

Cross-Validated Factor Structures and Factor-Based Subscales for PASS and PASSING

Robert J. Flynn, Manal Guirguis, Wolf Wolfensberger, and Errol Cocks

Abstract: Program Analysis of Service Systems (PASS 3) and Program Analysis of Service Systems' Implementation of Normalization Goals (PASSING) are used internationally to assess service quality in mental retardation and other fields. In two studies, we derived and cross-validated the factor structures of these program evaluation instruments. In Study 1, based on 626 PASS 3 evaluations, four factors provided a good fit to the cross-validation data. In Study 2, based on 633 PASSING evaluations, three factors furnished a good fit to the cross-validation data. Similarities between the two factor structures, parallels with the structural and functional service dimensions emphasized in ecological approaches to evaluation, and practical uses in program evaluation of subscales based on the factors were discussed.

The principles of normalization and social role valorization have been very influential in shaping service policies and practices in mental retardation and related fields over the past quarter-century (Flynn & Lemay, 1999). Heller, Spooner, Enright, Haney, and Schilit (1991) found that Wolfensberger's (1972) book on normalization was rated by a panel of 178 experts as the most influential work of the previous 50 years in the field of mental retardation and that his article on social role valorization (Wolfensberger, 1983a) was rated as the 17th most influential work. Kozleski and Sands (1992) identified normalization and social role valorization (which is a reconceptualization of normalization) as the philosophical ground within which other important movements took root, namely, deinstitutionalization, supported employment, community residential options, and increased community participation. In the United Kingdom, Pilling (1995) observed that normalization and social role valorization have brought about an enormous change in services, particularly for people with developmental disabilities.

Given the importance of normalization and social role valorization for policy and practice, it is not surprising that the main program evaluation instruments that they have inspired have also been prominent. Program Analysis of Service Systems—PASS 3 (Wolfensberger & Glenn, 1975, 1989) and Program Analysis of Service Systems' Implementation of Normalization Goals—PASSING (Wolfensberger & Thomas, 1983, 1989) are used in the United States, Canada, the United Kingdom, France, Spain, Switzerland, Australia, and New Zealand, in fields such as mental retardation, mental health, and aging, to assess the quality of residential, vocational, recreational, and other types of community services in terms of their consistency with normalization and social role valorization principles, respectively. We note at the outset that in North America, Australia, and New Zealand, social role valorization has tended to supersede normalization as a philosophical approach to services, with PASSING also replacing PASS 3 (hereafter referred to simply as PASS) as a program evaluation tool. On the other hand, in the United Kingdom, France,

and Switzerland, normalization and social role valorization have tended to coexist, with both PASS and PASSING continuing to be used. Table 1 provides a brief overview and comparison of the two instruments.

Like normalization and social role valorization themselves, PASS and PASSING are based in large part on several ecologically ori-

ented assumptions about person–environment relationships (Levine & Perkins, 1987): human service environments exert significant effects on the behavior and competencies of the people served; a good fit between environmental opportunities and service recipients' specific needs is most likely to lead to positive service outcomes; and, therefore, it is useful to evaluate

Table 1
Brief Overview and Comparison of PASS and PASSING

Characteristic	PASS	PASSING
Main purposes	(a) To evaluate the quality of human service programs, agencies, service systems, or even proposals for services not yet implemented. (b) To teach the principle of normalization and its implications very specifically.	(a) To evaluate the quality of human service programs, agencies, service systems, or even proposals for services not yet implemented. (b) To teach social role valorization and its implications very specifically.
Criteria used to access service quality	The adoption and implementation of desirable service practices and implications derived from: (a) the normalization principle (34 ratings); (b) administrative considerations (16 ratings).	The adoption and implementation of desirable service practices and implications derived from social role valorization (all 42 ratings)
Intended range of application	A wide range of types of human service programs, formal and informal, in many fields.	A wide range of types of human service programs, formal and informal, in many fields.
Estimated level of complexity	High: intellectual content and documentary presentation deemed more complex than in PASSING.	Medium: intellectual content and documentary presentation deemed less complex than in PASS.
Main intended users	Many human service professionals, some ordinary citizens. Learning and using instrument deemed to require more conceptual ability, judgment, and service background than does PASSING.	Many human service professionals, many ordinary citizens, some service recipients. Learning and using instrument deemed to require less conceptual ability, judgment, and service background than does PASS.
Estimated feasibility of training enough local evaluators for routine use	Moderately low: greater complexity and more demanding training requirements of instrument (compared with PASSING) make it more likely that highly trained outside evaluators will be needed for local evaluations.	Moderately high: lesser complexity and less demanding training requirements of instrument (compared with PASS) make it more likely that adequate numbers of local evaluators can be trained to conduct local evaluations.
Structure of the instrument (major categories within which ratings are situated and method of construction)	Five major rating categories; integration, appropriate interpretations and structures, model coherency, developmental growth orientation, and quality of setting. Rational method of scale construction used to derive categories and ratings, without use of factor or item analysis.	Seven major rating categories, with each of two basic dimensions—image-enhancement and personal competency-enhancement—subdivided, respectively, into all four and the first three of the following subcategories: physical settings, service-structured groupings and relationships, activities and other uses of time, and miscellaneous/other. Rational method of scale construction used to derive categories and ratings, without use of factor or item analysis.
Ratings (items)	50, each with 3 to 6 levels. Degree of service quality denoted by a given level (e.g., Level 2) varies across ratings. Rating levels describe service practices but only rarely mention the likely impact of these practices.	42, each with 5 levels. Degree of service quality denoted by a given level (e.g., Level 2) is uniform across ratings. Rating levels describe service practices and are phrased in terms of the likely impact of these practices on client's image or competencies.
Possible range of total score	–947 (very poor service quality) to +1,000 (near-ideal service quality).	–1,000 (very poor service quality) to +1,000 (near-ideal service quality).
Official versions available	English and French	English and French

service environments in order to improve them. PASS and PASSING are, in fact, environmental instruments and belong to a larger class of environmental measures, some of which they relate to in expected ways. Perry and Felce (1995), for example, in investigating a sample of 14 community residences in Wales, found that various PASS rating clusters correlated meaningfully with the Characteristics of the Physical Environment Scale (Rotegard, Bruininks, & Hill, 1981), the Group Home Management Schedule (Pratt, Luszcz, & Brown, 1980), the Community Oriented Programs Environment Scale (Moos, 1974), the Index of Participation in Domestic Life (Raynes, Sumpton, & Pettifer, 1989), and simple counts of the frequency of social contacts and community activities. Pilling and Midgley (1995) found that PASS and PASSING produced portraits of individual residential services that were consistent with those yielded by another environmental measure, An Assessment of Care Environments (Wolfson, 1995).

In a recent review of PASS and PASSING research, Flynn (1999) found that only a few researchers had investigated the factor structure of PASS or PASSING. Such research is needed, however, because the original structures of the two instruments were determined through the use of rational procedures only, without the use of factor or item analysis. Therefore, we had two interrelated purposes for conducting the present research: to (a) derive stable (i.e., cross-validated) and reasonably definitive factor structures for PASS and PASSING and, thereby, (b) provide factor-based PASS and PASSING subscales that would be useful for program evaluation purposes. We used larger samples than those previously available and employed cross-validation procedures involving the combined use of exploratory and confirmatory factor analysis.

A few studies have been conducted to determine the factor structure of PASS. Demaine, Silverstein, and Mayeda (1980) factor-analyzed PASS scores gathered on 98 mainly small, residential facilities serving persons with developmental disabilities in California. They found six interpretable factors: compliance with normalization principles, administrative policies pertaining to normalization principles, issues of normalization with regard to programming and physical setting, issues related to the administration of services, physical location and avail-

ability of services, and comfort and functional nature of the physical setting. Eyman, Demaine, and Lei (1979) found evidence of the predictive validity of these PASS factors, which they discovered to be related to changes on three dimensions of adaptive behavior among 245 residents with developmental disabilities of the 98 homes. Overall, a moderately strong and statistically significant canonical correlation of .46 was found between the six PASS factors and changes on the three adaptive behavior domains: Personal Self-Sufficiency, Community Self-Sufficiency, and Personal-Social Responsibility. With age and IQ statistically controlled, Eyman et al. (1979) found that residents who gained the most on Personal Self-Sufficiency lived in facilities with higher scores on the PASS factors of Environmental Blending of Facility With the Neighborhood, Location and Proximity of Services, and Comfort and Appearance, but lower scores on Ideology-Related Administration. Those making the most progress on Community Self-Sufficiency lived in homes with higher PASS scores on Administrative Policies, Location and Proximity, and Comfort and Appearance. Those showing the greatest gains on Personal-Social Responsibility lived in residences with higher PASS scores on Location and Proximity of Service and on Comfort and Appearance but lower scores on Ideology-Related Administration.

Flynn (1980) analyzed a sample of 256 American and Canadian programs that had been evaluated with PASS. Most were community programs, with 58% in the field of mental retardation. Flynn compared the quality of different types of community and institutional programs on four PASS subscales that he derived through factor and item analysis: Normalization-Program (19 items), Normalization-Setting (12 items), Administration (8 items), and Proximity & Access (4 items). With all scores linearly transformed to a common metric (i.e., the percentage of the maximum possible score), multivariate profile analyses conducted on the four PASS subscales provided evidence of the discriminant and construct validity of the subscales. For example, child development services (the only type of program in which integration of participants with and those without impairments was found) outperformed other types of services on Normalization-Program, as expected. Similarly, community residential services outscored other programs on Normaliza-

tion-Setting, and community programs scored higher than institutional services on Proximity & Access.

The factorial structure of PASSING has been examined in only one previous study. Dansereau, Duteau, Ely, and Flynn (1990) used PASSING to evaluate the quality of 38 Quebec community residences for adults with mainly psychiatric disabilities (72% had a diagnosis of schizophrenia or major affective disorder, and 14%, a diagnosis of mental retardation). In the course of this study, Dansereau et al. also factor-analyzed a larger sample of 213 mainly American and Canadian PASSING program evaluations (described in Flynn, LaPointe, Wolfensberger, & Thomas, 1991), deriving four PASSING subscales: Social Role Valorization-Program (15 ratings), Social Role Valorization-Setting (14 ratings), Beauty & Comfort (3 ratings), and Accessibility (3 ratings). They found that in their sample of 38 residential services, quality was best on the subscales of Social Role Valorization-Setting ($M = 52\%$ of the maximum possible score) and Accessibility ($M = 51\%$), both reflecting clients' physical rather than social integration. Quality was noticeably weaker on the Social Role Valorization-Program subscale ($M = 34\%$), assessing the content of a service, and on the Beauty & Comfort subscale ($M = 42\%$), measuring the interior and exterior aesthetic quality of the physical setting.

Several studies in which these factor-based PASSING subscales were used have provided evidence of their predictive and construct validity. Ely (1991) collected data on quality of life, physical integration, and social integration among 70 persons with psychiatric disabilities living in the 38 residences assessed by Dansereau et al. (1990). Despite the attenuating affect on correlations of relating the scores of more than one resident per home to the PASSING subscale scores attained by the homes, Ely still found that three of the PASSING subscales significantly predicted residents' degree of physical integration. The latter, defined as the frequency of residents' involvement in various activities outside their respective homes, was correlated with the PASSING subscales as follows: Social Role Valorization-Program, $r(68) = .26, p < .05$; Social Role Valorization-Setting, $r(68) = .27, p < .05$; and Accessibility, $r(68) = .36, p < .01$. The Social Role Valorization-Setting score for the home also significantly predicted the resident's satisfaction with his or her living situation, $r(68) = .30, p < .01$.

Pelletier (1992) reported on a PASSING assessment of 39 residential, vocational, and community-support programs chosen to be approximately representative of a larger regional mental retardation service system in Quebec. Using the PASSING subscales derived by Dansereau et al. (1990), Pelletier again found that service quality was higher on Accessibility ($M = 63\%$ of the maximum possible score) and Social Role Valorization-Setting ($M = 57\%$) than on Social Role Valorization-Program ($M = 37\%$) and Beauty & Comfort ($M = 48\%$).

With the same factor-based subscales, Flynn (1993) studied service quality in an augmented sample of 406 PASSING evaluations, which included but went beyond those used in Dansereau et al. (1990), Flynn et al. (1991), and Pelletier (1992). Fifty-two percent of the programs were located in Canada, 46% in the United States, and 2% in the United Kingdom. Three quarters served persons with mental retardation, and 70% served adults. Service quality was again higher on Accessibility ($M = 57\%$ of the maximum possible score) and Social Role Valorization-Setting ($M = 43\%$) than on Social Role Valorization-Program ($M = 25\%$) and Beauty & Comfort ($M = 34\%$).

The foregoing research suggests that interpretable factors are, in principle, derivable from item-level PASS and PASSING data and that such factors are likely to be able to generate data supportive of their predictive, discriminant, and construct validity. As noted earlier, the two present studies were intended to derive stable PASS and PASSING factor structures and provide useful subscales for program evaluation. Both studies were based on large samples and cross-validation of findings.

STUDY 1

Method

Sample

The sample consisted of 626 PASS 3 evaluations conducted between 1975 and 1987, including the 256 previously analyzed by Flynn (1980). All were "team-conciliated" (i.e., consensus-based) assessments made by a team of evaluators working under the guidance of an experienced team leader. Fifty-seven percent of the programs were located in the United States, 32% in Canada, 10% in France, and 2% in Switzerland and Mauritius. Thirty-eight percent of the programs were community residences;

18%, vocational programs; 16%, institutional residences; 9%, child-development programs; 7%, educational services; 5%, counseling, therapy, or day-treatment programs; 3%, recreational services; and 4%, other types of programs. Fifty-four percent of the programs served persons with mental retardation; 10%, persons with mental health difficulties; 7%, elderly persons; 4%, persons with physical disabilities; 4%, youths in difficulty with the law; 2%, persons with difficulties related to alcoholism or drug abuse; 6%, persons with other conditions; and 12% included subgroups with "mixed" (different) conditions. Fifty-eight percent of the programs served adults; 20%, school-age children or adolescents; 8%, children age 5 or less; 7%, elderly persons; and 7%, mixed age groups. Seventy-nine percent of the evaluations were conducted during PASS training workshops; 14%, during officially mandated evaluations; 3%, at the invitation of the agencies involved; and 4%, for other reasons. Finally, 82% were conducted with the English version of PASS and 18% with the French version, although virtually all the leaders of French-speaking teams had had extensive training and experience with the original English version.

Instrument

PASS (Wolfensberger & Glenn, 1975, 1989) is used by a team of external raters to assess a human service program on 50 different items or ratings, each of which has from three to six response categories or "levels," ranging (in almost all instances) from *very poor* to *near-ideal* service quality. Seventy-three percent of the total weighted score is due to 34 normalization-related ratings, with the other 27% due to 16 ratings that assess administrative issues. In an often lengthy "conciliation" session, PASS team members pool their information and come to a consensus on the level to be assigned to each rating. Later, the team forwards a written report containing its findings and recommendations. Total weighted PASS scores, calculated by summing the weighted scores across the 50 items, can range from -947 (very poor service quality), through zero (minimally acceptable quality), to +1,000 (near-ideal quality).

In the present sample, the internal consistency reliability (α) of the 50-item PASS scale was .93. Concerning the interrater reliability of PASS, Flynn and Heal (1981) found, with teams of trained raters, that interrater re-

liability was high, with estimated intraclass correlations for the complete instrument varying between .70 for a single rater chosen at random and .94 for the mean of 7 raters. Flynn and Heal (1981) also found that the level of agreement was high between team-conciliated scores and the simple average (mean) of team members' individual preconciated scores, with all intraclass correlations above .70 and most exceeding .90.

Data Analysis

We conducted an exploratory factor analysis as a necessary prelude to a confirmatory factor analysis in order to establish the number of factors underlying the PASS items and to identify those items that would provide especially good indicators of the factors (Bentler & Wu, 1995). Both of the factor analyses constituted new analyses of the cumulative sample of 626 PASS evaluations and were independent of and analytically more powerful than Flynn's (1980) previous analysis of 41% (256 out of 626) of the same data base. Version 5.1 of EQS for Windows (Bentler, 1993; Bentler & Wu, 1995; Byrne, 1994) was used for the exploratory and confirmatory factor analyses and to split the total sample of 626 PASS evaluations into random halves. The first half ($n = 313$) was used for the exploratory factor analysis and the second half ($n = 313$), for the confirmatory factor analysis, with the latter serving to cross-validate the results of the exploratory factor analysis. Both factor analyses used maximum likelihood estimation and a direct oblimin rotation, allowing an oblique solution and correlated factors.

Results

Exploratory Factor Analysis

Four eigenvalues were above the scree line (11.44, 3.24, 2.85, and 2.28). Four factors were thus rotated to a direct oblimin solution. Because follow-up replication of our exploratory factor analysis results was a primary research goal, we followed Bentler and Wu's (1995, p. 210) suggestion to retain only those items with loadings greater than .50 in absolute value as especially good indicators of the factor on which they loaded. As shown in Table 2, using this criterion, we retained 25 of the 50 PASS items in the exploratory factor analysis. None of the 25 items had a substantial secondary loading on

any other factor; only 6 secondary loadings were larger than .20 in absolute value, and the largest was .26.

Factor 1 (Program) is composed of 14 ratings tapping the content of a human-service program. A program scoring highly on this factor would, as much as possible, integrate participants with valued citizens; promote participants' age-appropriate and culture-appropriate personal appearance and activities; address and interact with participants in warm and respectful ways; facilitate their acquisition of age-appropriate personal possessions; promote their exercise of age-appropriate autonomy, rights, and sex-related behavior; and avoid overprotecting them from normative challenges and risks. Overall, a high-scoring program would implement a coherent, relevant, and effective intervention model, employing personnel and

procedures that would meet participants' major life needs satisfactorily.

Factor 2 (Setting) consists of 3 ratings related to the physical setting of a service. A program scoring highly on this subscale would have a building with an external appearance that was very consistent with the function of the service (e.g., residential or vocational). The setting would also fit in well with the neighborhood in which it was located, and the physical features of the setting would challenge participants to improve their behavioral competencies.

Factor 3 (Administration) is composed of 5 items related to the administrative aspects of a service. A program scoring highly on this factor would provide ample opportunity for service recipients, their families, and members of the public to take part in decision-making; make vigorous efforts to educate the general public

Table 2
Factor Loadings of the 25 Retained Items on the Four PASS Factors by Type of Factor Analysis

PASS 3 factors & items	Exploratory factor analysis (<i>n</i> = 313)				Confirmatory factor analysis (<i>n</i> = 313)			
	Pgm	Set	Adm	Acc	Pgm	Set	Adm	Acc
Factor 1: Program (Pgm)—14 items								
14. Socially integrative social activities	.55	—	—	—	.56	—	—	—
16. AA personal appearance	.63	—	—	—	.50	—	—	—
17. AA activities, routines, & rhythms	.59	—	—	—	.61	—	—	—
18. AA labels & forms of address	.65	—	—	—	.63	—	—	—
19. AA autonomy & rights	.62	—	—	—	.80	—	—	—
20. AA possessions	.52	—	—	—	.56	—	—	—
21. AA sexual behavior	.51	—	—	—	.60	—	—	—
23. CA personal appearance	.68	—	—	—	.44	—	—	—
24. CA activities, routines, & rhythms	.58	—	—	—	.60	—	—	—
27. Model coherency	.63	—	—	—	.66	—	—	—
29. Social overprotection	.59	—	—	—	.68	—	—	—
30. Intensity of relevant programming	.67	—	—	—	.68	—	—	—
33. Individualization	.59	—	—	—	.76	—	—	—
34. Interactions	.51	—	—	—	.62	—	—	—
Factor 2: Setting (Set)—3 items								
8. Function congruity image	—	.63	—	—	—	.55	—	—
9. Building-neighborhood harmony	—	.52	—	—	—	.60	—	—
28. Physical overprotection	—	.52	—	—	—	.77	—	—
Factor 3: Administration (Adm)—5 items								
37. Consumer & public participation	—	—	.52	—	—	—	.65	—
38. Education of the public	—	—	.56	—	—	—	.67	—
40. Ties to academia	—	—	.55	—	—	—	.54	—
47. Planning process	—	—	.51	—	—	—	.49	—
48. Program evaluation & renewal mechanisms	—	—	.54	—	—	—	.48	—
Factor 4: Accessibility (Acc)—3 items								
1. Local proximity	—	—	—	.63	—	—	—	.66
3. Access	—	—	—	.50	—	—	—	.70
4. Physical resources	—	—	—	.50	—	—	—	.69

Note. AA = age-appropriate; CA = culture-appropriate. In the final confirmatory factor analysis, the factor (F) intercorrelations were as follows: F1, F2 (Program, Setting), $r = .58$; F1, F3 (Program, Administration), $r = .54$; F1, F4 (Program, Accessibility), $r = .27$; F2, F3 (Setting, Administration), $r = .34$; F2, F4 (Setting, Accessibility), $r = .24$; and F3, F4 (Administration, Accessibility), $r = .00$ (i.e., Factors 3 and 4 were uncorrelated).

about the program; have strong ties to academic institutions for training and research; have well-developed planning procedures; and have well-thought-out program evaluation procedures.

Factor 4 (Accessibility) consists of 3 access-related ratings. A program scoring highly on this factor would be physically close to the local population it served; convenient for service recipients, their families, and the public to reach; and located close to plentiful community resources relevant to its mission and the needs of its clients.

Confirmatory Factor Analysis.

Table 2 shows that the factor loadings in the final confirmatory factor analysis were of the same order of magnitude as those from the exploratory factor analysis, thus cross-validating the latter. The final confirmatory factor analysis loadings were all in the excellent to good range, between .80 and .44 in magnitude. Only 3 out of 25 were lower than .50, with all 3 in the mid- to high .40s. (According to Stevens, 1986, loadings of .40 or more in absolute value are large enough to be useful in interpreting factors. Thus, we considered confirmatory factor analysis factor loadings in the .40 to .49 range to be good indicators and those equal to or greater than .50 to be especially good indicators of their underlying factors.)

Table 3 shows the fit of the null, baseline (i.e., initially hypothesized), and re-specified (i.e., intermediate and final) confirmatory factor analysis models to the cross-validation data.

Because of substantial kurtosis in several of the PASS items, we calculated robust fit statistics (Bentler, 1993), including the Satorra-Bentler scaled chi-square and the robust Comparative Fit Index (CFI). Based on our knowledge of the various PASS items and on the results of the exploratory factor analysis, we hypothesized, in our baseline model, that an adequate model of the cross-validation PASS data would consist of four correlated factors plus two sets of items with correlated error terms (because of partial overlap in item content; see Model 1 in Table 3).

As expected, the fit of the baseline model, CFI = .89, to the cross-validation data was an enormous improvement over that of the null model (in which there would be no correlation at all among any of the variables). Given that a CFI of .90 or more indicates an acceptable fit (Byrne, 1994), however, further improvement in the model appeared possible. The final model was obtained by post hoc model fitting and provided a good fit, CFI = .93, accounting for 93% of the overall covariation among the 25 PASS items in the confirmatory factor analysis sample. Compared with the baseline model, the final model was composed of three additional parameters (i.e., three pairs of significantly correlated errors were added to the baseline model, as noted in Models 2, 3, and 4, Table 3), minus one deleted parameter (i.e., we found that the hypothesized correlation between Factors 3 and 4 was not significantly different from zero, and this parameter was removed, as noted in Model 5, Table 3).

Table 3
Goodness-of-Fit Statistics for the PASS Confirmatory Factor Analysis Models

Model	Model components ^a	Satorra-Bentler scaled χ^2	df	Robust comparative fit index
0 (Null)	All correlations among variables equal to zero	3045.56 ^b	300	.00
1 (Baseline)	Four intercorrelated factors (Program, Setting, Administration, & Accessibility), and two sets of correlated errors (E23/E16 & E24/E17)	522.63	267	.89
2	Model 1, plus an additional set of correlated errors (E30/E27)	496.67	266	.91
3	Model 2, plus an additional set of correlated errors (E9/E8)	473.14	265	.92
4	Model 3, plus an additional set of correlated errors (E28/E20)	453.39	264	.93
5 (Final)	Model 4, minus a set of factor correlations (F3/F4)	455.57	265	.93

^aE = Error, F = Factor. ^bIndependence model χ^2 .

Factor-Based PASS Subscales

Factor-based subscales were formed by summing the items within each factor. All analyses involving the subscales were based on the entire sample of 626 PASS evaluations, with alpha coefficients ranging between .91 (for the Program subscale) and .69 (for Administration) (see Table 4). In multiple regression analyses, the subscales were used to predict the total score on the original 50-item instrument, in two metrics: the original weighted-score PASS metric and a percentage-score metric in which the weighted scores were linearly transformed into a percentage of the maximum possible weighted score (see Table 4 for the conversion formulas used). The PASS subscales predicted the total PASS score very closely, $R = .97, p < .001$.

Comparisons among the PASS subscale means were also carried out, by means of dependent t tests, in the percentage-score metric (which rendered the subscales commensurate). To control the overall error rate, we divided the alpha level by the number of comparisons (.05/6), such that each comparison was tested at the .0083 level of significance. Service quality was best on Setting and Accessibility, the two dimensions most closely related to physical integration, with the Setting mean (see Table 4) larger than the means for Program, $t(625) = 36.3, p < .001$, or Administration, $t(625) = 32.7, p < .001$, and the Accessibility mean also larger than those for Program, $t(625) = 32.1, p < .001$,

or Administration, $t(625) = 30.0, p < .001$. The means for Setting and Accessibility, however, were not different from each other, $t(625) = 1.90, p = .06$, nor were those for Program and Administration, $t(625) = 0.75, p = .45$.

Discussion

The exploratory and confirmatory factor analyses showed that PASS can be adequately and parsimoniously represented by four factors, Program, Setting, Administration, and Accessibility, which accounted for 93% of the variance in the cross-validation data. Also, the factor-based subscales explained 94% of the total PASS score variance in the full sample. These factors and subscales are similar to the four originally found by Flynn (1980). However, because they were based exclusively on items that had loadings of at least .50 in the exploratory factor analysis and were successfully cross-validated in the confirmatory factor analysis, the present PASS factors and subscales are more factorially valid than Flynn's (1980) earlier ones and should, therefore, replace them in future research.

STUDY 2

Method

Sample

The sample was composed of a total of 633 PASSING evaluations conducted between 1983

Table 4
Internal Consistencies (Alphas), Means, SDs, and Conversion Formulas for the PASS Subscales and Total Scale

PASS subscale or total scale	No. of items	N	Alpha	Weighted score (WS) ^a		Percentage score (PS) ^b		Computational formulas for converting WS to PS
				Mean	SD	Mean	SD	
Program	14	626	.91	-100.2	148.0	36.1	20.5	PS = ((360 + WS) / 720) x 100
Setting	3	626	.72	12.0	13.0	72.6	24.5	PS = ((26 + WS) / 52) x 100
Administration	5	626	.69	-21.0	30.3	36.8	19.0	PS = ((79 + WS) / 158) x 100
Accessibility	3	626	.71	30.7	33.8	70.6	22.7	PS = ((74 + WS) / 148) x 100
Total PASS ^c	50	626	.93	-110.7	283.7	43.0	14.6	PS = ((947 + WS) / 1947) x 100

^aThe weighted score (WS) for each PASS factor-based subscale = the sum of the WSs on the items (listed in Table 2) making up the factor. The WS for each PASS item is given on the PASS scoresheet (Wolfensberger & Glenn, 1975). ^bPercentage score (PS) = the total WS on a particular subscale, expressed as a percentage of the total maximum WS on that subscale (see conversion formulas). ^cThe following intercepts and raw regression coefficients, taken from prediction equations calculated on the entire sample of 626 PASS evaluations, may be used to estimate the Total PASS WS or PS for a service program from its PASS subscale scores:

Predicted Total PASS WS = -18.08 + (1.43 x Program WS) + (4.08 x Setting WS) + (1.69 x Administration WS) + (1.22 x Accessibility WS).

$R = .97; R^2 = .94; F(4, 621) = 2400.4, p < .001$; Standard Error of Estimate (SEE) = 70.15.

Predicted Total PASS PS = 4.29 + (0.53 x Program PS) + (0.11 x Setting PS) + (0.14 x Administration PS) + (0.09 x Accessibility PS).

$R = .97; R^2 = .94; F(4, 621) = 2400.4, p < .001$; SEE = 3.60.

and 1995 and included the 406 evaluations previously analyzed by Flynn (1993). All were "team-conciliated" (consensus-based) assessments made by a team of evaluators under the guidance of an experienced team leader. Fifty-four percent of the programs evaluated were located in the United States; 37%, in Canada; 6%, in Australia; 2%, in the United Kingdom; and 1%, in New Zealand. Forty-two percent were group residences; 25%, vocational services; 13%, institutional residences; 10%, adult day programs; 5%, child development programs; 3%, school-based education programs; and 1%, other types of programs. Thirty-nine percent of the programs served persons with mental retardation; 10%, persons with mental health difficulties; 7%, elderly persons; 3%, persons with physical disabilities; 5%, individuals with a variety of other conditions; and 36%, subgroups with "mixed" (different) conditions. Eighty-six percent were conducted during PASSING training sessions versus 14% during official evaluations. Eighty-six percent were conducted with the English version of PASSING and 14% with the French translation, although the leaders of virtually all the French-speaking PASSING teams were highly experienced with the original English-language version (Wolfensberger & Thomas, 1983).

Instrument

PASSING assesses two major social role valorization dimensions—client social image-enhancement and client competence-enhancement—in four program areas: physical settings (i.e., the service facility and neighborhood in which the latter is located), service-structured groupings and interpersonal relationships, service-structured activities, and miscellaneous other issues. Because PASSING was intended to be simpler to learn and use than PASS (see Table 1), it is not used to assess the administrative aspects of a service. The instrument consists of 42 items or ratings, each composed of five response options or "levels," ranging from *very poor* to *near-ideal service quality*. On the basis of detailed criteria in the PASSING manual, an external team of trained raters arrives at a consensus concerning the level of performance attained by the service on each rating. Total weighted PASSING scores, calculated by summing the weighted scores on all 42 ratings, can range from a minimum of $-1,000$ (extremely poor service quality), through zero (denoting

minimally acceptable service quality), up to a maximum of $+1,000$ (near-ideal quality).

In the present sample, the internal consistency reliability of the 42-item PASSING instrument was .94. Regarding the interrater reliability of PASSING, Flynn et al. (1991) found that even with raters in training, interrater reliability was generally high, with intraclass correlations in the .54 to .70 range for a single randomly chosen rater and in the .89 to .95 range for the mean of teams of 5 to 9 raters. Flynn et al. also found good agreement for the teams of 5 to 9 raters between team-conciliated scores and the simple average (mean) of team members' scores, with intraclass correlations in the .66 to .94 range.

Data Analysis

An exploratory factor analysis was conducted to establish the number of factors present in the 42 PASSING items and to identify those items that would furnish particularly good indicators of the factors. A confirmatory factor analysis was then carried out to cross-validate the results of the exploratory factor analysis. Both analyses were new examinations of the cumulative sample of 633 PASSING evaluations, analytically independent of and more powerful than the previous analyses conducted by Dansereau et al. (1990) and Flynn (1993) of 34% (213/633) and 64% (406/633), respectively, of the same data base. *EQS for Windows* (Bentler & Wu, 1995) was used to split the total sample of 633 PASSING evaluations into random halves, with the first half ($n = 316$) serving for the exploratory factor analysis and the second ($n = 317$), for the confirmatory factor analysis.

Results

Exploratory Factor Analysis

Table 5 shows the results of the exploratory factor analysis. Three eigenvalues were above the scree line (12.87, 3.19, and 2.0); thus, three factors were rotated to a direct oblimin solution, permitting correlated factors. Again, because replication of our exploratory factor analysis results was a key research objective, only PASSING items with loadings greater than .50 in absolute value were retained as especially good indicators of their respective factors (Bentler & Wu, 1995). None of the 26 items retained by this criterion had a substantial sec-

secondary loading on any other factor; only 2 secondary loadings were equal to or greater than .20 in absolute magnitude, and the largest was .22.

Factor 1 (Program), composed of 15 items, reflects the content of a service. A service scoring high on this factor would exhibit the following image-related and competency-related characteristics: groupings of clients that enhance their image (e.g., by not grouping people with different types of disabilities) and their competencies (e.g., by including only a small number of less competent clients with a majority of more advanced ones); encouragement of frequent personal interactions and relationships between clients and more valued and compe-

tent persons; participation by clients in valued, age-appropriate, and individualized activities in integrated settings; matching of staff role identities and clients' primary needs (as shown, for example, in the operation of a business employing adults with impairments by managers trained in manufacturing or business techniques); support for clients' exercise of age-appropriate choices and rights (e.g., choice of voting for adults or of type of play activities for toddlers); support for the acquisition by clients of personal possessions that enhance their image and improve their skills and productivity (e.g., possession of a personal computer for use in one's schooling or work); respectful ways of speaking to and about clients; encouragement

Table 5
Factor Loadings of the 26 Retained Items on the Three PASSING Factors by Type of Factor Analysis

PASSING factors & items	Exploratory factor analysis (n = 316)			Confirmatory factor analysis (n = 317)		
	Pgm	Set	Acc	Pgm	Set	Acc
Factor 1: Program (Pgm)—15 items						
14. Image projection of intraservice client grouping—social value	.63	—	—	.65	—	—
16. Image-related other integrative client contacts & personal relationships	.65	—	—	.62	—	—
18. Service worker—client image match	.59	—	—	.56	—	—
20. Image projection of program activities & activity timing	.59	—	—	.72	—	—
21. Promotion of client autonomy & rights	.72	—	—	.68	—	—
23. Image-related personal possessions	.64	—	—	.56	—	—
24. Image projection of personal labeling practices	.61	—	—	.61	—	—
35. Competency-related intraservice client grouping—composition	.65	—	—	.70	—	—
36. Competency-related other integrative client contacts & personal relationships	.71	—	—	.71	—	—
37. Life-enriching interactions among clients, service personnel, & others	.68	—	—	.67	—	—
38. Program support for client individualization	.72	—	—	.76	—	—
39. Promotion of client sociosexual identity	.55	—	—	.50	—	—
40. Program address of clients' service needs	.80	—	—	.82	—	—
41. Intensity of activities & efficiency of time use	.77	—	—	.81	—	—
42. Competency-related personal possessions	.59	—	—	.53	—	—
Factor 2: Setting (Set)—8 items						
1. Setting—neighborhood harmony	—	.58	—	—	.78	—
2. Program—neighborhood harmony	—	.66	—	—	.70	—
5. External setting appearance congruity with culturally valued analogue	—	.70	—	—	.79	—
7. External setting age image	—	.55	—	—	.59	—
9. Image projection of setting—physical proximity	—	.54	—	—	.46	—
10. Image projection of setting—history	—	.59	—	—	.52	—
12. Image projection of program-to-program juxtaposition	—	.63	—	—	.57	—
13. Service—neighborhood assimilation potential	—	.54	—	—	.66	—
Factor 3: Accessibility (Acc)—3 items						
28. Setting accessibility—clients & families	—	—	.68	—	—	.77
29. Setting accessibility—public	—	—	.70	—	—	.85
30. Availability of relevant community resources	—	—	.66	—	—	.55

Note. In the final confirmatory factor analysis, the factor (F) intercorrelations were as follows: F1, F2 (Program, Setting), $r = .49$; F1, F3 (Program, Accessibility), $r = .21$; and F2, F3 (Setting, Accessibility), $r = .34$.

of positive interactions among clients, service personnel, and members of the public; support for each client's uniqueness, individuality, and sociosexual role identity (e.g., as spouse, parent, child, loved one); a focus on clients' most pressing life needs, within the purview of the service; and use of the most effective and efficient interventions that are available, given current knowledge.

Factor 2 (Setting), consisting of 8 items, assesses aspects of the physical setting in which a service is located. A service setting scoring highly on this factor would blend in well with the surrounding neighborhood, and the nature of the program itself would be congruent with the neighborhood (e.g., a residential program in a residential neighborhood); the exterior of the setting would be consistent in appearance, size, and style with the exteriors of settings providing similar services for valued persons (e.g., places of work); the setting would be located near other locations that are positively imaged, by virtue of their appearance, current ownership, or history, and would avoid locations near other service programs for devalued people; and the service would be located in a neighborhood in which the general population and community resources were numerous enough to be able to integrate clients.

On Factor 3 (Accessibility), made up of 3 items, a program scoring highly would have a service setting that is speedily and conveniently accessible to its actual or potential clients, their families, and the general public by virtue of its closeness to population centers and the availability of means of transportation. The setting would also be within easy reach of a wide variety of community resources relevant to the needs of its clients (e.g., eating places, shops, libraries, post offices).

Confirmatory Factor Analysis

The factor loadings from the confirmatory factor analysis, carried out in the second random subsample, are displayed in Table 5. Ranging between .85 and .46, they were all in the excellent to good range and of similar magnitude to those from the confirmatory factor analysis, which they thus cross-validated. Twenty-five out of the 26 confirmatory factor analysis loadings were equal to or greater than .50, with the other equal to .46. (As in Study 1, we considered confirmatory factor analysis loadings of .40 to .49 to be good, and those equal

to or greater than .50 to be especially good, indicators of their respective factors.)

In Table 6, data are provided on the fit of the null, baseline, and re-specified (intermediate and final) models. Because of substantial kurtosis in several of the PASSING items, we calculated robust fit statistics (Bentler, 1993), including the Satorra-Bentler scaled chi-square and the robust CFI. From our knowledge of the PASSING items and from the findings of the exploratory factor analysis, we hypothesized, in the baseline model that an adequate representation of the 26 PASSING items in the cross-validation sample would consist of three correlated factors, plus two sets of items with correlated error terms (because of partial overlap in content; see Model 1 in Table 6).

As expected, the fit of the hypothesized model, CFI = .90, was a dramatic improvement over that of the null (completely uncorrelated) model, although further improvement in the model appeared possible. The final model, CFI = .94, provided a good fit to the confirmatory factor analysis cross-validation sample data, accounting for 94% of the overall covariation among the 26 retained PASSING items. Obtained by post hoc model fitting, the final model included three additional parameters, all of which were pairs of significantly correlated errors among PASSING items (see Models 2, 3, and 4, Table 6).

Factor-Based PASSING Subscales

Subscales were formed from the cross-validated PASSING factors by summing the items within each factor. In analyses based on the entire sample of 633 PASSING evaluations, alpha coefficients for the subscales were first calculated (see Table 7). They ranged from .92 (Program) to .79 (Accessibility). The subscales were then used in multiple regression analyses to predict the total score on the original 42-item instrument, again in the weighted-score and percentage-score metrics (see Table 7 for the conversion formulas used). The scores on the PASSING subscales predicted the total PASSING score very closely, $R = .97, p < .001$.

Comparisons were conducted among the PASSING subscale means (in the percentage-score metric, which rendered the subscales commensurate), with dependent *t* tests. The overall alpha was divided by the number of comparisons made (.05/3), with each comparison tested at the .017 level of significance. Service qual-

Table 6
Goodness-of-Fit Statistics for the PASSING Confirmatory Factor Analysis Models

Model	Model components	Satorra-Bentler scaled χ^2	df	Robust comparative fit index
0 (null)	All correlations among variables equal to zero	4100.52 ^a	325	.00
1 (baseline)	Three intercorrelated factors (Program, Setting, & Accessibility), and two sets of correlated errors (E23/E42 & E16/E36)	580.66	294	.90
2	Model 1, plus an additional set of correlated errors (E14/E35)	527.90	293	.92
3	Model 2, plus an additional set of correlated errors (E12/E13)	487.92	292	.93
4 (final)	Model 3, plus an additional set of correlated errors (E21/E39)	472.17	291	.94

Note. E = Error.

^aIndependence model χ^2 .

Table 7
Internal Consistencies (Alphas), Means, SDs, and Conversion Formulas for the PASSING Subscales and the Total Scale

PASS subscale or total scale	No. of items	N	Alpha	Weighted score ^a (WS)		Percentage score ^b (PS)		Computational formulas for converting WS to PS
				Mean	SD	Mean	SD	
Program	15	633	.92	-262.2	159.7	21.1	17.3	PS = ((462 + WS) / 924) x 100
Setting	8	633	.85	-6.1	55.4	47.4	24.1	PS = ((115 + WS) / 230) x 100
Accessibility	3	633	.79	8.3	44.2	55.2	27.6	PS = ((80 + WS) / 160) x 100
Total PASSING ^c	42	633	.94	-360.6	310.1	32.0	15.5	PS = ((1000 + WS) / 2000) x 100

^aThe WS for each PASSING factor-based subscale = the sum of the WSs on all the items (listed in Table 5) making up the factor. The WS for each PASSING item is given on the PASSING scoresheet (Wolfensberger & Thomas, 1983). ^bThe PS = the total WS on a particular subscale, expressed as a percentage of the total maximum WS on that subscale (see conversion formulas). ^cThe following intercepts and raw regression coefficients, taken from prediction equations calculated on the entire sample of 633 PASSING evaluations, may be used to estimate the Total PASSING WS or PS for a service program from its PASSING subscale scores:

Predicted Total PASSING WS = 30.37 + (1.45 x Program WS) + (1.82 x Setting WS) + (1.01 x Accessibility WS).
R = .97; *R*² = .95; *F* (3, 629) = 3791.0 (*p* < .001); Standard Error of Estimate (SEE) = 71.16.

Predicted Total PASSING PS = 3.45 + (0.67 x Program PS) + (0.21 x Setting PS) + (0.08 x Accessibility PS).
R = .97; *R*² = .95; *F* (3, 629) = 3791.0 (*p* < .001); SEE = 3.56.

ity was best on Accessibility, the mean for which (see Table 7) was larger than those for Program, $t(632) = 28.5, p < .001$, or Setting, $t(632) = 6.1, p < .001$. Service quality was also better on Setting than on Program, $t(632) = 28.8, p < .001$.

Discussion

The results of the exploratory and confirmatory factor analyses showed that PASSING can be adequately and parsimoniously represented by three factors—Program, Setting, and Accessibility—that accounted for 94% of the variance in the cross-validation data. Also, subscales based on these factors explained 95% of the variance in the total PASSING score in the entire sample. The three cross-validated

factors identified here were similar to three of the four factors found in the only previous PASSING factor analysis (Dansereau et al., 1990). A fourth, relatively minor, 3-item factor from that study, Beauty and Comfort, did not survive the exploratory factor analysis because each of the factor loadings composing it was lower than .50, the criterion we used to retain an item as an especially good indicator of a factor. The three factors derived here are more factorially valid than, and thus are to be preferred to, the four identified by Dansereau et al. (1990).

Our three empirically derived PASSING subscales are considerably more homogeneous than Wolfensberger and Thomas' (1983) five

rationally constructed "programmatically subscore areas." Our subscales have excellent to good internal consistency (alpha) coefficients: Program, .92; Setting, .85; and Accessibility, .79 (Table 7). These alphas are notably higher than those found by Flynn et al. (1991) for Wolfensberger and Thomas' (1983) five subscore areas: Image Projection, .80; Integrativeness, .66; Intensity, .62; and Felicity, .60. (For their fifth subscore area, Relevance, which consists of a single item, it was not possible to calculate an internal consistency coefficient).

GENERAL DISCUSSION

In the present research we achieved our two-fold aim of deriving stable (cross-validated) and relatively definitive factor structures for both PASS and PASSING and producing factor-based subscales for use in program evaluation. Before discussing our findings, however, we note two limitations of the research. First, the majority of the evaluations in both samples were conducted during training sessions rather than during official assessments. This seems unlikely to have had a serious negative impact on our findings, however, because PASS and PASSING training sessions are conducted according to standardized guidelines (Wolfensberger, 1983b), under the supervision of experienced team leaders. As a result, good interrater reliability and good agreement between team-conciliated and averaged individual scores appear to be achieved even under training conditions (Flynn et al., 1991). Second, our final confirmatory factor analysis models involved a small amount of post hoc (i.e., exploratory) model fitting. Further cross-validation of our factor-analytic results would, therefore, be useful.

The PASS and PASSING factors of the same names are similar, though not identical, in content. Some differences between the same-named factors are to be expected, in light of the inter-instrument differences noted earlier (Table 1). These include differences in item content (due partly to the fact that PASS is based on normalization, whereas PASSING derives from social role valorization, a reconceptualization of normalization) and differences in item construction (PASS ratings have three to six levels, for example, whereas PASSING ratings all have five levels). Other differences in item content between the same-named factors stem from our demanding criterion for re-

taining items during the exploratory factor analysis: Both members of a pair of conceptually similar PASS and PASSING ratings were retained only if both had absolute factor loadings of at least .50.

Despite such differences, there is a striking overall similarity between the cross-validated PASS and PASSING factor structures. They cover similar program, setting, and accessibility-related dimensions of service quality (with PASS alone including an additional, administration-related, dimension) and are predictively very similar (both sets of subscales had multiple correlations of .97 with the total scores of their respective instruments). The fact that both sets of subscales capture almost all of the variance in their parent instruments, even though they are composed, respectively, of only 25 of the original 50 PASS items and 26 of the 42 PASSING items, also suggests that the factor analysis eliminated a considerable amount of redundancy among ratings. (Much of this redundancy originated in the intended use of PASS and PASSING as instruments for teaching the specific implications of normalization and social role valorization, respectively, in addition to their use as tools for evaluating service quality.)

The PASS and PASSING factors and subscales also resemble the structural and functional service dimensions that ecologically oriented evaluation approaches have defined as central (e.g., Felce, 1988; Landesman, 1988; Meador, Osborn, Owens, Smith, & Taylor, 1991). According to Meador et al. (1991), *structural* features include the physical characteristics of the service facility (e.g., size, site, convenience of location, adequacy of furnishings, and utilities), the socioeconomic status and population density of the neighborhood, and the experience and training of staff. *Functional* features, on the other hand, are those that are involved in the day-to-day operation of the program, such as the frequency and quality of interactions between staff and residents, the types of activities in which residents engage, the network of relationships with individuals and agencies in the community, the meeting of clients' needs, the opportunity for habilitation, the independence afforded clients, and the administrative organization and resource-allocation pattern in the program. Our Setting and Accessibility subscales can, therefore, be seen as measuring structural aspects of services, and our

Program and Administration subscales can be seen to be measuring functional aspects. This parallelism, which no doubt originates in a common concern with fundamental human service issues and in a shared environmental perspective, is supportive of the construct validity of our factors and subscales.

Meador et al. (1991) distinguished between structure and function because they found that group homes with very similar structural features differed greatly in terms of their functional features. Our findings similarly suggest that structural and functional aspects of services are relatively distinct, with good quality on the former being no guarantee of good quality on the latter. We observed only weak to moderate correlations between our structural (Setting, Accessibility) and functional (Program, Administration) factors, and service quality was considerably better for the former than for the latter. Other evaluations carried out with PASS and PASSING have yielded similar results. Picard (1988), Dansereau et al. (1990), and Pelletier (1992) all found wide program-related (i.e., functional) variations among residential services that had many setting- and accessibility-related (i.e., structural) similarities, with the structural aspects being of considerably higher quality.

Our research results have a number of practical implications for persons interested in program evaluation. First, our factor-based subscales provide standardized measures for assessing core dimensions of service quality that are directly derived from normalization and social role valorization, which are likely to remain important influences on the design and evaluation of community services. These subscales are also compatible, however, with a broader, ecological perspective on evaluation, in assessing structural and functional dimensions of services. On both counts, our subscales seem well-suited to helping service agencies meet the central challenge of the next decade that Felce and Perry (1997) recently identified, namely, the attainment of high-quality services. According to these authors the same priority must now be placed on improving the pertinence, sophistication, supportiveness, and effectiveness of community services (i.e., on function) that has previously been put on ensuring the appropriateness of their size, location, accessibility, and staffing (i.e., on structure). Concretely, this

means that service personnel must focus on understanding people's needs; on responding in relevant and intense ways to their needs; and on helping them achieve greater self-direction, personal development, and social integration. In terms of our subscales, success by an agency in meeting Felce and Perry's challenge might be seen, for example, in year-to-year gains in scores on the Program and Administration subscales and the maintenance of already satisfactory scores on the Setting and Accessibility subscales.

Second, our results should enhance the practical usefulness of both PASS and PASSING as evaluation instruments. Because of their high levels of factorial validity and homogeneity, our subscales are likely to prove more sensitive as measures than the rationally constructed rating clusters and domains used up until now. Also, the computational formulas provided in Tables 4 and 7 can be used to convert weighted subscale scores to a common metric (i.e., the percentage of the maximum possible score), permitting service quality to be compared easily across subscales. These features should facilitate the kinds of comparisons for which the instruments were originally created (e.g., among programs within the same agency, within a single program from one year to the next, or among different kinds of services within a given region). Finally, the subscales should help PASS and PASSING evaluators organize their field assessments more tightly around the program, setting, accessibility, and administration dimensions. This promises to afford greater unity and clarity throughout the evaluation process, from identifying key evaluation questions, gathering data, and identifying service strengths and limitations to formulating recommendations, delivering feedback, writing the report, and providing follow-up consultation. In a later companion paper, we hope to add to the utility of the present results by furnishing more detailed normative data derived from comparing the subscale scores of different types of services in several service fields and countries.

References

- Bentler, P. M. (1993). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Bentler, P. M., & Wu, E. J. C. (1995). *EQS for Windows user's guide*. Encino, CA: Multivariate Software.

- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.
- Dansereau, J., Duteau, C., Ely, P., & Flynn, R. J. (1990). *Évaluation des programmes résidentiels en santé mentale dans l'Outaouais*. Hull, Québec: Conseil régional de la santé et des services sociaux de l'Outaouais.
- Demaine, G. C., Silverstein, A. B., & Mayeda, T. (1980). Validation of PASS 3: A first step in service evaluation through environmental assessment. *Mental Retardation*, 18, 131-134.
- Ely, P. W. (1991). *Quality of life and social integration of psychiatrically disabled citizens in community residences*. Unpublished doctoral dissertation, University of Ottawa.
- Eyman, R. K., Demaine, G. C., & Lei, T. (1979). The relationship between foster care environments and resident changes in adaptive behavior: A path model. *American Journal of Mental Deficiency*, 83, 330-338.
- Felce, D. (1988). Behavioral and social climate in community group residences. In M. P. Janicki, M. W. Krauss, & M. M. Seltzer (Eds.), *Community residences for persons with developmental disabilities: Here to stay* (pp. 133-147). Baltimore: Brookes.
- Felce, D., & Perry, J. (1997). A PASS 3 evaluation of community residences in Wales. *Mental Retardation*, 35, 170-176.
- Flynn, R. J. (1980). Normalization, PASS, and service quality assessment: How normalizing are current human services? In R. J. Flynn & K. E. Nitsch (Eds.), *Normalization, social integration, and community services* (pp. 323-357). Baltimore: University Park Press.
- Flynn, R. J. (1993). Intégration et évaluation de programmes: Comparaisons internationales. In S. Ionescu, G. Magerotte, W. Pilon, & R. Salbreux (Eds.), *L'intégration des personnes présentant une déficience intellectuelle* (pp. 5-15). Actes du IIIème congrès de l'Association Internationale de Recherche scientifique en faveur des personnes Handicapées Mentales. Trois-Rivières: Université du Québec à Trois-Rivières et AIRHM.
- Flynn, R. J. (1999). A comprehensive review of research conducted with the program evaluation instruments PASS and PASSING. In R. J. Flynn & R. Lemay (Eds.), *A quarter-century of normalization and social role valorization: Evolution and impact* (pp. 309-341). Ottawa: University of Ottawa Press.
- Flynn, R. J., & Heal, L. W. (1981). A short form of PASS 3: A study of its structure, interrater reliability, and validity for assessing normalization. *Evaluation Review*, 5, 357-376.
- Flynn, R. J., LaPointe, N., Wolfensberger, W., & Thomas, S. (1991). Quality of institutional and community human service programs in Canada and the United States. *Journal of Psychiatry and Neuroscience*, 16, 146-153.
- Flynn, R. J., & Lemay, R. A. (Eds.). (1999). *A quarter-century of normalization and social role valorization: Evolution and impact*. Ottawa, ON: University of Ottawa Press.
- Heller, H. W., Spooner, F., Enright, B. E., Haney, K., & Schilit, J. (1991). Classic articles: A reflection into the field of mental retardation. *Education & Training in Mental Retardation*, 26, 202-206.
- Kozleski, E. B., & Sands, D. J. (1992). The yardstick of social validity: Evaluating quality of life as perceived by adults without disabilities. *Education and Training in Mental Retardation*, 27, 119-131.
- Landesman, S. (1988). Preventing "institutionalization" in the community. In M. P. Janicki, M. W. Krauss, & M. M. Seltzer (Eds.), *Community residences for persons with developmental disabilities: Here to stay* (pp. 105-116). Baltimore: Brookes.
- Levine, M., & Perkins, D. V. (1987). *Principles of community psychology: Perspectives and applications*. New York: Oxford University Press.
- Meador, D. M., Osborn, R. G., Owens, M. H., Smith, E. C., & Taylor, T. L. (1991). Evaluation of environmental support in group homes for persons with mental retardation. *Mental Retardation*, 29, 159-164.
- Moos, R. (1974). *The Social Climate Scales: An overview*. Palo Alto, CA: Consulting Psychologists Press.
- Pelletier, J. (1992). *Évaluation de la qualité des services du réseau de la déficience intellectuelle de la région 04 à l'aide de la méthode PASSING*. Gatineau, QC, & Gloucester, ON: Les Communications OPELL.
- Perry, J., & Felce, D. (1995). Objective assessments of quality of life: How much do they agree with each other? *Journal of Community and Applied Social Psychology*, 5, 1-19.
- Picard, D. (1988). *La réinsertion sociale des personnes handicapées mentales en familles d'accueil (5 tomes)*. Québec, QC: Centre de Services Sociaux de Québec.
- Pilling, D. (1995). Do PASS and PASSING pass? A critique of PASS/ING. In D. Pilling & G. Watson (Eds.), *Evaluating quality in services for disabled and older people* (pp. 50-60). London, UK, & Bristol, PA: Kingsley.
- Pilling, D., & Midgley, G. (1995). PASS/ING and ACE in action: Similarities and differences in evaluating services. In D. Pilling & G. Watson (Eds.), *Evaluating quality in services for disabled and older people* (pp. 84-105). London, UK, & Bristol, PA: Kingsley.
- Pratt, M. W., Luszcz, M. A., & Brown, M. E. (1980). Measuring dimensions of the quality of care in small community residences. *American Journal of Mental Deficiency*, 85, 188-194.
- Raynes, N. V., Sumpton, R. C., & Pettifer, C. (1989). *The Index of Participation in Domestic Life*. Manchester: University of Manchester, Department of Social Policy and Social Work.
- Rotegard, L. L., Bruininks, R. H., & Hill, B. K. (1981). *Environmental characteristics of residential facilities for mentally retarded people*. Minneapolis: University of Minnesota, College of Education, Department of Psychoeducational Studies.

-
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum.
- Wolfensberger, W. (1972). *The principle of normalization in human services*. Toronto: National Institute on Mental Retardation.
- Wolfensberger, W. (1983a). Social role valorization: A proposed new term for the principle of normalization. *Mental Retardation*, 21, 234-239.
- Wolfensberger, W. (1983b). *Guidelines for evaluators during a PASS, PASSING, or similar assessment of human service quality*. Toronto: National Institute on Mental Retardation.
- Wolfensberger, W., & Glenn, L. (1975). *Program analysis of service systems: A method for the quantitative evaluation of human services (3rd ed.)* Vol. I: Handbook. Vol. II: Field manual. Toronto: National Institute on Mental Retardation.
- Wolfensberger, W., & Glenn, L. (1989). *Analyse de programmes pour les systèmes de services (PASS): Méthode d'évaluation quantitative des services humains*. Manuel pratique (Version européenne de 1989 traduite de la 3e édition anglophone de 1975). Genève, Suisse: Comité Européen pour le Développement de l'Intégration Sociale (CEDIS).
- Wolfensberger, W., & Thomas, S. (1983). *PASSING (Program Analysis of Service Systems' Implementation of Normalization Goals): Normalization criteria and ratings manual (2nd ed.)*. Toronto: National Institute on Mental Retardation.
- Wolfensberger, W., & Thomas, S. (1989). *PASSING (Programme d'analyse des systèmes de services: Applications des buts de la valorisation des rôles sociaux): Manuel des critères et des mesures de la valorisation des rôles sociaux (2ème éd.)*. (M. Roberge, traduction, & J. Pelletier, adaptation). Toronto: Institut G. Allan Roeher & Les Communications OPELL.
- Wolfson, P. (1995). ACE: An assessment of care environments. In D. Pilling & G. Watson (Eds.), *Evaluating quality in services for disabled and older people* (pp. 78-83). London, UK, & Bristol, PA: Kingsley.
- Received 3/16/98, first decision 5/8/98, accepted 8/5/98.
Editor-in-Charge: Steven J. Taylor
-
- Authors:** ROBERT J. FLYNN, PhD, Professor, Community Services Research Unit, School of Psychology, 11 Marie Curie, University of Ottawa, Ontario, Canada K1N 6N5 (E-Mail: rflynn@uottawa.ca). MANAL GUIRGUIS, BSc, Doctoral Student, Clinical Psychology Program, School of Psychology, 145 Jean-Jacques Lussier, University of Ottawa, Ontario, Canada K1N 6N5. WOLF WOLFENSBERGER, PhD, Research Professor and Director, Training Institute for Human Service Planning, Leadership, & Change Agency, Syracuse University, 230 Euclid Ave., Syracuse, NY 13244-5130. ERROL COCKS, PhD, Professor of Learning Disabilities, Clinical Psychology Research Centre, University of Northumbria at Newcastle, Coach Lane Campus, Newcastle upon Tyne, England NE7 7XA.