
**Distinguishing Process and Outcome Indicators in Pay for Performance
Funding Models: Are People “better off?”**

Scott Spreat, Ed.D.

James W. Conroy, Ph.D.

¹Woods Research Institute, 30 Martin Gross Drive, Langhorne, PA, USA, 19047

²Center for Outcome Analysis, 426B Darby Road, Haverford, PA, USA, 19083

doi: 10.51505/IJEBMR.2023.7913

URL: <https://doi.org/10.51505/IJEBMR.2023.7913>

Received: Aug 29, 2023

Accepted: Sep 18, 2023

Online Published: Sep 27, 2023

Abstract

Pay for performance funding models are entering the intellectual disability field. Providers will be reimbursed for their performance on key metrics. Concern is raised that it is essential for client/patient welfare must be included and emphasized as the primary performance metric. Pay for performance models must be based on an assessment of whether the client/patient is "better off."

Keywords: Intellectual Disability; outcomes; Program Evaluation; Funding Models

1.0 Introduction

In the closing days of the 1980 presidential campaign, Ronald Reagan faced President Jimmy Carter in a second debate. During the debate, Reagan asked Carter and the American public whether things were better now than they had been at the start of Carter's term as President. This simple, common sense type of question, combined with a number of other factors, helped to assure a Reagan victory in November.

The expectation that endeavors, whether a presidency or a school or a social service program, should yield some sort of positive outcome is neither new nor unique. Hogan (2007) traced the history of social program outcome assessment, noting a beginning in the 18th century when William Farrish introduced quantitative grading to assess students' performance and the effectiveness of teaching strategies. At the height of President Lyndon Johnson's Great Society initiative, American Psychological Association President Donald Campbell (1969) argued that all social programs should be submitted to the same level of rigorous empirical validation as are scientific experiments. The essence of Campbell's argument was that social programs should be expected to collect and report evidence of the success of their efforts. At the core, Campbell sought for social programs to be evaluated by real outcomes rather than intent. The call was for an accountability for outcomes.

A confluence of factors was responsible for the introduction of outcome evaluation to the Intellectual Disability field. Innovators like the late Marc Gold (1978) demonstrated that people with intellectual disability could learn complex skills and perform them with accuracy. Richard Foxx and Nathan Azrin (1972) were among the behavior analysts who demonstrated the ability to substantially modify dangerous behaviors. Evidence grew that intellectual disability was not a static condition, and that new skills could be learned and inappropriate behaviors eliminated.

This sort of work essentially revealed possibilities and helped to elicit increased expectations for intellectual disability supports and services, and these increased expectations were submitted to objective, empirical investigation. Research efforts evolved from mere description of the condition of intellectual disability to assessment of therapeutic outcomes. Growth and change was now recognized as possible. Perhaps the greatest change agent within the field was litigation. Cases such as *Wyatt v. Stickney* and the *Haldemann v. Pennhurst* led to massive changes within the system, moving people from state institutions to smaller community based homes operated by private agencies. This massive social experiment drew the attention of both independent researchers and the National Institute of Mental Health.

The NIMH funded study of the closure of Pennhurst Center in Spring City, PA was designed to answer the simple question – “Are people better off?” The intent of the project was to collect objective data regarding the welfare of persons transferred from a large, overcrowded state institution to small group homes scattered throughout Pennsylvania. A broad based approach to the evaluation of the Judge Raymond Broderick's 1978 order was adopted. Key outcome indicators included: development of personal skills, changes in behavior, inclusion in the community, access to services, and satisfaction. These factors were designed to allow the investigators from Temple University Developmental Disabilities Center and the Human Services Research institute (Conroy & Bradley, 1985) to offer an objective opinion regarding the impact of the social experiment -- in a sense, to answer whether the individuals who were moved were "better off."

The Pennhurst Study, (Conroy & Bradley,1985) revealed that skills tended to increase modestly, participants spent greater amounts of time in community settings outside of their home, they received a higher amount of service hours, stakeholders report greater degrees of happiness, and employment tended to be more integrated than older sheltered workshop models. Conroy & Bradley's (1985) findings have been replicated across a variety of sites and settings, and a relatively common pattern of results has emerged. The only negative factors that emerged were some difficulties in obtaining certain forms of medical care (a problem that has declined somewhat) and a substantially reduced number of hours of vocational engagement (a problem that has remained) (Spreat & Conroy, 2015).

Perhaps the degree of success that was associated with most of the deinstitutionalization efforts in intellectual disability was linked to a shift in focus from providing proof that the model works to more of a management focus. While the Pennhurst study (Conroy & Bradley, 1985) did address goals, resources, services, and outcomes, this thoroughness has not been retained. At times, we seem mired in analysis of timeliness of reports and documentation of service hours, rather than the truly important questions regarding human welfare. The intellectual disability field is now giving consideration to funding programs based on performance on a variety of indicators. This approach derives largely from the healthcare industry. Healthcare providers are paid for their services based on their performance on a designed number of indicators. Typical indicators are process, quality, and efficiency (NEJM, 2018). Outcomes for patients seem to be less of an issue within healthcare. Pennsylvania has proposed the adoption of a selective contracting model that is largely based on a pay for performance model (Office of

Developmental Programs, 2023). The proposed model is designed to provide fiscal rewards to agencies that are recognized by their superior performance on key metrics; weaker performing agencies are not to be punished. A key concern with the Pennsylvania model is the selection of performance metrics. Eleven specific metrics are listed, but not one metric addresses the basic question regarding intellectual disability services/supports - "Are people better off?" There is nothing on attainment of treatment goals. There is nothing about self-determination. There is nothing about social integration, except slight mention with regard to employment (which oddly enough, is not regulated under the residential regulations in question). There is nothing regarding modification of socially devalued behaviors such as aggression and self-injury. There is nothing about quality of life. The net impact is that a better performing agency in terms of the proposed metrics might be supporting clients who are no "better off" than those clients supported by a weaker agency. Whether the general healthcare model or the Pennsylvania IDD model, outcomes seem to have been neglected in the interest of easier to measure process indicators. We will know whether services are provided, but not necessarily whether they have helped the service recipient. A similar pattern was evident in the small number of other states that have implemented this pay for performance approach.

The focus on program evaluation in the intellectual disability field has evolved over the past 40+ years. Early program evaluation efforts were focused on whether individuals with intellectual disability became more capable in response to therapeutic interventions. For example, Spreat & Isett (1980) studied adaptive behavior changes over time as an indicator of program outcome. Over time, the service model shifted from a "repair shop" model in which deficits or shortcomings were corrected to one of providing supports. Primary outcomes of concern became things like exercise of choice, societal inclusion, and general quality of life. Robert Schalock (2000) was a major proponent of this increased focus on quality of life, despite the measurement challenges. This change has been most clearly recognized by the National Core Indicator Project, which focuses heavily on life quality indices. The Council on Quality and Leadership has similarly recognized the shift in focus, however, their data collection strategy makes the process of aggregation more difficult than the National Core Indicator project.

We have a growing paradox in which some data collectors are interested in process measures, while others focus on human outcomes. Clearly the collection of data on process indicators is essential to understanding the factors that are associated with the identified outcomes. The importance of these process indicators, however, is substantially reduced when not considered within the context of outcomes. Without solid assessment of personal outcomes, none of the process indicators have any real meaning, or for that matter, any validity. For this reason, it seems essential that any program evaluation effort must attempt to directly answer the basic question, "Are people better off?"

2.0 Discussion

The argument that we put forth is largely supportive of the pay for performance approach, however, we are suggesting that the key metrics must address people-based outcomes, in addition to the process based metrics. Stufflebeam (1971) has written that program evaluation is a complex process, covering far more than outcome and process indicators. Stufflebeam (1971)

suggested that program evaluation consisted of four essential components. Under Stufflebeam's (1971) CIPP model, the evaluator must 1) identify the goals of the project [this is sometimes called context], 2. Assess the availability of resources to achieve the goals [called inputs], 3. Assess the implementation of the program [called processes], and 4. Assess the outcomes of the program [called products). A similar set of evaluation components was present in Carter and Newman's (1976) work on cost-outcome evaluation, and each of these key evaluative elements was evident in Conroy & Bradley's (1985) evaluation of the closure of Pennhurst. In simple terms, the evaluation of any social experiment must determine the adequacy of the project goals, the sufficiency of the resources (i.e., money, staffing, etc.) to accomplish the goals, the extent to which the intervention was actually implemented, and finally, whether the goals were achieved. All of this should be accomplished within the context of an experimental design that substantially reduces the possibility of alternative explanations for the observed outcomes. It must be recognized, however, that the validity of goals, resources, and inputs derives from the outcomes. If the outcomes are unacceptable, the values on goals, resources, and inputs have no meaning

Ultimately, any pay for performance scheme will be deficient if it fails to assess each of Stufflebeam's (1971) key measures, and any pay for performance scheme will be an absolute failure if it fails to assess people-based outcomes. Pennhurst study actually did address the four indicators. With proper guidance, Pennsylvania's performance contracting can re-establish the primacy of human outcomes, and the importance of learning whether people are truly "better off."

References

- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24(4), 409–429. <https://doi.org/10.1037/h0027982>
- Carter, D. & Newman, F. (1976). A client-oriented system of mental health service deliver and program management: A workbook and guide. DHEW Publication No. (ADM) 76-307, superintendent of Documents, U.S. Government Printing Office, Washington, D.C.
- Conroy, J. W., & Bradley, V. J. (1985). The Pennhurst longitudinal study. A report of five years of research and analysis. Philadelphia: Temple University Developmental Disabilities Center.
- Foxx, R. & Azrin, N. (1972). Restitution: A method of eliminating aggressive-disruptive behavior of retarded and brain damaged patients. *Behaviour Research and Therapy*, Volume 10, Issue 1, Pages 15-27, ISSN 0005-7967, [https://doi.org/10.1016/0005-7967\(72\)90003-4](https://doi.org/10.1016/0005-7967(72)90003-4).
- Gold, M. (1978). Try another way: A Manual. Austin, Texas: Marc Gold & Associates.
- Halderman v. Pennhurst State School & Hospital, 446 F. Supp. 1295 (E.D. Pa. 1978)U.S. District Court for the Eastern District of Pennsylvania - 446 F. Supp. 1295 (E.D. Pa. 1978), March 17, 1978

- Hogan, R. (2007). The historical development of program evaluation: exploring past and present. *Online Journal for Workforce Education and Development*, 2(4), article 5. Available at <https://opensiuc.lib.siu.edu/ojwed/vol2/iss4/5>.
- NJM Catalyst. (2018). What Is Pay for Performance in Healthcare? *NEJM Catalyst*, 3/1/2018. <https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0245>
- Office of Developmental Programs. (2023). Overview of selective contracting residential and support coordination services: concept paper. Harrisburg, PA: Office of Developmental Programs, Department of Human Services.
- Schalock, R. L. (2000). Three Decades of Quality of Life. Focus on Autism and Other Developmental Disabilities, 15(2), 116–127. <https://doi.org/10.1177/108835760001500207>
- Spreat, S. & Conroy, J. (2015). Longitudinal investigation of vocational engagement. *Journal of Policy and Practice in Intellectual Disabilities*, 12(4), 266-271.
- Spreat, S. and Isett, R. (1980). An examination of performance decrement in institutionalized mentally retarded persons. *Evaluation and the Health Professions*, 3, 385-397.
- Stufflebeam, D. (1971). The relevance of the CIPP evaluation model for educational accountability. Paper presented at annual meeting of the American Association of School Administrators. Atlantic City, NJ, 2/24/1091.
- Wyatt v. Stickney, 344 F. Supp. 373 and 387 (M.D. Ala. 1972), aff'd sub. nom., Wyatt v. Aderholt, 503 F. za. 1305 (5th Cir. 1974).