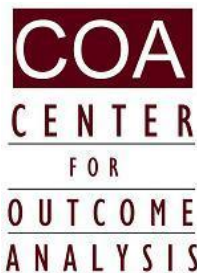


# Personal Life Quality Protocol

## Short, Reliable Outcome Measurement Tools for Quality Tracking in Developmental Disabilities Systems



*Center for Outcome Analysis*

[www.eoutcome.org](http://www.eoutcome.org)

610.246.5961 email [jconroycoa@gmail.com](mailto:jconroycoa@gmail.com)

# Table of Contents

<b>Overview</b>	<b>3</b>
<b>Commentary on Other Approaches to Quality</b>	<b>5</b>
<b>Behavior</b>	<b>6</b>
<b>Perceived Quality of Life – Individual Interviews</b>	<b>10</b>
<b>Person-Centered Planning</b>	<b>13</b>
<b>The Integrative Activities Scale</b>	<b>15</b>
<b>Choice Making and Autonomy in Daily Life</b>	<b>17</b>
<b>Productivity and Meaningful Day Activities</b>	<b>19</b>
<b>Close Relationships</b>	<b>21</b>
<b>Psychometrics of Quality of Life Measurement</b>	<b>23</b>

## Overview

These simple short one page scales were designed to detect changes in qualities of life when people move from large institutions to small community homes. They have been shortened over the years to be more and more practical to use with hundreds or thousands of people. These are the most refined and tested instruments for this purpose extant. They have been used in more than 30 large scale studies, and have been adapted for use in other countries including translations into French, Georgian, Korean, and Spanish. They have also been adapted for use with children, elders, and people with other mental and physical disabilities. They naturally extend from research purposes into long term monitoring and tracking of quality of life and services.

Each scale can be used independently – they are “modular.” Each monitoring or tracking project can use the entire set, or can select the ones most appropriate for local interests. Other measures from other sources can be added to these tools with no ill effect.

The elements of quality of life addressed in these scales are “things that matter” to people with disabilities themselves. Over the course of more than 150,000 face to face interviews with these instruments, we have painstakingly learned what matters to people and to their families. Each individual, with or without disabilities, self-defines quality of life in idiosyncratic ways, with varying combinations and emphasis among the components – and this may change within an individual across the life span. Tracking qualities of life must not be portrayed as easy or simple. And scientific humility requires recognition of the fact that we cannot measure every aspect of quality of life. For example, self-esteem and romantic love are far beyond our ability to tap reliably among people who have limited means of linguistic expression. Nevertheless, after decades of work, the essence of what we do know has been distilled into a handful of relatively simple measurement tools about the things that people say matter the most for their well-being, satisfaction, fulfillment, and, ultimately, happiness.

Via studies with tens of thousands of unique individuals and across many years, the most important dimensions that have emerged include having friends, relationships, being “part of” the surrounding community; being engaged in meaningful daily activities – including working; having some money to do things; making choices, running one’s own life; becoming more and more independent; being able to behave in ways that do not harm or disturb self or others; being supported in a person-centered and respectful ways; and of course being healthy and free of pain.

The entire package contained here can be collected in an average of 60 minutes after training and practice. Sometimes, with some people, it will take much longer, but the average turns out at about an hour.

The scales are used as interviews with each person, but since many of the people we have tracked in dozens of studies do not use verbal language, the help of a third party is appropriate – as long as it is someone who knows the person very well on a day to day basis.

Many studies of reliability have been performed over the years. The scales, though short, perform quite well on internal consistency reliability, test-retest reliability, and inter-rater reliability. The most recently rewritten scale, on Closest Relationships, has not been rigorously tested yet – and yet it may well be the most important aspect of quality of life. This should remind us that our measurement approaches are far from perfect, that they are still evolving and improving, and that the tools offered here can be combined with other tools at will. (Because this scale is relatively new, it has not been formatted in the “Then and Now” layout. We will do so after rigorous application and testing.)

The scales here are set up as “Then and Now” form, so that people can be asked what their lives were like “Then” – when still living in an institution, for example – and “Now” – living in a community based home. When real “Before and After” or “Pre-Post” measurements are possible, this is the much preferred method. In genuine Pre-Post studies, the column’s for “Then” can simply be removed to avoid confusion. The scales are collected when people are still in the pre-change situation, e.g. institution, and then done all over again at least 6 months and preferably a year after moving.

## Commentary on Other Approaches to Quality

1. The Personal Outcome Measures of the Council on Quality & Leadership for People with Disabilities. These are most useful for accreditation visits that teach leaders and workers to understand and emphasize the most important elements of good lives. However, psychometrically, they have little merit. The approach boils down into 25 “Yes/No” items that are not at all useful for accurate measurement of changes over time. The developers have offered some limited reliability investigations, but never about the most important question: will two different POMs raters who visit the same program produce similar ratings of quality? This is inter-rater reliability. Without inter-rater reliability, “we may just as well stay home and flip coins” – because the entire enterprise will be unscientific, based on the various raters’ predilections and biases. Some raters are “Hard Graders” and some are “Easy Graders” – hence POM scores received depend on which rater shows up, not on the real quality of life of the person or in the program being visited. The POMs have been used to fully accredit large public institutions in the U.S. and elsewhere. We cannot seriously believe that they accurately measure the values and qualities of life that are central to modern human service ideals.

2. Behavioral Measures. Independent functioning (also called adaptive behavior, self-care, activities of daily living, and so on) is the easiest of all dimensions to measure accurately. There are hundreds of scales available and nearly all of them are simple and reliable. Any can be substituted for the ones we include here – though the ones here are among the simplest and most rigorously tested. Challenging behavior is an essential measure too, and there are decades of study about how to collect data on this dimension accurately. Subjectivity influences the measurement in known and significant ways, particularly cross-culturally. Schemes have adopted measures based on frequency or on severity and have been compared. Overall, the instruments for independent and challenging behaviors presented here are among the easiest to collect and most reliable. It should be noted that reliability in the dimension of challenging behavior has always been found to be lower than for independent functioning measures, but ratings are now within acceptable levels, and it is such an important element of life and supports that it must be included.

3. National Core Indicators of the Human Services Research Institute and others. The NCI instrument is simply not designed to track individual / group changes over time. They are not intended to, nor should they be used, to assess changes in quality of life during deinstitutionalization and community living. The NCI items are too simple and too crude to be used to find out – for example – whether people gain in their power to determine their own lives – since there are only four “Yes/No/Don’t Know” items in the NCI package, rather than a carefully designed and tested psychometric scale like the Decision Control Inventory here. The NCI is designed for large scale cross-system comparisons, not for detailed study of the benefits of system changes from one year to the next.

## **Behavior**

The behavioral measures are derived from Warren Bock's Minnesota Developmental Programming System, which is easier to use than the shortened forms of the original AAMR Adaptive Behavior Scale (Nihira, Foster, Shellhaas, & Leland, 1974). The two instruments are correlated at .94, hence they are practically interchangeable. The first part contains 80 quick and simple ratings of adaptive behavior, and the second, 20 items on the frequency of challenging behaviors. The measures were shortened on the basis of the mathematical criteria of factor structure and reliability. According to Arndt (1981), the best way to treat these type of data is as two simple additive scales, one reflecting adaptive behavior and the other challenging behavior. The adaptive behavior sum score has been found to be highly reliable (Isett & Spreat, 1979; Spreat, 1980; Devlin, 1989). Devlin found interrater reliability of .95 and test-retest reliability of .96. For the maladaptive behavior section, interrater reliability was .96 and test-retest was .78.

In some of our data sets (California, Florida, and New Hampshire), the California behavior scales called the Client Development Evaluation Report were used. This behavior measure is composed of 52 items. The CDER adaptive behavior measure has been reported to have good reliability under certain circumstances (Harris, 1982).

Adaptive behavior is usually measured by interviewing the person and/or a third party who knows the individual very well on a day to day basis.

# Behavior Part 1: Bock Developmental Scale

<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>99</b>
<b>Needs total support to accomplish</b>	<b>Needs major support to accomplish</b>	<b>Needs minor support to accomplish</b>	<b>Needs no support to accomplish</b>	<b>Unknown, N/A</b>

#	THEN	NOW	Kind of Behavior
1	0 1 2 3 99	0 1 2 3 99	Sits without support
2	0 1 2 3 99	0 1 2 3 99	Moves from lying down on stomach to a sitting position
3	0 1 2 3 99	0 1 2 3 99	Pulls self to standing position using person or prop for support
4	0 1 2 3 99	0 1 2 3 99	Walks 5 feet ( may use braces or crutches)
5	0 1 2 3 99	0 1 2 3 99	Walks upstairs and downstairs, putting both feet on each step (may use wall or handrail for support)
6	0 1 2 3 99	0 1 2 3 99	Walks a straight line 10 feet
7	0 1 2 3 99	0 1 2 3 99	Runs
8	0 1 2 3 99	0 1 2 3 99	Walks upstairs and downstairs alternating feet (may use wall or handrail for support)
9	0 1 2 3 99	0 1 2 3 99	Jumps up, both feet off the floor at once
10	0 1 2 3 99	0 1 2 3 99	Stands on tiptoe for 10 seconds
11	0 1 2 3 99	0 1 2 3 99	Eliminates when on the toilet (bowel or bladder)
12	0 1 2 3 99	0 1 2 3 99	Goes to the bathroom with a reminder
13	0 1 2 3 99	0 1 2 3 99	Has bowel control
14	0 1 2 3 99	0 1 2 3 99	Remove clothing, sits on the toilet, and eliminates and replaces clothing
15	0 1 2 3 99	0 1 2 3 99	Has bowel and bladder control
16	0 1 2 3 99	0 1 2 3 99	Goes to the bathroom independently
17	0 1 2 3 99	0 1 2 3 99	Flushes toilet after use
18	0 1 2 3 99	0 1 2 3 99	Obtains help with and toileting problems
19	0 1 2 3 99	0 1 2 3 99	Ask the location of the bathroom in new location
20	0 1 2 3 99	0 1 2 3 99	Chooses the correct restroom in a public place
21	0 1 2 3 99	0 1 2 3 99	Extends and withdraws arms and legs while being dressed and undressed
22	0 1 2 3 99	0 1 2 3 99	Removes slip-over shirt
23	0 1 2 3 99	0 1 2 3 99	Places toothbrush in mouth and begins brushing motion
24	0 1 2 3 99	0 1 2 3 99	Puts shoes on correct feet
25	0 1 2 3 99	0 1 2 3 99	Soaps and rinses hands
26	0 1 2 3 99	0 1 2 3 99	Blows nose in tissue or handkerchief
27	0 1 2 3 99	0 1 2 3 99	Dries entire body with towel after bathing
28	0 1 2 3 99	0 1 2 3 99	Puts on outer wear without reminder
29	0 1 2 3 99	0 1 2 3 99	Washes, rinses and dries hair
30	0 1 2 3 99	0 1 2 3 99	Changes dirty clothing without reminder
31	0 1 2 3 99	0 1 2 3 99	Drinks, without spilling, from a glass or cup with assistance
32	0 1 2 3 99	0 1 2 3 99	Picks up food with fingers and puts food in mouth
33	0 1 2 3 99	0 1 2 3 99	Uses spoon to pick up and eat food
34	0 1 2 3 99	0 1 2 3 99	Picks up a glass and drinks from it without spilling
35	0 1 2 3 99	0 1 2 3 99	Eats a complete meal with little or no spilling (may use only fingers and spoon)
36	0 1 2 3 99	0 1 2 3 99	Uses a fork to pick up and eat food
37	0 1 2 3 99	0 1 2 3 99	Eats, supervised in public without calling attention to eating behavior
38	0 1 2 3 99	0 1 2 3 99	Eats a complete meal with little or no spilling using all normal dishes and utensils
39	0 1 2 3 99	0 1 2 3 99	Serves self in a family-style setting
40	0 1 2 3 99	0 1 2 3 99	Order and eats in public dining facility
41	0 1 2 3 99	0 1 2 3 99	Obeys a simple instruction such as , “come here”
42	0 1 2 3 99	0 1 2 3 99	Stops and activity upon request such as “No.” or “Stop.”
#	THEN	NOW	Kind of Behavior

43	0 1 2 3 99	0 1 2 3 99	Says or indicates, "Yes," or "No." In response to questions such as, Do you want to go out?"
44	0 1 2 3 99	0 1 2 3 99	Names 10 common objects when asked, "What is this?"
45	0 1 2 3 99	0 1 2 3 99	Says first and last name when asked
46	0 1 2 3 99	0 1 2 3 99	Expressed feelings, desires or problems in complete sentences = (subject - verb) = such as, " I am
47	0 1 2 3 99	0 1 2 3 99	Speaks in phrases or sentences clearly enough to be understood by someone not familiar with the person
48	0 1 2 3 99	0 1 2 3 99	Carries on a conversation with another person(s) for 10 minutes
49	0 1 2 3 99	0 1 2 3 99	Says address of residence clearly when asked
50	0 1 2 3 99	0 1 2 3 99	Invites others to participate in an activity such as going for a walk or going to a movie
51	0 1 2 3 99	0 1 2 3 99	Scribbles with chalk, pencil or crayon
52	0 1 2 3 99	0 1 2 3 99	Turns the pages in a book one at a time
53	0 1 2 3 99	0 1 2 3 99	Marks on a chalkboard or paper in circles and lines
54	0 1 2 3 99	0 1 2 3 99	Traces with pencil or crayon around the outside of a six-inch circular object in a continuous motion
55	0 1 2 3 99	0 1 2 3 99	Draw a line connecting 3 dots on a piece of paper
56	0 1 2 3 99	0 1 2 3 99	Follows printed material left to right
57	0 1 2 3 99	0 1 2 3 99	Writes or prints first and last name with no example to look at
58	0 1 2 3 99	0 1 2 3 99	Reads aloud the alphabet from A to Z (may look at letters)
59	0 1 2 3 99	0 1 2 3 99	Reads aloud sentences with 5 common words
60	0 1 2 3 99	0 1 2 3 99	Reads for information or entertainment
61	0 1 2 3 99	0 1 2 3 99	Separates one object from a group upon request, "Give me 1 block, ect."
62	0 1 2 3 99	0 1 2 3 99	Sorts coins from other small metal objects
63	0 1 2 3 99	0 1 2 3 99	Answers correctly when asked, "Is it day or night?"
64	0 1 2 3 99	0 1 2 3 99	Points to the short or long line when asked
65	0 1 2 3 99	0 1 2 3 99	Chooses the correct number of objects up to 5 upon request, "Give me 1 block, 2 blocks, etc.
66	0 1 2 3 99	0 1 2 3 99	Answers what day of the week it is now
67	0 1 2 3 99	0 1 2 3 99	Counts from 10 to 20.
68	0 1 2 3 99	0 1 2 3 99	Tells or identifies birth date: month, day and year
69	0 1 2 3 99	0 1 2 3 99	Exchanges the correct number of mixed coins for a quarter
70	0 1 2 3 99	0 1 2 3 99	Counts the change from a purchase of one dollar or less.
71	0 1 2 3 99	0 1 2 3 99	Participates in a single activity for 10 minutes (if protected from interruption)
72	0 1 2 3 99	0 1 2 3 99	Goes to public places in a supervised group without calling unfavorable attention to behavior
73	0 1 2 3 99	0 1 2 3 99	Attends to an assigned task or activity for one-half hour (may need to be encouraged)
74	0 1 2 3 99	0 1 2 3 99	Sweeps a floor with a broom, picks up sweepings in a dustpan, and empties pan.
75	0 1 2 3 99	0 1 2 3 99	Sets a table with plates, cups, forks, spoons, and knives (need not be a formal setting)
76	0 1 2 3 99	0 1 2 3 99	Crosses residential street intersections, looking and waiting for traffic to clear before crossing
77	0 1 2 3 99	0 1 2 3 99	Prepares a meal of a sandwich and cold beverage
78	0 1 2 3 99	0 1 2 3 99	Obeys lights and "WALK," "DON'T WALK" signals at a light controlled intersection
79	0 1 2 3 99	0 1 2 3 99	Initiates self-involvement in a hobby not including reading or watching TV
80	0 1 2 3 99	0 1 2 3 99	Uses public transportation on one local route such as from residence to work and back

1-10 Gross Motor Development, 11-20 Toileting Skills, 21-30 Dress/Grooming Skills, 31-40 Eating Skills, 41-50 Language Skills, 51-60 Read Write Skills, 61-70 Quantitative Skills, 71-80 Independent Living Skills



## Behavior Part 2: Challenges

This scale should be asked of whoever know the person best on a day-to-day basis.

### Problem Severity: Respondent's opinion

<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>None or None Known</b>	<b>Mild</b>	<b>Moderate</b>	<b>Major</b>	<b>Extreme</b>

	THEN	NOW	Kind of Behavior
1	0 1 2 3 4	0 1 2 3 4	Self-Injury
2	0 1 2 3 4	0 1 2 3 4	Assaulting Others
3	0 1 2 3 4	0 1 2 3 4	Threatening Others
4	0 1 2 3 4	0 1 2 3 4	Damaging Property
5	0 1 2 3 4	0 1 2 3 4	Screams or Cries or Yells Inappropriately
6	0 1 2 3 4	0 1 2 3 4	Unusual or Repetitive or Stereotyped Behaviors
7	0 1 2 3 4	0 1 2 3 4	Social Withdrawal
8	0 1 2 3 4	0 1 2 3 4	Depressive Symptoms
9	0 1 2 3 4	0 1 2 3 4	Suicidal Actions, Tendencies, Thoughts
10	0 1 2 3 4	0 1 2 3 4	Hallucinations/Delusions
11	0 1 2 3 4	0 1 2 3 4	Poor Grooming/Cleanliness
12	0 1 2 3 4	0 1 2 3 4	Eating Disorders
13	0 1 2 3 4	0 1 2 3 4	Hyperactivity or Mania
14	0 1 2 3 4	0 1 2 3 4	Anxiety or Panic
15	0 1 2 3 4	0 1 2 3 4	Inappropriate, Illegal, or Dangerous Sexuality
16	0 1 2 3 4	0 1 2 3 4	Substance Abuse
17	0 1 2 3 4	0 1 2 3 4	Stealing
18	0 1 2 3 4	0 1 2 3 4	Lying
19	0 1 2 3 4	0 1 2 3 4	Setting Fires
20	0 1 2 3 4	0 1 2 3 4	Running Away

## Perceived Quality of Life – Individual Interviews

Every “data collection” visit begins with the person. Every visit must include an attempt to speak directly with the focus person, preferably in private. There are hundreds of such interview questionnaires extant. We do not believe that any have proven superior – either in content or psychometrics – to the simple approach that has emerged from our 40 years of work asking these crucial questions.

The way the questions are asked, however, is very important. “Yes/No” questions should be avoided. They are extremely unreliable. The problem of “Response Acquiescence” was noted as early as the 1980s by Sigelman and colleagues (“When in Doubt, Say Yes” is the classic. Much more accurate five point responses can be obtained in a very simple and clear interview method, used and documented by COA research over decades. The general instructions used by COA are reproduced below as an example of procedure.

### INSTRUCTIONS

- These questions must be directed toward the person, but can be answered with help from whoever knows the person best on a day to day basis.
- Preferably in private.
- Keep it informal. Begin with the usual social niceties that you would expect from any visitor to your home. How are you, telling about yourself, comments on the home, etc.
- Any item with 5-point scale answers should be thought of as a "YES-NO" or "GOOD-BAD" 2-point scale, with a chance to get more detail if the person is able. Example: Ask "How do you feel about living here?" and the person answers "Good" then you probe "Would you say Good or Very Good?" If the person answers "I don't know," or "Not sure," or some indefinite answer, probe with "Do you feel on the good or bad side?" If no preference, stick with "Fair," which we will interpret to mean "In Between."
- Tell the person this interview is VOLUNTARY. Say that he/she does NOT HAVE TO talk to you. Even if he/she agrees to the interview, he/she can stop at any time, for any reason.

This scale can be used very appropriately and successfully with families. Comparisons of perceptions can yield crucial insights into the meaning of “better off” to the people themselves and to their relatives. The two are not always the same.

## COA's Quality of Life Changes – A Scale of Perceptions

The simple 14 item scale of quality of life captures the perceptions of the person (and/or whoever knows the person best on a day to day basis) about qualities of life. It can be set up to collect perceptions about life NOW, plus perceptions about life THEN, at some previous time. This can be useful for people who have moved into a new home but no one collected “pre-move” or “baseline” data. This approach is not as accurate as genuine “before and after” measurements, but is quite useful as perceptions of quality, and has been shown to mirror the “before and after” data fairly well.

(There is also an option to collect ratings of how important each dimension of quality of life is to each person. This can reveal the fundamental truth that people differ greatly on what is important to them – which is why quality of life is difficult to “get a handle on” – because it is actually different for everyone. With the importance ratings, it is possible to “weight” each person’s quality of life perceptions by how important each dimension is to that specific individual.)

## Quality of Life Perceptions (To Be Answered by the Person or Whoever Knows the Person Best)

Copyright © J.W. Conroy 2017

Ask the person to rate the qualities of his/her own life “THEN” and “NOW.” For people living at this setting, this means trying to remember what life was like THEN, before they moved here, versus right now. **If the person can't answer, accept answers from whoever knows the person best.** You must find someone who the person will allow to answer, or who knows the person on a day to day basis better than anyone else.

Each quality item is approached as two “Either-Or” questions. For example, the first Either-Or question on the first item is “Would you say your health is good or bad?” (In between is implied, if the person says “neither” or “OK” or “neither” or any similar response. But answers like that have to be checked by probing with “Oh, so it’s in between, not really good or bad?”) Once the person answers, for example, “good,” the follow-up is a second Either-Or question: “Would you say good or very good?”

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>9</b>
<b>Very Bad</b>	<b>Bad</b>	<b>In Between</b>	<b>Good</b>	<b>Very Good</b>	Don't know, N/A

THEN	NOW	
		1 Health
		2 Running my own life, making choices
		3 Family relationships
		4 Relationships with friends
		5 Getting out and getting around
		6 What I do all day
		7 Food
		8 Happiness
		9 Comfort
		10 Safety
		11 Treatment by staff/attendants
		12 Health care
		13 Privacy
		14 Overall quality of life

15. How many of these 14 questions were answered by the Focus Person?

\_\_\_\_\_ (from 0 to 14)

## Person-Centered Planning

The “Elements of the Person-Centered Planning Process” scale taps the degree to which a person’s individual planning process follows the general guidelines of person-centered planning.

Most modern support systems now practice some variety of the “person-centered planning” process first described and elaborated by Beth Mount and colleagues John and Connie Lyle O’Brien.<sup>1</sup> As they stated, the emergence of Person-Centered Planning was founded on:

- Seeing people first rather than relating to diagnostic labels;
- Using ordinary language and images rather than professional jargon;
- Actively searching for a person’s gifts and capacities in the context of community life;
- Strengthening the voice of the person and those who know the person best in accounting for their history;
- Evaluating their present conditions in terms of valued experiences;
- Defining desirable changes in their lives.

The Elements of the Person-Centered Planning Process scale cannot capture all the intensely personal and subtle elements of what it means to put the person’s dreams and hopes at the center of all support planning and delivery – but it does appear to work quite well as a gross index. Moreover, it is in fact sensitive to improvements over time in these “best practices,” as seen in dozens of studies of deinstitutionalization and self-determination. (These are cited at [www.eoutcome.org](http://www.eoutcome.org), ), and many of them can be downloaded; the rest can be requested from COA.)

We compute the scale so that its lowest possible score is 0, and the highest is 100. That way, it is easy to interpret, like a number “grade” or percentile.

---

<sup>1</sup> Mount, B. (1987). *Personal futures planning: Finding direction for change*. (Doctoral dissertation, University of Georgia). Ann Arbor, MI: UMI Dissertation Information Service. Connie Lyle O’Brien and John O’Brien (2000). *The Origins of Person-Centered Planning: A Community of Practice Perspective*. Atlanta: Responsive Systems Associates.

# Elements of the Person-Centered Planning Process, Before and Now

Copyright © James W. Conroy, 2017

Ask the person to rate each element “BEFORE” and “NOW.” BEFORE means before the person got involved in the new program, initiative, or agency being evaluated. If this is routine monitoring for quality, use “A YEAR AGO” instead of “BEFORE.” Phrase each question as “True or Not True” followed by the second probe, such as, “OK, True, but would you say Mostly True or Completely True?”

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>9</b>
<b>Not true at all</b>	<b>A little bit true</b>	<b>Half true</b>	<b>Mostly true</b>	<b>Completely true</b>	Don't know, N/A

<b>How True BEFORE? (Or A Year Ago?)</b>	<b>How True NOW?</b>	<b>Plain wording</b>	<b>More detail and jargon</b>
1B	1N	Planning really includes my dreams.	<i>Strong efforts are made to understand the focus person's long term dreams. (As opposed to short term goals set by others.)</i>
2B	2N	Planning tries to build networks of support from family, friends and community.	<i>The planning process emphasizes building a network of supports from informal, unpaid, or general community sources</i>
3B	3N	Planning meetings are comfortable and relaxed for me.	<i>The meetings are comfortable and relaxed for the focus person. (As opposed to formal and "official.")</i>
4B	4N	Planning meetings happen when we need them, not on some fixed schedule.	<i>Planning sessions are scheduled flexibly, as needed. (As opposed to a regular set schedule, such as annually.)</i>
5B	5N	We decide how to do the planning – we don't have to follow a bunch of rules and regulations.	<i>The planning process is defined by group preferences. (As opposed to defined or regulated by a set of standards, rules, laws, or regulations.)</i>
6B	6N	We try to be creative in planning – thinking of new ideas, new ways to get things done, different approaches.	<i>The planning process encourages creativity, new ideas, different ways of thinking.</i>
7B	7N	Our planning can handle disagreements, we can get past them.	<i>The planning process allows for conflicts and disagreements, and try to resolve them.</i>
8B	8N	Our planning is flexible – we will try a different way if something is not working.	<i>The planning process is flexible, allowing for changes in approach when things do not work.</i>
9B	9N	If others in the planning group can't agree, I have the final word (as long as it's not dangerous or unhealthy).	<i>The person has ultimate authority if able and willing to exercise it. (He or she could overrule the entire group on an issue, within safety limits.)</i>
10B	10N	Cooperation is important in our planning – no one group is 'in charge.'	<i>Did the planning process emphasize cooperation among all participants? (As opposed to professional authority.)</i>
11B	11N	Our planning works a lot on my relationships – friends, colleagues – and includes romance if I want it to.	<i>Does the planning process emphasize the person's relationships? (As opposed to emphasizing skill development, or behavior, or services.)</i>
12B	12N	Money and figuring out how to spend it is a big part of our planning.	<i>Does the planning process take money into consideration? (Does the group discuss what supports cost, and what alternatives there are?)</i>
13B	13N	Our planning group has full control over the money that's used to support me.	<i>Does the planning group have control over the resources (money) devoted to supporting the focus person?</i>
14B	14N	Non-professionals (myself and my freely chosen allies) have most of the power over planning decisions.	<i>Do the unpaid group members have the real power? (As opposed to paid staff and professionals.)</i>
15B	15N	My planning process is person-centered.	<i>Do you consider this plan to be "person-centered"?</i>

## **The Integrative Activities Scale**

The scale used to assess integration was taken from the Harris poll of Americans with and without disabilities (Taylor, Kagay, & Leichenko, 1986). It measured how often people visit with friends, go shopping, go to a place of worship, engage in recreation, and so on, in the presence of non-disabled citizens. The scale tapped only half of the true meaning of integration; if integration is composed of both presence and participation, then the Harris scale reflects only the first part. Presence in the community is a necessary but not sufficient condition for participation in the community. The scale simply counts the number of “outings” to places where non-disabled citizens might be present. The scale is restricted to the preceding month.

Because the scale was developed by Harris, and was used nationally with both disabled and non-disabled Americans, we have national data for comparison. This scale was also used in the National Consumer Survey of 1990 (Conroy, Feinstein, Lemanowicz, Devlin, & Metzler, 1990) with 13,075 Americans with developmental disabilities. Thus there is a very rich national basis for comparison of individual and group experiences of integrative activities.

The interrater reliability of this scale was reported to be very low when the two interviews were separated by 8 weeks, but very high when the time interval was corrected for (.97). The Integrative Activities Scale is shown on the following page.

## Integrative Activities Scale

Copyright © Center for Outcome Analysis, 2017

**ABOUT HOW MANY TIMES** did this person do each of the following in **ONE MONTH**? **ONLY COUNT ACTIVITIES WHEN THE PERSON WAS IN THE PRESENCE OF NON-DISABLED CITIZENS. (Rough estimates are fine.** If the past month was not typical, ask about the average month during the past year. Write DK if "Don't Know.")

**BEFORE** means in the previous living situation, or before the program being evaluated began. **I this is part of routine monitoring, use "A YEAR AGO" instead of "BEFORE."**

**NOW** means within the past 4 weeks.

<b>BEFORE (In previous situation – OR – A Year Ago</b>	<b>NOW (Past 4 Weeks)</b>	
1B	1N	Visit with close friends, relatives or neighbors
2B	2N	Visit a grocery store
3B	3N	Go to a restaurant
4B	4N	Go to a place of worship
5B	5N	Go to a shopping center, mall or other retail store to shop
6B	6N	Go to bars, taverns, etc.
7B	7N	Go to a bank
8B	8N	Go to a movie
9B	9N	Go to a park or playground
10B	10N	Go to a theater or cultural event (including local school & club
11B	11N	Go to a post office
12B	12N	Go to a library
13B	13N	Go to a sports event
14B	14N	Go to a health or exercise club, spa, or center
15B	15N	Use public transportation (May be marked "N/A")
16B	16N	Other kinds of "getting out" not listed above

17. **ACCESS TO TRANSPORTATION:** If this person wanted to go somewhere on the spur of the moment (beyond walking distance), how many times out of 10 would he/she be able to? If this person does not communicate such wants, phrase the question as "If someone unpaid wanted this person to be able to go somewhere on the spur of the moment BBBB" Count only trips that are within 1 hour of home.

18. \_\_\_\_\_ times out of 10 BEFORE

19. \_\_\_\_\_ times out of 10 in the past month, NOW



## **Choice Making and Autonomy in Daily Life**

The scale of choice making is called the Decision Control Inventory. It is composed of 35 ratings of the extent to which minor and major life decisions are made by paid staff versus the focus person and/or unpaid friends and relatives. Each rating is given on a 5 point scale, where 0 means the choice is made entirely by paid staff/professionals, 5 means the choice is made entirely by the focus person (and/or unpaid trusted others), and 3 means the choice is shared equally. This is the same scale used by the Robert Wood Johnson Foundation in its National Evaluation of Self-Determination in 29 states. The interrater reliability of the Inventory was reported as .86 (Conroy, 1995). The most current version of the Decision Control Inventory is shown on the following page.

## Decision Control Inventory, Before and Now

Copyright © J.W. Conroy 2017

Ask the person and/or the person's chosen ally to say who actually makes decisions in each area as shown, from 1 to 5. Use the "Two Either-Or Questions" approach. If decisions are made entirely by PAID PERSONNEL (program staff, Case Manager, agency officials, doctors, etc.), enter "1" for that area. If decisions are made entirely by the PERSON AND/OR TRUSTED FRIENDS, RELATIVES, ADVOCATES, etc., enter "5." If decisions are equally shared, enter "3."

### WHO MAKES DECISIONS?

1	2	3	4	5	99
<b>All or Nearly All Decisions Made by Paid Folks</b>	<b>Mostly Made by Paid Folks</b>	<b>Equally Shared Decisions</b>	<b>Mostly Made by Person and/or Freely Chosen Allies</b>	<b>All or Nearly All Made by Person and/or Freely Chosen Allies – relatives, friends, advocates</b>	<b>D/K, N/A</b>

BEFORE	NOW	FOOD
		What foods to buy for the home when shopping
		What to have for breakfast
		What to have for dinner
		Choosing restaurants when eating out
		<b>CLOTHES AND GROOMING</b>
		What clothes to buy in store
		What clothes to wear on weekdays
		What clothes to wear on weekends
		Time and frequency of bathing or showering
		<b>SLEEP AND WAKING</b>
		When to go to bed on weekdays
		When to go to bed on weekends
		When to get up on weekends
		Taking naps in evenings and on weekends
		<b>RECREATION</b>
		Choice of places to go
		What to do with relaxation time, such as choosing TV, music, hobbies, outings, etc.
		Visiting with friends outside the person's residence
		Choosing to <u>decline</u> to take part in group activities
		Who goes with you on outings?
		Who you hang out with in and out of the home?
		<b>SUPPORT AGENCIES AND STAFF</b>
		Choice of which service agency works with person
		Choice of Case Manager (or other term such as SSA, SC, etc.)
		Choice of agency's support persons/staff (N/A if family)
		Choice of support personnel: option to hire and fire support personnel
		<b>ECONOMIC RESOURCES</b>
		What to do with personal funds
		How to spend residential funds
		How to spend day activity funds
		<b>HOME</b>
		Choice of house or apartment
		Choice of people to live with
		Choice of furnishings and decorations in the home
		<b>WORK OR OTHER DAY ACTIVITIES</b>
		Type of work or day program
		Amount of time spent working or at day program
		Type of transportation to and from day program or job
		<b>OTHER</b>
		Express affection, including sexual
		"Minor vices" - use of tobacco, alcohol, caffeine, explicit magazines, etc.
		Whether to have pet(s) in the home
		When, where, and how to worship

## **Productivity and Meaningful Day Activities**

Productivity can be reflected by earnings, by the amount of time engaged in daytime activities that were designed to be productive (adult day activities, vocational training, workshops, supported and competitive employment). The scale captures hours in each kind of activity, and also how much money was earned, if any. In recent years, we have added the column at the right, which indicates whether the person was completely segregated from the general public, as in a sheltered workshop, or had some level of contact, no matter how small.

This simple form captures the core of the information needed to monitor, track, and evaluate changes in work and daily engagement in activities. As sheltered workshops and other segregated settings are considered for downsizing and conversion, we must find out “Whether the people are better off.” That is the fundamental question of all health and social program evaluations. In the realm of day activities and employment, the primary quality of daily life questions can be answered with “What are you doing every day, for how long, at what pay if any, and with what degree of contact with non-handicapped citizens?”

## Time, Money, & Integration – THEN & NOW

Copyright © James W. Conroy, 2017

- **THEN** means back then, before the person started with this program of support, no matter when that was. (When you fill out these columns, leave the NOW columns blank – they will be completed later.)
- **NOW** means NOW about 6 months to a year in the Initiative. Please answer the NOW questions about the past week – or a recent “typical week” if last week was unusual.

**HOURS:** Estimate how many hours per week are or were worked, on average, in each kind of work setting

**EARNINGS:** Estimate how much money per week the person earned or earns from each kind of activity on average

**INTEGRATION:** Write the number for HOW INTEGRATED the person was THEN and NOW:

Completely segregated	Never in the presence of people without disabilities	1
Mostly segregated	Some or a little of the time in the presence of people without disabilities	2
In between	Between 2 and 4	3
Mostly integrated	Often in situation where people without disabilities are, or might be, present	4
Completely integrated	Nearly always in a situation where people without disabilities might be present	5

Type of Day Activity	# Hours Work Per Week THEN	# Hours Work Per Week NOW	\$ Earned Per Week THEN	\$ Earned Per Week NOW	Integration THEN	Integration NOW
Self-Employed: Has His/Her Own Business						
Regular Job (Competitive Employment)						
Supported Employment						
Enclave or Work Crew Employment						
Sheltered Employment or Workshop Employment						
Pre-Vocational Program or Vocational Rehabilitation Program						
Day Habilitation Program (Adult Day Program, Non-Vocational Day Program)						
Senior Citizen Program						
Partial Hospitalization Program - Mental Health Oriented						
Volunteer Work						
Public School						
Private School						
Adult Education - GED, Adult Ed, Trade School, etc.						
Physical activity – exercise, gym, zumba, etc.						
Community Experience						
Other _____						
<b>TOTAL HOURS</b>			xxx	xxx	xxx	Xxx

## **Close Relationships**

Measurement of relationships and community connections is notoriously difficult in our field, and has received relatively little scientific attention. The world research literature on measurement of relationships and intimacy is completely dominated by scales that focus on verbal interaction. Yet we know that about half of the people in residential settings do not use verbal language at all, and many others have significant limitations.

The scale following aims to collect data about the person's five closest relationships. It captures their nature – including paid or unpaid – and the intensity of the connection. Human relationships can be measured via intensity, duration, and frequency (though no one would claim that measures everything about our immensely complex and important relationships).

With this kind of simple scale, we can find out whether a person can even name five close relationships – and many cannot. We can find out whether they are relatives or not – and whether they are paid or not. We also learn whether relationships include participation in individual planning, and how long they have lasted. (Many people experience important relationships with paid staff that turn out to be short term because of turnover.) We also learn how often people have contact of any kind with their closest allies.

# Closest Relationships Inventory

James W. Conroy, 2017

This table is intended to get at the types and characteristics of a few of your closest relationships.

- A “close relationship” or friend is anyone you define that way.
- If there are fewer than 5 close relationships, just describe however many there are.
- If there are close relationships with more than 5 people, please try to count only the closest 5.
- This scale may be left empty, if you has no close friends; please indicate this with a large “X.”
- “Contact” can include phone, letter, computer, or even just waving or saying hello.

## Closest Relationships: Now

ID or code	<u>Relationship</u> (Present or Former) 1. Relative 2. Paid person in this service program 3. Paid person not at this service program 4. Other paid (Case manager, nurse, etc.) 5. Roommate or Housemate 6. Co-worker or schoolmate 7. Friend “in the system” 8. Friend not “in the system” 9. Other	<u>Gender</u> of this friend  1. Male 2. Female	<u>Romance, Intimacy</u> Is this relationship romantic?  0. No 1. Maybe 2. Yes  99 = D/K	<u>Disability Status</u> Does this friend have a disability?  0. No 1. Yes, minor 2. Yes, moderate 3. Yes, major 99 = D/K	<u>Duration</u> About how long has the person known this friend?  (Years - use fractions and decimals if needed, as in 2.5 years, or 2 ½ years) 99 = D/K	<u>Frequency</u> About how many times has the person had ANY contact with this friend, in the past four weeks (28 days)? Maximum 28 for people seen every day. 99 = D/K
<b>1</b>	1 2 3 4 5 6 7 8 9	1 2	0 1 2 99	0 1 2 3 99	Years	Times in Past 28 Days
<b>2</b>	1 2 3 4 5 6 7 8 9	1 2	0 1 2 99	0 1 2 3 99	Years	Times in Past 28 Days
<b>3</b>	1 2 3 4 5 6 7 8 9	1 2	0 1 2 99	0 1 2 3 99	Years	Times in Past 28 Days
<b>4</b>	1 2 3 4 5 6 7 8 9	1 2	0 1 2 99	0 1 2 3 99	Years	Times in Past 28 Days
<b>5</b>	1 2 3 4 5 6 7 8 9	1 2	0 1 2 99	0 1 2 3 99	Years	Times in Past 28 Days

# **Psychometrics of Quality of Life Measurement**

## **General Background**

The quality of a test is judged by three major standards: (1) validity, (2) reliability, and (3) practicality.

Validity reflects how well a test measures what it is intended to measure. For example, a test of reading comprehension could lose validity if it allows too little time for taking the test. It might actually measure reading speed rather than comprehension.

Reliability refers to the consistency of results achieved by the test. To establish reliability, a test may be given to the same group several times. If very similar results are obtained each time, the test may be considered highly reliable.

Practicality involves the cost and convenience of the test. If a test requires too much expense or effort, it may be impractical. It also may be impractical if the results are too difficult to interpret.

(Excerpted from the World Book, author James Crouse, Ph.D., Professor of Education and Sociology, University of Delaware.)

## Psychometric Policy of the Center for Outcome Analysis

The objectivity of the evaluation depends upon what dimensions of quality are measured, and how they are measured. The principal scientific criteria of good measurement devices (referred to as protocols, instruments, or tests), are validity and reliability.

Validity in general concerns the degree to which a test measures what it is intended to measure. Validity is usually divided into several categories (AERA, APA, & NCME, 1985):

- Face validity - do 'experts' in the area of interest tend to agree that a test actually measures the underlying dimension of interest?
- Predictive validity - do test scores predict some future behavior or performance?
- Concurrent validity - do test scores correlate well with other accepted tests of the same performance?
- Content validity - does the test give a broad representation of the elements and subcategories of performance in the desired content area?
- Construct validity - the extent to which the test conforms to the requirements of the theoretical constructs the test is designed to measure.

A valid test measures the desired 'underlying dimension.' However, validity is often difficult to establish, because rigorous knowledge of the nature and properties of the 'underlying dimension' to be measured is required. With regard to the measures relevant to quality of life for people with developmental disabilities, the validity of many measures has been well established in the literature (Conroy, 1980; Conroy, Efthimiou, & Lemanowicz, 1981; Conroy et al, 1987; Flynn & Heal, 1981; Devlin, 1989; Harris, 1982; Isett & Spreat, 1979; Schalock, Keith, Hoffman, & Karan, 1989; Spreat, 1980; Spreat, 1982).

Reliability, on the other hand, concerns the degree to which a test is stable and consistent over time, across different users of the test, and across the items or subsections of the test. Stanley (1971) contended that reliability is no less important than validity, and perhaps is more important. A measure that has strong validity, but on which widely different results are obtained by different data collectors, or from one week to the next, is without value. "Unreliability places a question mark after the score and causes any judgement based on it to be tentative to some extent" (Stanley, 1971, p. 358). The more unreliable the instrument the more tentative the judgement.

Classically, three kinds of reliability are defined: test-retest, interrater, and internal consistency.

Test-retest reliability is the degree to which a measurement is stable over time. Test-retest reliability is assessed by administering the same test to the same individuals in exactly the same way twice, with some time interval between the two.

Interrater reliability is the degree to which measurements collected by different people agree with one another. A good measure should bring about consistent results regardless of the "rater" or test giver. Interrater reliability is assessed by having two different people collect the same measure at approximately the same time.



Internal consistency is the degree to which all of the items, parts, or subscales of a measurement tend to measure the same underlying dimension, and therefore to be correlated with one another. Internal consistency is assessed by several methods, including split half (the test is randomly divided into two sets of items which are tested for correlation), and item-total correlations (which examine the degree to which each item in a measure is correlated with the measure's total score).

In the Center for Outcome Analysis approach, we attempt to measure changes in many dimensions of quality of life. This information will be used to detect areas in which movement from institution to community appears to be helping, harming, or leaving unchanged the qualities of peoples' lives. This, in turn, may affect the policies and decisions of the government and service providers. Because the information can potentially influence how people are supported, it is extremely important that the data be accurate, objective, and scientific.

General background reference on Psychometrics:

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.

## References from the Original Work on Interviewing People with Intellectual Disabilities:

Sigelman, C.K., Budd, E.C., Spanhel, C.L., Schoenrock, C.J. (1981). When in doubt, say yes: Acquiescence in interviews with mentally retarded persons. *Mental Retardation*, 19, 53-58.

Sigelman, C.K., Budd, E.C., Winer, J.L., Schoenrock, C.J., & Martin, P.W. (1981). Evaluating alternative techniques of questioning mentally retarded persons. *American Journal of Mental Deficiency*, 86, 511-518.

Sigelman, C.K., Schoenrock, C.J., Spanhel, C.L., Hromas, S.G., Winer, J.L., Budd, E.C., & Martin, P.W. (1980). Surveying mentally retarded persons: responsiveness and response validity in three samples. *American Journal of Mental Deficiency*, 84, 479-486.

Sigelman, C.K., Schoenrock, C.J., Budd, E.C., Winer, J.L., Spanhel, C.L., Martin, P.W., Hromas, S.G., & Bensberg, G.J. (1983). *Communicating with mentally retarded persons: Asking questions and getting answers*. Lubbock: Research & Training Center in Mental Retardation, Texas Tech University.

Sigelman, C.K., Schoenrock, C.J., Winer, J.L., Spanhel, C.L., Hromas, S.G., Martin, P.W., Budd, E.C. and Bensberg, G.J. (1979). *Issues in interviewing mentally retarded persons: An empirical study*. Lubbock, Texas: Research and Training Center in Mental Retardation, Texas Tech University.

Sigelman, C.K., Winer, J.L., Schoenrock, C.J., Hromas, C.L. (1978). *Performance of the mentally retarded in interviews: Responsiveness to questions*. Lubbock, Texas: Research and Training Center in Mental Retardation, Texas Tech University.

Spanhel, C.L., Sigelman, C.K., Schoenrock, C.J., Winer, J.L., Hromas, S.G. (1978). *The feasibility of interviewing the mentally retarded: responsiveness, reliability, and response bias*. Lubbock, Texas: Research and Training Center in Mental Retardation, Texas Tech University.

Winer, J.L., Sigelman, C.K., Schoenrock, C.J., Spanhel, C.L., Hromas, S.G. (1978). *The performance of mentally retarded children on repeated and alternative format interview questions*. Lubbock, Texas: Research and Training Center in Mental Retardation, Texas Tech University.

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

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**

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