The Catalogue of Goals for Residential Treatment: An Individualized Outcome Measure for Residential Children's Programs

S. Mark Pancer
Wilfrid Laurier University

RÉSUMÉ

Cet article esquisse le dévelopement d'un Catalogue d'objectifs de traitement en résidence (COTR) et son emploi dans l'évaluation du programme d'un camp d'enfants ayant des difficultés d'apprentissage et de comportement. Le COTR comprend 34 objectifs, qui se divisent en sept catégories majeurs. Chaque objectif est accompagné de quatre énoncés qui délimitent les niveaux de réussite s'échelonnant de pauvre à bon. Les équipes du camp ont sélectionné des objectifs du catalogue (COTR) pour chaque enfant. Au début et à la fin du camp, chaque enfant a été évalué par rapport aux objectifs choisis et son progrès a été déterminé. Les changements évalués à l'aide du COTR ont été comparés avec d'autres méthodes plus habituelles d'évaluation du comportement d'enfants. Le COTR semble être plus sensible que les autres méthodes d'évaluation. L'article examine aussi dans quelle mesure le COTR est adéquat à sa clientèle-cible. Enfin il prend en considération la validité psychométrique, l'efficacité administrative, et l'éfficacité du COTR à la gestion des programmes et des services.

ABSTRACT

This article outlines the development of the Catalogue of Goals for Residential Treatment (CGRT), and how it was used in the evaluation of a residential summer camp program for children with learning disabilities and behaviour problems. The CGRT consists of 34 goals, grouped into seven major categories. Each goal is accompanied by four statements that describe a range of functioning within that goal area, from poor to good. In using the CGRT, staff teams selected goals for each child from the Catalogue. Ratings on those goals were made at the beginning and end of the camp session, and a goal prograss score was computed for each child. Changes in the CGRT were significantly correlated with other, more standardized measures of children's emotional adjustment, and the CGRT appeared to be more sensitive to change than other measures. The CGRT was also evaluated according to its appropriateness for the client group. Its psychometric soundness, its administrative efficiency, and its usefulness in service delivery and program management.

* I would like to express my appreciation to Marc Wilchesky, Valerie Davis, Harold Minden, Carole MacLeod, Annette Hebb-Grier, and all of the staff at Camp Towhee and the Integra Foundation for their enthusiastic support of this research, and to Geoff Nelson, for his comments on an earlier version of the paper. I would also like to thank Joe Novak and Paul Dumouchel for their translation of the abstract into French.

* Requests for reprints and the complete CGRT manual should be sent to S. Mark Pancer, Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario Canada, N2L 3C5.

The Catalogue of Goals for Residential Treatment: An Individualized Outcome Measure for Residential Children's Programs

Individuals who attempt to evaluate children's mental health programs are confronted with a daunting array of problems. By no means the least of these is the selection and/or construction of appropriate outcome measures. Ciarlo et al. (1981), in their discussion of criteria for the development, selection and use of client outcome measures, list eleven major criteria (and several more sub-criteria) that the "ideal" outcome measure should meet. In my own discussion of outcome measures (Pancer, 1983), I suggested four kinds of standards that should be applied to such measures: they should be appropriate to the program and client group involved; they should be psychometrically sound; they should be administratively efficient; and they should provide information that is useful to the program's managers and service providers. These standards often conflict with one another. For example, the measure which is easiest and least costly to administer may not be the most psychometrically sound, or appropriate. In reality, there are few measures which come close to the ideal of meeting all of these standards, and one is continually faced with the prospect of making compromises with regard to one or more of them. The standard which is least desirable to compromise, I would suggest, is that of appropriateness. While it is possible to use measures which are a little less reliable, somewhat harder to administer, or less useful to service providers, it is essential that the measures selected be appropriate to the client population being served, and directly related to the program's goals and objectives.

The primary objective of many mental health programs is to reduce the psychological and behavioural problems of their clients. These problems may be of many different kinds, however. In children's treatment programs, for example, there may be children who are too aggressive and domineering in their interactions with others, and other children who are socially isolated and fearful of contacts with others. While the generic goal — to reduce problems — still applies to both kinds of children, very different goals must be pursued for individual children. This creates a tremendous problem in the selection of appropriate outcome measures. Standardized instruments of proven reliability and validity tend to apply the same yardstick to children who, at the individual level, may have many problems, and very different treatment goals. Yet if one uses different instruments to assess outcome for all the different kinds of problems, how does one aggregate these to provide a meaningful picture of the overall program effects?

A variety of solutions has been proposed for this conundrum. Initially, one of the most promising appeared to be Goal Attainment Scaling (GAS) (Kiresuk & Lund, 1979; Kiresuk & Sherman,(1968). In using this technique, goals are established for each service recipient individually, a range of five possible outcomes (from most unfavourable to the best anticipated) is specified, and weights are assigned to each goal. The indicator for each goal that best describes the client at the beginning and the end of treatment is then noted. These are then aggregated to produce a score which is indicative of client goal attainment, and is comparable to the scores of other clients.

Unfortunately, Goal Attainment Scaling has not lived up to its original promise (Calysyn & Davidson, 1978; Cytrynbaum, Ginath, Birdwell & Brandt, 1979). While it has demonstrated some usefulness as a therapeutic tool (Galano, 1977; Weinstein & Ricks, 1977), it has proven to be a technique which is costly to use and of questionable reliability and validity (Cytrynbaum et al., 1979). In what follows, I will be describing a goal-based outcome measure which attempts to redress some of the problems of Goal Attainment Scaling, while retaining some of its more useful features.

The Catalogue of Goals for Residential Treatment (CGRT) was designed originally for use at a camp for children with learning disabilities. Camp Towhee, operated by the Integra Foundation, is a co-educated, residential summer camp serving learning disabled children 8 to 12 years of age from across the province of Ontario. In addition to having been diagnosed as learning disabled, these children also manifest concommitent social and behavioural problems. The camp program lasts six weeks, and is designed to help the child in both the academic/learning and social/behavioural areas. In the mornings, campers work on specific learning problems in small groups with other children who have similar problems. The rest of the time, they participate in a group-centered camping program which includes a wide range of activities such as swimming, arts and crafts, canoeing and even ice-skating.

There is a strong emphasis in the program on teaching adaptive social skills in everything the child does. During the first week of camp, an assessment is made of the child's problem areas, and a set of treatment goals is devised. Throughout the summer, counsellors and other camp staff help the child work on these problem areas, and general social functioning, through the use of a variety of techniques, including modelling, reinforcement and instruction in interpersonal problem-solving.

Camp Towhee had used a goal-based treatment process for a number of years, but had never attempted to use these goals in evaluating treatment outcomes. When they conducted their first formal evaluation in the summer of 1983, it was decided that at least one of the measures adopted should make use of the individual goals set for the campers. Goal Attainment Scaling was ruled out, primarily because of the amount of time it would have taken for training in and actual use of the procedure. Instead, a catalogue of commonly used goals was constructed so that program staff could select appropriate goals for each camper from the catalogue.

Method

Research Participants and Evaluation Design

All 62 children attending camp in the summers of 1983 and 1984 were participants in this evaluation (a total of 124 children — 15 girls and 109 boys). They ranged from 8 years, 7 months to 12 years, 10 months in age, with a mean of 10.38 years. All measures were administered within one week of the beginning of camp, and one week before the end of camp. All of the measures were used in the 1983 evaluation, but only two of the measures — the CGRT and a child behaviour rating scale — were used in the 1984 evaluation.

Construction of the Goal Catalogue

The goals contained in the CGRT were derived primarily from the behavioural and social goals assigned to children the previous summer. A list was made of all the goals assigned that summer, and additional goals were added as a result of discussions with camp staff and management. Goals in the list were then examined for redundancy and clarity of expression, and redundant goals were deleted. In addition, the goal statements were amended to enhance the clarity and consistency of the language. The resulting 34 goals were grouped into seven major categories, according to the kind of behavioural or social skill they addressed: peer interaction, emotional expression, tasks and activities, behaviour/compliance, group interaction, personal care, and games and play skills.

A set of four statements that described a range of functioning was constructed for each goal in the catalogue. For example, one of the most frequently selected goals in the CGRT was "to increase compliance with reasonable adult requests". The statements accompanying this goal were: "almost always resists reasonable adult requests (1); frequently does not comply with requests (2); complies with the majority of reasonable requests (3); complies readily with reasonable adult requests (4)". In using the Catalogue, the staff team, consisting of the supervisor and three counsellors for each cabin group, decided on the goals for each of the six children in their group, and then selected goals in the catalogue that most closely matched the ones they had developed. The code numbers of these goals were recorded on a Goal Record Sheet. The staff team then selected the statement for each goal that best described the camper's behaviour at that point in time. This rating was made one week into the camp session, and again four weeks later, during the last week of the program. If a goal was devised which did not appear in the Catalogue, it was written in full on the record sheet. Since indicator statements were not available for written-in goals, the team was asked to assign a 1, 2, 3 or 4 to the goal, depending upon whether the child was performing very poorly, poorly, adequately, or well in that goal area at the time of rating. Post-test rating for all goals were made by the same staff team, without access to the pretest ratings made at the begining of the camp session. It was hoped that this would prevent the team from recalling their initial ratings, and reduce any bias on the part of the raters. A goal attainment score was calculated for each camper by determining the difference between the initial and final goal ratings, and dividing by the number of goals assigned.

Other Measures

Most of the other measures used in evaluating the program were standardized measures available from the published literature. With the exception of a self-esteem measure completed by the children, all measures were administered to the counsellors. Each counsellor completed the measures independently for the two children in his/her group for which he/she had case responsibility. These ratings were made within one week of the beginning of the program, and a week before the end of the program. Final rat-

ings were made without access to the initial ratings made approximately four weeks earlier.

The Child and Adolescent Adjustment Profile (CAAP) Scale. (Ellsworth, 1979).

The CAAP Scale is a standardized scale for measuring adjustment of children and adolescents aged 3 to 18 years. It consists of 20 items divided into 5 subscales of 4 items each. The subscales assess peer relations, dependency, hostility, productivity and withdrawal behaviour. The CAAP Scale is designed to be completed by parents, teachers, counsellors or other staff who work with children.

The short form of the CAAP was initially derived from a longer scale, the Personal Adjustment and Role Skills Scale (PARS II) (Ellsworth, 1977). Items were selected for use in the CAAP scale if they a) had high factor loadings following factor analysis of PARS II items, b) showed high reliability ratings, c) were judged to be clinically important for adjustment, d) were sensitive to pre-post treatment changes, and e) discriminated between groups known to differ in adjustment.

Reliability for the five subscales of the CAAP range from .80 to .90 (as indicated by coefficient Alpha) and from .78 to .89 (for test-retest reliability) (Ellsworth, 1979). Validation studies have shown that the CAAP discriminates amongst groups of normal, clinic and probationer youths.

The Behaviour Rating Form (BRF). (Coopersmith, 1967).

The BRF contains two sets of items, one set of 10 items (Self-Esteem Behaviours, SEB) providing an appraisal of behaviours that are associated with poise and assurance, and reactions to criticism and failure. The second part of the BRF provides an index of defensive behaviours such as bragging or bullying. Only the SEB was used in the present evaluation. Correlations between the BRF and other indices of performance include: .41 with total WISC IQ, .37 with standardized achievement tests, and .25 with sociometric ratings (Coopersmith, 1975).

Self-Esteem Inventory (SEI). (Coopersmith, 1967).

The SEI measures the evaluative attitudes toward the self in social, academic, family and personal domains. The inventory (short form) consists of 25 short statements (e.g., "I'm a lot of fun to be with") which are answered "like me" or "unlike me". The scale has been designed for use with individuals ranging form 8 years to adulthood, and is self-administered. The many studies using the SEI indicate that it has high reliability (Fullerton 1972; Taylor & Reitz, 1968) and validity (Coopersmith, 1967; Shaver & Robinson, 1973; Taylor & Reitz, 1968). The measure was administered to each child individually by the counsellor who had case responsibility for that child.

Social Participation Skills Checklist.

This measure was designed by Mills, Pancer & Favaro (1982) to assess a child's participation in group activity situations. It consists of fifteen descriptive phrases that counsellors rated according to how well they described the child's behaviour, from "not at all" (1) to "very well" (5). Included in the scale are items such as "can accept things not going his/her way" and "shares toys, equipment". Initial validation studies indicate that this is a reliable measure which is sensitive to change in children's group participation behaviour over short periods of time.

Inventory of Personal and Social Skills (IPASS).

The IPASS is a behaviour rating scale designed to assess all campers in the seven skill areas represented in the Catalogue of Goals. Each goal in the Catalogue has a corresponding item in the IPASS (e.g., "complies with reasonable adult requests"), which the counsellor rates the child as having never (1), rarely (2), sometimes (3), often (4), or consistently (5) performed during the previous week. While the CGRT provides an indication of how the child is functioning only in relation to the goals selected, the IPASS indicates the level of functioning in all areas represented in the Catalogue.

Results

In total, 273 goals were set for the children in 1983, and 248 in 1984, a mean of 4.4 and 4.0 goals per child, respectively. Of these, 15.3% (in 1983) and 11.3% (in 1984) were goals which did not appear in the Catalogue and had been written on the Goal Record Sheets. A wide range of goals were selected from all seven sections of the catalogue, with the exception of the "game and play skills" section. Table 1 contains the percentages of goals selected from each category of the CGRT, and Table 2 lists the most frequently selected goals.

Table 1

Percentage of Goals Selected from Different Categories

	Percentage of Goals Selected	
Category	**1983**	**1984**
Peer interaction	23.8	26.6
Emotional expression	12.1	10.9
Tasks and activities	19.4	21.4
Behaviour/Compliance	13.6	11.3
Group interaction	11.4	14.9
Personal care	4.0	3.2
Game and play skills	.4	.4

Table 2

Most Frequently Selected Goals

Goal	Percentage of Campers for Whom Goal was Chosen
To increase compliance with reasonable adult requests	21.8
To increase ability to initiate, maintain and complete tasks	21.0
To increase self-confidence and self-esteem	21.0
To increase frequency of peer interaction	20.2
To decrease teasing, name-calling	20.2
To increase tolerance for frustration	19.4
To develop positive leadership skills	18.5

Table 3

Comparison of Initial and Final Ratings

Measure	Initial Rating	Final Rating	t	df
Goals (1983)	1.86	3.02	14.26***	61
Goals (1984)	1.83	3.11	13.16***	61
Goals (total)	1.84	3.07	19.24***	123
CAAP (1983)	48.98	52.93	2.66**	61
SEB (1983)	31.42	32.16	.98	61
Social participation skills (1983)	46.43	49.33	2.17*	60
SEI (1983)	14.66	15.49	1.88	58
IPASS (1983)	146.83	152.59	2.30*	60
IPASS (1984)	142. 26	155.38	7.29***	56
IPASS (total	144.59	153.97	5.92***	115

* $<.05$

** $<.01$

*** $<.001$

Note. The CGRT and IPASS were the only measures used in the 1984 evaluation.

A comparison of initial and final goal ratings indicates significant improvements in almost all areas over the course of treatment. Table 3 contains the initial and final scores for all measures, along with related t-values. As the table indicates, changes in CGRT scores were of a much greater magnitude than were changes in other measures.

In order to examine the stability of the goals scores over time, Pearson correlations were computed between the initial and final goal ratings. These yielded highly significant positive correlations of .83 (in 1983) and .91 (in 1984).

Validity of the measure was assessed by correlating pre-post changes in the CGRT scores with changes in the other measures. These correlations appear in Table 4. As can be noted from the table, the CGRT demonstrates substantial correlations with related measures of children's emotional and behavioural adjustment, with the exception of the Self-Esteem Inventory.

Table 4

Correlations of CGRT Changes with Changes in Other Measures

Measure	Correlation with CGRT	N
CAAP (1983)	.48	61
SEB (1983)	.37	62
SEI (1983)	.14	59
Social Participation Skills (1983)	.42	61
IPASS (1983)	.47	59
IPASS (1984)	.49	57
IPASS (total)	.46	116

Note. All correlations are significant at the $p < .001$ level, with the exception of that with the Self-Esteem Inventory, which is not significant.

Discussion

When set against the standards of appropriateness, psychometric soundness, administrative efficiency, and clinical utility, the CGRT fared quite well. There is little question about its appropriateness for the group involved, because the goals in the Catalogue had been derived directly from goals that had been assigned to members of the same group one year earlier. The fact that fewer than 15 percent of the goals assigned to campers were non-Catalogue goals suggests that the goals in the Catalogue were sufficiently comprehensive to represent the great majority of behaviours targetted for improvement.

The evidence at hand suggests that the instrument also possesses reasonable psychometric qualities. Initial and final ratings on the selected goals were highly correlated. While this correlation is not as pure a measure

of temporal stability as one would want (because of the fact that there was a treatment program that occurred between the two administrations of the measure), it does indicate that, once specific goals are chosen, they can be rated in a stable fashion. Inter-rater reliability was not assessed, because it was decided that the entire treatment team should be involved in the establishment of goals and the rating of goal progress. The vadility of the goal ratings is substantiated by the high correlations manifested between changes in the CGRT and changes in related child behaviour rating scales such as the the Child and Adolescent Adjustment Profile. The measure's sensitivity to change appears to be even greater than that shown by related standardized measures.

Administratively, the CGRT appears to be relatively easy and inexpensive to use. It required only a small amount of time (less than two hours) to train staff to use the Catalogue in selecting goals and recording goal progress. There was evidence that the CGRT may have helped save time by reducing the amount of narrative description required by the counsellors to describe each child's level of adjustment. The descriptions provided by the goal indicators seemed to provide enough information without the need for additional material.

Clinically, the CGRT proved useful in a number of ways. It provided an individualized picture of each child's progress that helped the treatment team decide which behaviours were in greatest need of attention. More generally, by examining the most frequently chosen goals for all children, it was possible to assess the treatment needs of the entire group, and plan staff training and treatment strategies accordingly. This attention to the needs of clinical staff in selecting outcome measures can greatly enhance the extent to which evaluation results are utilized (Pancer, 1985a, b). Perhaps the greatest indication of the usefulness of the CGRT is the fact that camp management and staff decided to continue using it even after the formal camp evaluation had been completed.

There are, of course, limitations associated with any measure of treatment outcome. The reliability of the CGRT is difficult to assess in the same way as it is determined for other measures. While test-retest and inter-rater reliability can be determined in a straightforward manner once the goals are selected, it is by no means certain that different raters would select the same goals for a particular child. Any measure of inter-rater reliability, then, would have to incorporate a statistic reflecting the extent to which different raters were likely to select the same goals. A further problem with the measure, in its current use, is that the individual who administers the treatment also rates the child's goal progress. The bias that this may introduce could be reduced by having an independent judge rate the child's level of functioning, but the independent judge would have to be familiar enough with the child to make the ratings.

Perhaps the greatest concern expressed by those using the CGRT was the fact that it would make it too easy for them to select goals for the children in their care. Rather than selecting a goal that was exactly suited to a child's needs, they would choose one that was merely "close enough" from those listed in the catalogue. It was hoped that the provision of a "write-in" option would lessen the occurrence of such a possibility.

What does the CGRT offer relative to other, more standardized measures of child functioning? While it may not match the reliability or ease of use of some standardized measures, it provides the capacity to assess each child along dimensions that are maximally appropriate to his/her needs and course of treatment. Its greatest asset is its ability to provide a measure of goal attainment that is appropriate for individual children, but can be aggregated in a meaningful way with the scores of other children, so that the effectiveness of an entire program can be assessed. For programs that deal with children having similar needs and/or problems, a standardized child behaviour rating scale might be more appropriate as a measure of program outcome. The CGRT would be most useful for programs that deal with children having a wide varity of needs and treatment objectives.

References

Calysyn, R.J.& Davidson, W. (1978). Do we really want a program evaluation strategy based solely on individualized goals?: A critique of Goal Attainment Scaling. In T.D. Cook (ed.), *Evaluation Studies Review Annual. Vol. 3.* Beverly Hills, Calif.: Sage.

Ciarlo, J.A., Broskowski, A., Cox, G.B., Goldman, H.H., Hargreaves, W.A., Mintz, J., Waskow, I.E.& Zinober, J.W. (1981). "Ideal" criteria for development, selection, and use of client outcome measures. In J.A. Ciarlo, D.W. Edwards, T.J. Kiresuk, F.L. Newman & T.R. Brown, *The assessment of client/patient outcome techniques for use in mental health programs.* National Institute of Mental Health, Contract No 278-80-0005 (DB).

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco: Freeman.

Coopersmith, S. (Ed.) (1975). *Developing motivation in children.* San Francisco: Albion.

Cytrynbaum, S., Ginath, Y., Birdwell, J. & Brandt, L. (1979). Goal attainment scaling: A critical review. *Evaluation Quarterly, 3,* 5-40

Ellsworth, R.B. (1977). *PARS II Adjustment Scale.* Roanoke, VA.: Institute for Program Evaluation.

Ellsworth, R.B. (1979). *CAAP Scale: The measurement of child and adolescent adjustment.* Roanoke, VA.: Institute for Program Evaluation.

Fullerton, W.S. (1972). *Self-disclosure, self-esteem and risk-taking: A study if their convergent and discriminant validity in elementary school children.* Ph.D Dissertation: University of California, Berkeley.

Galano, J. (1977). Treatment effectivenesas as a function of client involvement in goal-setting and goal-planning. *Goal Attainment Review, 3,* 17-32.

Kiresuk, T.J. & Lund, S.H. (1975). Process and outcome measurement using goal attainment scaling. In J. Zusman & C.R. Wurster (Eds.), *Program Evaluation: Alcohol, Drug Abuse and Mental Health Services.* Lexington, Mas., D.C. Health, pp. 213-228.

Kiresuk, T.J. & Sherman, R. (1968). Goal Attainment Scaling: A general method for evaluating community mental health programs. *Community Mental Health Journal,4,* 443-453.

Mills, S., Pancer, S.M. & Favaro, P. (1982). *Development of a social skills scale for children aged 5-12 years.* Paper presented at the meeting of the Canadian Psychological Association, Montreal.

Pancer, S.M. (1983). *Outcome measures for children's mental health programs.* Paper presented at the meeting of the Canadian Psychological Association, Winnipeg.

Pancer, S.M. (1985a). Program evaluation: A participatory approach. *Canadian Journal of Community Mental Health, 4,* 5-7.

Pancer, S.M. (1985b). Program vs. evaluation: Reconciling the needs of service providers and program managers. *Canadian Journal of Community Mental Health, 4,* 83-92.

Shaver, P. R. & Robinson, J.P. (1973). *Measures of social psychological attitudes.* Ann Arbor. MI,: Survey Research Center, Institute for Social Research.

Taylor, J. & Reitz, W. (1968). *The three faces of self-esteem.* London, Ontario, Canada: Department of Psychology, The University of Western Ontario, Research Bulletin No. 80.

Weinstein, M. & Ricks, F. (1977). Goal Attainment Scaling: Planning and outcome. *Canadian Journal of Behavioural Science, 9,* 1-11.