TUTORIAL

The Use of Evaluation in Treatment Programs for Children with Autism

Kristin L. Miller¹

Published online: 18 July 2016

© Association for Behavior Analysis International 2016

Abstract Program evaluation is the use of planned activities to monitor process, outcomes, and impact of a health program or intervention. The application of program evaluation to behavioral analytic treatment programs for children with autism is a useful and necessary activity to inform practitioners and other stakeholders of the efficacy of these programs and to promote adherence to best-practice treatments. A brief survey of behavioral providers in California and Texas and search of the behavioral literature suggest that the practice of program evaluation is underutilized among providers of behavioral services. Current organizational practices primarily involve reporting on individualized consumer goals. The purpose of this paper is to provide an introduction to evaluation processes and procedures to promote the implementation of some or all of these components. Areas discussed include defining the population served and program stakeholders, describing the program and intervention, selecting evaluation goals and objectives, ethical considerations, and reporting.

Keywords Program evaluation · Outcomes · Impact · Best-practices

Program evaluation has been utilized in public health and health care interventions for a number of years to assess the value and impact of specific work activities (NSF, 2002). Evaluation is used across direct service interventions, community collaboratives, research initiatives, advocacy work, and training programs. In all of these contexts, evaluation is a set of imbedded, systematic activities designed to collect

information about the characteristics, outcomes, and impacts of a program to make judgments about the program's effectiveness and to inform future program activities (CDC, 2011). In program evaluation, the reference to "program" refers to a set of planned activities that are designed to produce a specific result, rather than reference to an individualized learning program (as in an individualized education program) or goal. Thus, the evaluation is conducted at the systems level.

Program evaluation provides an empirical answer to the question, "How are we doing?" While a service provider may have a general feel about program processes and outcomes based on dimensions such as staff turnover, staff performance, and consumers' individual outcomes, a more systematic measure of these and other dimensions coupled with standardized measurement can reveal overall strengths and weaknesses of the program. This is important for the alignment of the service with best practices, the sustainability of the organization, and accountability to stakeholders (McMahon & Cullinan, 2014). Importantly, Durlak and DuPre (2008) found that programs with strong adherence to best practice demonstrated mean effect sizes two to three times higher than those with poor adherence.

As an example, a program evaluation was conducted to evaluate the effectiveness of trainings and intervention provided by the Contra Costa County Childcare Council to 17 Bay Area childcare facilities to increase the health outcomes of children birth to five (University of California Division of Agriculture and Natural Resources, 2015). Intervention strategies focused on health education, technical assistance, and the promotion of organizational policy to increase the nutritional value of meals served, to promote healthy meal practices and to increase physical activity. A number of dimensions were measured through parent and provider surveys and through direct observation. The evaluation found that many facilities were already implementing many of the best-practice

✉ Kristin L. Miller
Kristin.miller@edsupportservices.org

¹ Ed Support Services, 1942 Embarcadero, Oakland, CA 94606, USA

strategies, with notable improvements in the areas of reduced screen time, increases in fruit and vegetable consumption and nutritional education, increases in the availability of exercise equipment and adult-led physical activity, and decreases in processed meals. However, there were other areas in which improvements were not seen, such as children not being involved in food preparation and clean-up. Children often lacked free access to water, which may have reduced their water consumption, and adults did not typically sit and eat with children and instead stood over them prompting to eat, potentially leading to overeating. These findings helped to inform the Childcare Counsel about areas that require additional focus in future program efforts.

Evaluation can take many forms, and evaluators often propose methods under categorical titles, such as formative (pre-implementation/pilot phase), developmental (on-going and throughout intervention), and summative (post-intervention), which inform the scope, purpose, type, and schedule of data collection and the response to information collected (Patton, 2011). For the evaluation of intervention programs, specific questions (similar to research questions) are formulated to address the implementation of the intervention, program effectiveness, efficiency, cost-effectiveness, and attribution (CDC, 2011). These questions guide the evaluation process.

While evaluation is an institutionalized best practice across many publically funded health programs (e.g., Ballard & Grawitch, 2016; Bishop-Fitzpatrick, Jung, Ilsung, Trunzo, & Rauktis, 2015; Gaugler, 2015; Truiette-Theodorson, Tuck, Bowie, Summers, & Kelber-Kaye, 2015), the incorporation of evaluation at the program level into treatment programs for children with autism is still emerging. In 1987, Strain called for more stringent documentation and description of community-based intervention. In 1999, citing the lack of evaluation-based culture in treatment programs for children with autism, Jordan called for the incorporation of on-going evaluation into community-based intervention work. Recently, McMahon and Cullian (2014, p. 3689) stated that “although autism researchers have identified a common set of practices that form the basis of quality programming in ASD...little is known regarding the implementation of these practices in community settings” and suggest more rigorous evaluation practices and dissemination to provide information regarding the adoption and implementation of programs in community settings.

Indeed, there are few published program evaluations in the area of treatment for individuals with autism. An APA PsycInfo search using the terms “program evaluation” and “autism” yielded 114 records. References that included research designs, strictly narrative descriptions of programs, theoretical articles, and articles that included no actual evaluation or studied a population other than individuals with autism were eliminated. The remaining literature included nine dissertations, all of which were comprehensive evaluations that included quantitative and quantitative

measures, and nine articles, four of which focused on the validation of program intervention components and five of which were more extensive and layered evaluations. Most, but not all, evaluated behavior analytic treatment models. A review of these articles depicts the range of scope and focus of evaluation that is consistent with the program evaluation literature base across fields.

Two evaluations were formative, providing information regarding the effectiveness of pilot programs. McDougall, Servais, Meyer, Case, Dannenhold, Johnson, and Riggan (2009) conducted a preliminary evaluation of a program designed to build capacity to serve students with autism spectrum disorder in public schools using qualitative and quantitative data. Stakeholder satisfaction data were collected for participants and administrators. Educators who received support reported positive outcomes and stakeholder data were positively correlated with gains in school-wide outcomes. Heitzman-Powell, Buzhardt, Rusinko, and Miller (2014) used formative evaluation to measure knowledge of behavioral strategies and concepts taught in a web-based training program for parents of children with autism as well as intervention feasibility and costs vs. benefits. Findings were used to improve the program design.

Two evaluations were conducted to validate program models. McClannahan and Krantz (1997) described the incorporation of program evaluation by external reviewers as a method of validating best-practice educational services, staff training, administration, and program outcomes for children and adults with autism spectrum disorder served at Princeton Child Development Institute. This review serves as an on-going method of evaluation for the leadership team (e.g., developmental evaluation). Swartz, Sandall, McBride, and Boulware (2004) used program evaluation to validate the Project Developmentally Appropriate Treatment for Autism (DATA) school-based intervention model through the collection of stakeholder satisfaction surveys across district staff and parents and quantitative data measuring child gains.

Four evaluations focused on identifying programmatic elements that led to best outcomes for the participants and/or improved service delivery. Isaacs and Ling (2003) evaluated a Canadian Children and Youth Services Division intake and short-term intervention program for children with autism using efficacy and stakeholder satisfaction measures. Qualitative, quantitative (outcome), and demographic measures were collected, and results were used to improve the program delivery and shorten the waitlist for long-term services. Dyer, Martino, and Parvenski (2006) evaluated the River Street Autism Program, correlating outcome data with treatment hours (intensity), duration of time in program, and family participation. Measures of staff training and case supervision were also reported. Outcome data found positive correlations between greater intensity, time in program, and family participation. Perry and Condillac (2010) collected

qualitative and quantitative data to measure child outcomes, parent knowledge of procedures and principles, and parent satisfaction and stressors for families who attended a parent-training program for families of children with autism. The information collected highlighted program strengths and weaknesses. Vinton (2012) evaluated perceptions of a caregiver-directed pilot program for adults with developmental disabilities. Surveys were distributed to caregivers, project coordinators, and consumers. Focus groups were also conducted among a smaller subset of stakeholders. The results of the evaluation validated goodness of fit for the service, participant choice, and participant satisfaction.

Evaluation can also be used to conduct cost-benefit analyses. Trudgeon and Carr (2007) conducted a cost-benefit analysis of stressors vs. resources provided to participants in a home-based early intervention program for families with children with autism via the collection and analysis of qualitative data.

Given the limited number of published studies on program evaluation with behavioral providers, an email survey was sent to providers in California and Texas listed in the Texas Association for ABA Directory of Behavioral Consultants and Autism Speaks ABA California Provider Directory to attempt to gain more information from practitioners. Of the 239 organizations listed, valid emails were found for 209. The response rate was 10 %; 20 surveys were completed. Though the return rate did not allow for statistical analysis of data, trends did emerge that provide some preliminary data about current practices in the field.

Forty-five percent of surveys returned were from providers serving under 50 clients, 15 % from providers serving between 51 and 101 clients, 20 % from providers serving between 101 and 300 clients, and 20 % from providers serving over 300 clients.

Most providers (18/20) reported using internal staff to collect data for evaluation purposes, and all providers reported using individualized client goals as benchmarks for client progress. Six providers also reported collecting some data on process goals, four providers on outcome goals, and three providers on impact goals.

No providers reported presenting evaluation data in publications or presentations or making the results available on their websites. Four providers reported using the information to make internal improvements, and four providers reported sharing data with stakeholders.

When selecting contingencies that competed with conducting more thorough and complete evaluation, 80 % of providers reported that financial limitations and the variability of client assessment/measurement tools were barriers to compiling outcomes and collecting impact data. Sixty percent of providers stated that limited knowledge of program evaluation was a barrier, and 55 % noted that program evaluation is not a current organizational priority. Twenty percent of providers responded that the lack of personnel and other factors are additional barriers.

The information gleaned from this survey suggests that behavioral providers could benefit from additional information regarding program evaluation and strategies to promote its utilization and practice without imposing undue strains on an organization's finances and personnel. The purpose of this paper is to provide an introduction to evaluation processes and procedures to promote the implementation of some or all of these components. A full tutorial is outside of the scope of a journal article; however, there are many print and online resources that provide comprehensive instruction on the topic.

Some Design Elements

The following sections provide an overview of elements generally included in a program evaluation with specific attention to areas that may be unfamiliar to behavior analysts.

Description of the Population, Provider, and Capacity

In programs designed to promote the health of a population, the first step is conducting a needs assessment, which is the process of identifying, analyzing, and prioritizing the needs of a priority population (Grunbaum, Gingiss, Orpinas, Batey, & Parcel, 1995). The needs assessment answers the questions, “*Who* are we serving (clear description of the population), *why* are we serving them (justification for intervention based on the literature), and *how* are we serving them (description of the intervention components)?” Although behavior analysts typically work within sectors in which categorical funds are pre-designated to specific populations and intervention components, it is useful to conduct consumer surveys for a variety of reasons, including learning more about the variations in the population to be served such as cultural and language differences and knowledge of behavioral concepts before services begin. This critical step also ensures that appropriate and adequate resources are allocated to match organizational capacity with the need of the population. A complete description of best-practice needs assessment is outside of the scope of this article; however, the CDC provides a variety of resources on the topic (e.g., <http://www.cdc.gov/policy/ohsc/chna>).

Defining stakeholders is another step that may be missed by behavioral providers who mistakenly view their only stakeholders as the consumers of the service. In fact, the list of stakeholders is much wider, encompassing organizational staff and the Board of Directors, funding agencies, partner agencies, legislators, affiliate university programs (as in when agency staff are receiving their practical training in the field), members of the general public who are interested in the efficient and effective use of public resources, and the field of Applied Behavior Analysis. Considering the full range of stakeholders involves reflecting on how outcome and impact measures may be expanded to capture the wider impact of

services and prompts the engagement of individuals representing these sectors in the evaluation. Including a variety of stakeholders in the evaluation process will provide a different orientation and give valuable insights and opinions as to what is important. Ultimately, stakeholders will be the program's advocates, lend credibility to the program, help to institutionalize the practice, and provide funding.

After defining the population of interest, stakeholders, and justifying the need for service, a program evaluation begins with a clear description of the organization and its service. Behavioral service providers generally operate with a current business or program plan that provides this information including the following.

- A program rationale based on epidemiological data, the economic impact through return on investment (ROI) calculations, and scientific evidence for the effectiveness of the intervention
- Mission and vision statements
- Agency goals and objectives
- Theory and/or logic models that specify the contingent relations for behavior change and provide a basis for measuring outcomes and impacts (Nutbeam & Harris, 1999; Julian, 1997);
- A comprehensive description of the intervention and ethical considerations
- Identification and allocation of resources, including program personnel and budget
- Clearly defined protocols for baseline and intervention assessments
- Other activities including marketing, research, community outreach, and involvement in public policy (McKenzie, Neiger, & Thackeray, 2013)

While many of these elements are familiar to agency staff, some may lack familiarity with theory and logic models to visually describe intervention processes. A logic model is a pictorial representation linking program activities, processes, theoretical assumptions, and principles with the outcomes and impacts of the program. Developing a logic model provides a clear description of assumptions and key events (inputs, activities, and outputs) that can be monitored and provides a common understanding between all program staff and the staff assigned to evaluate the program's effectiveness (W.K. Kellogg Foundation, 2004). Additionally, a well-conceived logic model is useful for identifying indicators for measurement. An example of a logic model is depicted in Fig. 1 (Shakman & Rodriguez, 2015).

Evaluation Focus

As stated earlier, program evaluation answers questions regarding organizational health and intervention effectiveness

and impact, and thus these questions must be defined to focus the evaluation. Jordan (1999) noted some general questions that may be relevant to providers of behavioral services:

1. Is this intervention successful, for whom and under what circumstances?
2. Is intervention X superior to intervention Y?
3. Is the intervention being implemented with fidelity?
4. What are the barriers to program implementation?
5. What is the cost-benefit analysis of receiving the intervention?
6. What is the overall benefit to the community?

These questions are not an exhaustive list nor are these areas mutually exclusive. It is likely that the service model will guide the evaluation team in selecting the most relevant questions, but it is important to incorporate other measures that are of interest to the wider stakeholder community, which is why it is important to involve stakeholders in the evaluation process.

Program Goals and Objectives

After defining the evaluation questions of interest, the heart of program evaluation begins with a careful outline of program goals and objectives. Goals define what the program intends to accomplish. Objectives are short-term measures that fall into three categories: process, outcome, and impact.

Process objectives describe the activities related to program implementation such as creating a board or advisory committee, hiring and training staff, developing the intervention and program materials, and purchasing equipment. Some of these activities are one-time events, and others may be on-going or recur at particular intervals. Data collected are generally descriptive in nature, such as reporting on dates and the nature of activity.

The process level of analysis can provide important information about the fidelity of the intervention such as turnover rates of direct service staff, amount of on-going training delivered to direct service staff, data describing clinical case-loads, and lapses in supervision by clinical staff due to variables such as resignations. Reporting on deviations from the intended program plan helps providers understand trends and challenges that can inform systems-level change to provide a higher quality service and improve client outcomes.

Outcome objectives are measures related to the effects of intervention at the end of intervention (or pre-designated mid-points). Outcome measures are sometimes confused with impact measures. Consider, for example, the evaluation of the language abilities through standardized assessment of preschool-aged children with autism who were enrolled in 2 years of early intensive behavioral intervention programming at the rate of 25 h per week. This is an empirical measure of the

Context: Implement four pillars of effective teaching: teacher evaluation, teacher preparation, teacher induction, and professional development.

Long-term goals: To support teachers' continuous improvement and to improve student learning and other outcomes.

Evaluation goal: To evaluate the implementation of the state teacher evaluation framework in a subset of schools and districts in 2012/13.

Strategies	Participants		Implementation outcomes	Intermediate outcomes	Impact and long-term outcomes
Design and implement new teacher evaluation systems that employ multiple measures including:	Teachers		District teacher evaluation plans correspond with state guidelines.	Based on information gained from evaluation of implementation:	Alignment of evaluation and professional support systems
• Teacher observations.	Classroom and school specialists		District teacher evaluation systems have been implemented as designed with regard to multiple measures of student learning and teacher performance including:	• Program changes and modifications to framework for teacher evaluation.	Improvement in teacher practice.
• Evaluator and teacher conferences.	School and district administrators		• Teacher observations.	• Department of Education provides guidance, training, and support materials for districts.	Improvement in student-learning outcomes, including:
• Student-learning objectives.	Department of Education		• Evaluator and teacher conferences.	• Individual district modifications to teacher evaluation systems.	• Higher achievement on standardized assessments.
• Professional growth plans that place teachers in "tracks" based on experience.	Students		• Student-learning objectives.	• Systems differentiate among teachers based on level of experience and performance.	• Decreased dropout rate.
• Multiple rating categories.	Parents		• Professional growth plans.		
• Teaching standards based on the Danielson framework. ^a	Other stakeholders (such as school committees, school board)		District teacher evaluation systems have been implemented with fidelity with regard to use of multiple rating categories.		

a. Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Assumptions: The Department of Education assumes that educator evaluation is one of four essential pillars of effective teaching. To maximize long-term outcomes, all pillars need to be in place and integrated. The department also assumes that if districts implement reformed teacher evaluation systems that map onto the Blueprint for Effective Teaching in New Hampshire (<http://www.education.nh.gov/teaching/documents/phase1report.pdf>), then district-level systems will lead to long-term outcomes indicated above. It hypothesizes that achievement of outcomes will be influenced by professional climate: Schools with a favorable climate will be more likely to have successful implementation.

Fig. 1 Example of a logic model

immediate effectiveness of the program components at exit rather than longer-term measures. For outcome objectives, data collected is often quantitative but may also include qualitative measures, such as surveys and interviews of program participants to measure client satisfaction and social validity.

Meaningful evaluation does not necessitate the use of experimental design. However, a level of description and evaluation of the intervention that exceeds the current practice of reporting on individual consumers' program goals independent of standardized or normative measures is critical. Individuals' progress on goals does not allow for the systematic analysis of overall program outcomes and impacts. If providers have the opportunity to incorporate comparison

groups into an evaluation, for example, subpopulations of individuals who received a different level of service or individuals yet to be served in the program because they are wait-listed, this may be useful but it is not necessary. While a comparison-type design is outside of the capacity of most providers, conducting simple quasi-experimental pre-post design without a control group is a first step toward reporting outcomes and impacts. The use of this type of design is appropriate when it is not ethical to assign participants to a control group and when program staff have no control over variables such as the timing of participants entering the program and/or cannot control factors such as participants' ages and developmental levels upon entry.

Service providers who are currently relying on the exclusive use of individualized goals for outcome reporting may struggle with conceptualizing how to incorporate more standardized measures that allow for group comparisons. Populations served need to be clearly delineated; even if similar intervention procedures are used across groups, measurement tools are likely to vary. Many providers begin services with a behavioral assessment that generally includes a norm-referenced assessment and/or curriculum assessment as well as functional assessment of problem behavior. Larger providers have often developed their own assessments and curriculum, and providers of all sizes use tools such as the Assessment of Basic Language and Learning Skills (Partington, 2005) or assessments such as the Battelle Developmental Inventory and the Vineland Adaptive Behavior Scales-II. Periodic reassessment using these tools are opportunities to measure participants' gains over time, and when these gains are compiled and quantified by some dimension such as percent gain, some basic program outcomes can be demonstrated (Jordan, 1999). Bacon, Dufek, Schreibman, Stahmer, Pierce, and Courchesne (2014) demonstrated support for the use of curriculum-based assessment with the positive correlation of Adapted Student Learning Profile scores (a curriculum assessment) with three other measures, the Mullen Scales of Early Learning, the Weschler Preschool and Primary Scales of Intelligence, and the Vineland Adaptive Behavior Scales for 45 young children with autism.

Invariably, other issues related to intervention effectiveness will arise, such as client attrition and parental adherence to the prescribed recommendations, which are important to capture for a most robust analysis identifying variables that lead to an effective program. Hayes, Barlow, and Nelson-Gray (1999, p.272) discuss the importance of moving beyond "gross measures of participation... to show that those who participated needed the care delivered, and others did not." This question is particularly relevant to define and when parents make their child available for a small fraction of the recommended therapeutic intervention or when children who require minimal treatment to address their deficits are in a program for the maximum number of therapeutic hours allowed by the funding source.

Impact objectives measure the long-term effectiveness of the intervention. For example, at a 5-year follow-up, were children exposed to *X* intervention less likely to require specialized services in the school environment? It is these measures that demonstrate the generalized effects that spread across behaviors, settings, time, and persons and ultimately justify the service to funding sources. Data collected are again likely to be a mix of quantitative and qualitative.

Following the identification of goals, objectives are developed. Objectives are activities that span a particular period that begins with program initiation and, for research programs,

ends with dissemination. For service providers, this cycle may involve evaluating cohorts and is likely to be an ongoing process in which unpredictable variables, such as changes in funding and service structures, must be incorporated in to the design in real time. This type of evaluation is called Developmental Evaluation (Patton, 2011).

Table 1 is an example of a program goal with process, impact, and outcome objectives for a parent training program to reduce the level of disparity in language and cognitive development in children from low-SES families by the age of three by increasing interactive parent-child communication. The actual program plan would be much more extensive, with process goals that spanned all of the necessary activities, and additional benchmark impact and outcome goals.

Goal: to reduce the level of disparity in language and cognitive development in children from low-SES families by the age of three by increasing interactive parent-child communication through a program called Language for Learning

Objectives:

Each objective provides a measure for one or more evaluation questions. Description of the data to be collected, a reference of the data source, a description of method of analysis, and person responsible for collecting the data must be defined for each. This rounds out the evaluation plan. Two examples from the objectives in Table 1, one process objective and one outcome objective, are presented in Table 2. Additionally, when possible, data from one source can be correlated or cross-referenced with data from another source measuring the same dimension as a method of enhancing the validity of information collected (Dymond, 2001).

Activities are often enumerated with the use of a Gantt or project management chart, which allows for the program planner to specify the month and year in which the activity will occur, who is responsible for implementing the program; collecting the data; and analyzing, compiling, and disseminating the data. Table 3 demonstrates the use of a timeline to conduct some of the program activities outlined in Table 1.

Conducting the Evaluation

Up to this point, there has been discussion on defining indicators and creating a plan for evaluation. The next stage is implementation of this plan.

Data collected are interpreted in the context of the program goals and objectives as outlined above. Evaluators should also consider the effects of other variables such as the choice of evaluation tools influencing the data collected (and what is not collected), client maturation, client exposure to other interventions and activities, bias of program staff conducting

Table 1 Example program goal with process, impact, and outcome objectives

Process	Outcome	Impact
1. By June 1, 2013, an advisory committee consisting of individuals from participating institutions and stakeholders will be formed.	1. By February 1, 2014, 80 % of program participants attending sessions 1 and 2 will name and demonstrate two strategies to increase parent-led language opportunities.	1. One year after program completion, 70 % of children whose parents completed at least 80 % of parent training sessions will demonstrate mid-average scaled scores (95–105) or above on the Bayley Scales of Infant Development (BSID)-III and the Preschool Language Scale (PLS), 5th Edition at program completion.
2. By July 1, 2013, the advisory committee will commence a monthly meeting schedule.	2. By March 1, 2014, 80 % of program participants attending sessions 1, 2, and 3 will name and demonstrate two strategies to increase child-led language opportunities.	
3. By August 1, 2013, assessment materials will be obtained and the assessor identified.	3. By April 1, 2014, 80 % of program participants will report feeling confident in using at least three of the strategies taught in parent trainings.	
4. By August 1, 2013, best practice strategies for increasing language development for children ages birth to 3 will be identified based on a literature review.		
5. By September 1, 2013, a parent training curriculum based on these strategies will be developed.		
6. By October 1, 2013, program materials (manual, handouts, incentives, etc.) will be created/obtained.		
7. By October 1, 2013, parent training session evaluations will be created.		
8. By November 1, 2013, a schedule of parent training sessions will be created and meeting locations will be identified.		
9. By December 1, 2013, program participants will be identified.		
10. By December 15, 2013, vendors for food and child care providers for the meetings will be identified.		
11. By January 20, 2014, all children in the program will have undergone baseline assessments.		
12. By January 20, 2014, the first monthly parent training meeting will be held.		
13. By January 20, 2014, program participants will complete session evaluations.		
14. By January 25, 2014 session evaluations will be analyzed. Feedback will be incorporated into future sessions.		
15. By June 30, 2014, the second LENA assessment will be conducted.		
16. By December 30, 2014, the third LENA assessment will be conducted.		
17. By June 30, 2015, impact assessments (PLS-5/BSID-III) will be conducted for children whose parents have completed the program.		
18. By June 30, 2016, program outcomes and impacts will be reported to stakeholders and attendees at a regional or national conference.		

measurement, and potential practice effects with repeated measurement.

Evaluation will necessitate the proper allocation of personal resources for monitoring evaluation activities. Generally,

this means accounting for these individuals' time. Additionally, a person must be assigned as a project manager to take the lead in keeping the team organized and compiling information in a timely manner. This allows program staff to

Table 2 Example evaluation plan

<i>Process objective 1:</i> By June 1, 2013, an advisory committee, consisting of individuals from participating institutions and stakeholders, will be formed.	<i>Data to be collected:</i> list of advisory committee members <i>Data source:</i> internal organizational documents <i>Method of analysis:</i> descriptive <i>Person responsible:</i> clinical director
<i>Outcome Objective 1:</i> By February 1, 2014, 80 % of program participants attending sessions 1 and 2 will name and demonstrate two strategies to increase parent-led language opportunities.	<i>Data to be collected:</i> participants' verbal responses <i>Data source:</i> written record <i>Method of analysis:</i> score correct responses per participant. Compile percent of participants who can name/demonstrate two strategies. <i>Person responsible:</i> trainer

make adjustments to the intervention and evaluation process when needed. As stated earlier, some programs rely on external evaluators or consultants to facilitate the process; however, internal staff must still complete much of the work. In terms of budget, The National Science Foundation (2002) estimates that evaluation activities account for about 10 % of a program's budget, but this cost may involve some initial investments (such as the purchase of assessment tools) that are one-time costs.

Just as in research, data must be compiled, described, and interpreted. However, program evaluations are less constrained by specific conventions and thus allow for more descriptive narrative around the findings that can provide context for the results.

Ethics

Finally, it should be stated that any evaluation process that involves collecting consumer information must follow all ethical guidelines of the profession including the protection of confidentiality and obtainment of consent from program participants in order for their data to be compiled into overall program outcomes. However, providers should not be dissuaded to conduct evaluation by confusion between research and evaluation. The goal of serving clients in a behavioral intervention program is to improve their functioning not to study and demonstrate the relations between events. Clients are enrolled in the program based on their needs and preferences. Thus, clients are not research participants and a

program evaluation is not subject to a review by an IRB (Hayes et al., 1999).

Reporting

Just as evaluations vary, so do methods of reporting. Reports should aim to meet the following purposes: make program improvements, demonstrate accountability, demonstrate effectiveness, and provide recommendations. A standard report should contain the following components: title page and table of contents, executive summary, purpose of the report, needs assessment data if available, background and description of the organization, stakeholders and priority population, general description of the program, goals, methodology, data collection and analysis, limitations of the evaluation, interpretations and conclusions, and recommendations for revisions. Appendices may include information about the measurement instruments and raw data. The disseminations of reports can vary from professional publication or presentation to the distribution of an executive summary to stakeholders and linking the final report on the service provider's website. For a variety of examples of program evaluation methods, tools, and final reports, *Evaluation and Program Planning* is a multidisciplinary scholarly journal dedicated to this topic.

Outside Consultants

The preceding information is meant to provide basic information about program evaluation and serve as a general framework for evaluation conducted by internal

Table 3 Example evaluation timeline

	Jan 2014	Feb 2014	Mar 2014	Apr 2014	May 2014	Jun 2014	Dec 2014	Jun 2015
Stakeholder meeting with staff: executive director/clinical director	X	X	X	X	X	X		
Parent trainings/training evaluations staff: trainer	X	X	X	X	X			
LENA assessment staff: clinical director	X					X	X	
BSID-III/PLS-5 assessments staff: clinical director								X

staff. It should be said that the use of outside consultants lends rigor and credibility, which may be important when results are distributed to stakeholders demanding accountability, such as funding agencies. The American Evaluators Association (www.eval.org) provides a list of independent consultants and evaluation firms by area of expertise and location. Evaluators are individuals who are generally not associated with any particular field or orientation and can bring a wider perspective to the process. Additionally, public health departments, the CDC, websites such as Community Toolbox (<http://ctb.ku.edu>), and universities can offer technical assistance.

Getting Started: a Basic Evaluation

Providers that have limited capacity need not be dissuaded. Smaller and medium capacity providers can and should still conduct some elements of a full-scale evaluation that will yield information regarding the efficacy of the intervention and organizational strengths and weaknesses. At minimum, agency staff can describe the population, provider, capacity, and intervention; compile participant outcomes using existing measures; compile existing process measures (e.g., staff hiring, training, turnover, etc.); conduct a consumer satisfaction survey; and report on these dimensions. The results of this level of evaluation can be shared internally and on the agency's website.

Conclusion

The use of program evaluation to assess the outcomes and impacts of behavior analytic services for children with autism is a method to confirm current best practices and to demonstrate accountability to stakeholders. Results of program evaluation can lead to the adoption of long-term strategies that increase quality, efficiency, and access to services. The use of program evaluation is still emerging in behavioral practice, and there are real and perceived barriers in conducting evaluations at the organizational level. Increasing knowledge of evaluation procedures and including components of evaluation practices into everyday service provision are important in the age in which services are increasingly funded by medical providers. Hayes et al. (1999) argue that “the analysis of population-based delivery systems requires that we move our questions and concepts to another level of analysis (p. 267).” It is incumbent upon practitioners to be accountable to the consumers of their services, their funders, and the academic discipline.

Compliance with Ethical Standards

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

- Ballard, D. W., & Grawitch, M. J. (2016). *The psychologically healthy workplace: building a win-win environment for organizations and employees*. Washington: American Psychological Association. doi:10.1037/14731-012.
- Bacon, E. C., Dufek, S., Schreibman, L., Stahmer, A., Pierce, K., & Courchesne, E. (2014). Measuring outcome in an early intervention program for toddlers with autism spectrum disorder: use of a curriculum-based assessment. *Autism Research and Treatment*. doi:10.1155/2014/264704Epub2014Mar10.
- Bishop-Fitzpatrick, L., Jung, N., Ilsung, N., Trunzo, A. C., & Rauktis, M. E. (2015). Outcomes of an agency-developed treatment foster care model for adolescents. *Journal of Emotional and Behavioral Disorders*, 23(3), 156–166. doi:10.1177/1063426614530470.
- Centers for Disease Control and Prevention (U.S. Department of Health and Human Services). Office of the Director, Office of Strategy and Innovation. (2011). *Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*. Atlanta: GA
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350. doi:10.1007/s10464-008-9165-0.
- Dyer, K., Martino, G. M., & Parvenski, T. (2006). The River Street Autism Program: a case study of a Regional Service Center behavioral intervention program. *Behavior Modification*, 30(6), 925–943. doi:10.1177/0145445506291395.
- Dymond, S. (2001). A participatory action research approach to evaluating inclusive school programs. *Focus on Autism and Other Developmental Disabilities*, 16(1), 54–63. doi:10.1177/10883576010600113.
- Gaugler, J. E. (2015). Describing and evaluating novel programs and therapies for older persons. *Journal of Applied Gerontology*, 34(5), 547–551. doi:10.1177/0733464815585681.
- Grunbaum, J. A., Gingiss, P., Orpinas, P., Batey, L. S., & Parcel, G. S. (1995). A comprehensive approach to school health program needs assessment. *Journal of School Health*, 65(2), 54–59. doi:10.1111/j.1746-1561.1995.tb03345.x.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner* (2nd ed.). Needham Heights: Allyn & Bacon.
- Heitzman-Powell, L. S., Buzhardt, J., Rusinko, L. C., & Miller, T. M. (2014). Formative evaluation of an ABA Outreach Training Program for parents of children with autism in remote areas. *Focus on Autism and Other Developmental Disabilities*, 29(1), 23–38. doi:10.1177/1088357613504992.
- Isaacs, B., & Ling, A. (2003). Evaluation of a short-term service for children and youth with developmental disabilities. *Journal on Developmental Disabilities*, 10(1), 93–100.
- Jordan, R. (1999). Evaluation practice: problems and possibilities. *Autism*, 3(4), 411–434. doi:10.1177/1362361399003004008.

- Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning*, 20(3), 251–257. doi:10.1016/s0149-7189(97)00002-5.
- McClannahan, L. E., & Krantz, P. J. (1997). Princeton Child Development Institute. *Behavior and Social Issues*, 7(1), 65–68. doi:10.5210/bsi.v7i1.301.
- McMahon, J., & Cullinan, V. (2014). Education programmes for young children with Autism Spectrum Disorder: an evaluation framework. *Research in Developmental Disabilities*, 34, 3689–3697. doi:10.1016/j.ridd.2014.09.004.
- McDougall, J., Servais, M., Meyer, K., Case, S., Dannenhold, K., Johnson, S., et al. (2009). Preliminary evaluation of a school support program for children with autism spectrum disorders: educator and school level outcomes and program processes. *Exceptionality Education International*, 19(1), 32–50.
- McKenzie, J. F., Neiger, B. L., & Thackeray, R. (2013). *Planning, implementing, and evaluating health promotion programs: a primer* (6th ed.). Boston: Pearson Education.
- National Science Foundation. *The 2002 User Friendly Guide to Project Evaluation*. Arlington, VA. <http://www.nsf.gov/pubs/2002/nsf02057/start.htm>
- Nutbeam, D., & Harris, E. (1999). *Theory in a nutshell: a guide to health promotion theory*. Sydney: McGraw-Hill.
- Partington, J. W. (2005). *The assessment of basic language and learning skills revised*. Walnut Creek: Behavior Analysts Inc.
- Patton, M. Q. (2011). *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.
- Perry, A., & Condillac, R. A. (2010). The TRE-ADD preschool parent training program: program evaluation of an innovative service model. *Journal on Developmental Disabilities*, 16(3), 8–16. doi:10.1037/t00362-000.
- Shakman, K., & Rodriguez, S. M. (2015). *Logic model for program design, implementation and evaluation: Workshop toolkit*. U.S. Department of Education: National Center for Education Evaluation and Regional Assistance
- Strain, P. (1987). Comprehensive evaluation of intervention for young autistic children. *Topics in Early Childhood Special Education*, 7(2), 97–110. doi:10.1177/027112148700700210.
- Swartz, I., Sandall, S. R., McBride, B. J., & Boulware, G.-L. (2004). Project DATA (Developmentally Appropriate Treatment for Autism): an inclusive school-based approach to educating young children with autism. *Topics in Early Childhood Special Education*, 24(3), 156–168. doi:10.1177/02711214040240030301.
- Trudgeon, C., & Carr, D. (2007). The impacts of home-based early behavioral intervention programmes on families of children with autism. *Journal of Applied Research in Intellectual Disabilities*, 20(4), 285–296. doi:10.1111/j.1468-3148.2006.00331.
- Truiette-Theodorson, R., Tuck, S., Bowie, J. V., Summers, A. C., & Kelber-Kaye, J. (2015). Building more effective partnerships to improve birth outcomes by reducing obesity: the B'more Fit for healthy babies coalition of Baltimore. *Evaluation and Program Planning*, 51, 53–58. doi:10.1016/j.evalprogplan.2014.12.007.
- University of California Division of Agriculture and Natural Resources. (2015). *Monument HEAL Zone strategy level evaluation: child care final report*. Concord: Contra Costa County Childcare Counsel.
- Vinton, L. (2012). Caregivers' perceptions of a consumer-directed care program for adults with developmental disabilities. In G. H. S. Singer, D. E. Biegel, & P. Conway (Eds.), *Family support and family caregiving across disabilities* (pp. 101–119). New York: Routledge/Taylor & Francis Group.
- Kellogg Foundation, W. K. (2004). *Using logic models to bring together planning, evaluation, and action: logic model development guide*. Battle Creek: W. K. Kellogg Foundation.