliability study of measures assessing the impact of deinstitutionalization. *Research in Developmental Disabilities, Vol. 20, No. 6*, pp. 387-400.

Published version of the report above.

Dodder, R., Foster, L., & Bolin, B. (1999). Measures to monitor developmental disabilities quality assurance: A study of reliability. *Education and Training in Mental Retardation and Developmental Disabilities, 34, 1*, 66-76.

Report of a conservative exploration of interrater and test-retest reliability of seven major scales developed by Conroy et al. Found acceptable reliabilities overall and recommended continued utilization of the scales in quality assurance activities.

Harris, C. (1982). *An interrater reliability study of the Client Development Evaluation Report*. Final report to the California Department of Developmental Services.

Found the behavior scales of the CDER to display acceptable reliabilities, with the adaptive behavior section showing exceptionally high interrater reliability.

Isett, R., & Spreat, S. (1979). Test-retest and interrater reliability of the AAMD Adaptive Behavior Scale. *American Journal of Mental Deficiency*, 84, 93-95.

Calculated test-retest and interrater reliabilities for all domains of the American Association on Mental Deficiency Adaptive Behavior Scale. Part 1 domains evidenced generally adequate estimates of both within- and between-rater variability. The domains on Part 2 of the scale were less reliable than those of Part 1, particularly with reference to interrater reliability. The low Part 2 interrater reliability coefficients raise questions concerning the use of Part 2 of the instrument.

Jagannathan, R., Camasso, M., Lerman, P., Hall, D., & Cook, S. (1997). *The New Jersey Client Assessment Form: An Analysis of Its Stability Over Time*. Newark, NJ: Developmental Disability Planning Institute, New Jersey Institute of Technology.

An independent Rutgers University research group adopted the COA instruments to continue study of the deinstitutionalization process begun by COA in New Jersey. The Rutgers group reported high stability (test-retest) and internal consistency for the instruments.

Lemanowicz, J., Feinstein, C., & Conroy, J. (1980). *Reliability of the Behavior Development Survey: Services received by clients*. Pennhurst Study Brief Report 2. Philadelphia: Temple University Developmental Disabilities Center/UAP.

Compared data collected by Temple University group to data collected by Pennhurst human resources staff on the type and amount of services received by people. The Temple group collected data by staff interview plus records scrutiny. The Pennhurst staff collected data by direct observation and time sampling. The definitions of each service differed in some cases, but the total amount of developmentally oriented services received by each person was correlated at the level of .92 between the two methods.