

*MULTIPLE-RESPONDENT ANECDOTAL ASSESSMENTS: AN ANALYSIS
OF INTERRATER AGREEMENT AND CORRESPONDENCE WITH
ANALOGUE ASSESSMENT OUTCOMES*

CARLA M. SMITH, RICHARD G. SMITH, JOSEPH D. DRACOBLY, AND
AMY PETERSON PACE

UNIVERSITY OF NORTH TEXAS

We evaluated interrater agreement across multiple respondents on anecdotal assessments and compared cases in which agreement was obtained with outcomes of functional analyses. Experiment 1 evaluated agreement among multiple respondents on the function of problem behavior for 27 individuals across 42 target behaviors using the Motivation Assessment Scale (MAS) and the Questions about Behavioral Function (QABF). Results showed that at least 4 of 5 respondents agreed on the primary maintaining consequence for 52% (22 of 42) of target behaviors with the MAS and 57% (24 of 42) with the QABF. Experiment 2 examined correspondence between the anecdotal assessment results and functional analysis results for 7 individuals for whom at least 4 of 5 respondents showed agreement in Experiment 1. Correspondence with functional analysis results was observed in 6 of 7 cases with the QABF and in 4 of 7 cases with the MAS. Implications of these outcomes for the utility of anecdotal assessments are discussed.

Key words: functional analysis, anecdotal assessment, Motivation Assessment Scale, Questions about Behavioral Function

Functional assessment procedures are designed to identify environmental variables that influence an individual's problem behavior, including antecedents that evoke the behavior and consequences that reinforce or maintain the behavior (Neef & Peterson, 2007). Three general assessment methods have emerged in the literature: descriptive assessment, in which information is obtained by direct observation in the natural environment (e.g., Bijou, Peterson, & Ault, 1968); anecdotal assessments, such as structured interviews or checklists (e.g., Durand & Crimmins, 1988b); and functional analysis (FA), in which environmental events are systematically manipulated to test the effect on behavior (e.g., Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994). The current

study focuses on FA and anecdotal assessment procedures.

Functional analysis has been widely studied and is considered to be the gold standard of functional assessment (Applegate, Matson, & Cherry, 1999; Hanley, Iwata, & McCord, 2003). The validity of FA has been demonstrated repeatedly through research showing that it results in identification of the maintaining reinforcer and serves as a basis for the development of effective function-based treatments (Asmus, Vollmer, & Borrero, 2002; Iwata, Vollmer, & Zarcone, 1990). Whereas functional relations between the environment and behavioral events are empirically demonstrated using FA, descriptive and anecdotal assessments infer those relations through observed correlations between environmental events and problem behavior (descriptive assessment) or caregiver reports (anecdotal assessment; Asmus et al., 2002). The FA, however, has several disadvantages, including the level of expertise required to implement procedures and interpret outcomes, the possibility that high-intensity behavior episodes will

Thanks to Shahla Ala'i-Rosales and Manish Vaidya for helpful suggestions. Joseph Dracobly is now at the University of Kansas.

Address correspondence to Richard G. Smith, Department of Behavior Analysis, 1155 Union Circle, Box 310919, University of North Texas, Denton, Texas 76203 (e-mail: Rick.Smith@unt.edu).

doi: 10.1901/jaba.2012.45-779

occur during assessment, and the possibility that treatments developed in a clinical setting may not be effective in the natural environment (Sturmey, 1995).

Anecdotal, or indirect, assessment involves the use of interviews, rating scales, checklists, or questionnaires to determine possible sources of reinforcement that maintain problem behavior. Anecdotal assessments are conducted by obtaining information from an individual who is presumably familiar with the circumstances surrounding the problem behavior. Respondents may include teachers, parents, direct-care staff, or, in some cases, the individual whose behavior is being assessed (Neef & Peterson, 2007). Anecdotal assessments have the advantage of being efficient, inexpensive, and easy to administer (Paclawskyj, Matson, Rush, Smalls, & Vollmer, 2001). Although widely used in practice, anecdotal assessments have well-known limitations. For example, they rely on respondents' memories and opinions instead of direct observation of the behavior in question (Neef & Peterson, 2007). In addition, anecdotal assessments provide information that permits hypotheses about the function of aberrant behavior but do not directly test these hypotheses. Finally, the reliability and validity of indirect assessments have been questioned (Bihm, Kienlen, Ness, & Poindexter, 1991).

One of the most widely used and studied anecdotal assessments, the Motivation Assessment Scale (MAS; Durand & Crimmins, 1988b), provides an illustrative example of some of the issues that surround indirect assessment. The developers used the MAS to assess the self-injurious behavior (SIB) of 50 developmentally disabled children. Teachers who worked with the children for the academic year served as primary respondents for the MAS, and their outcomes were compared with those from teacher aides from the same classrooms. Pearson correlation coefficients showed significant correlations between raters for the individual questions. Durand and Crimmins (1988a) concluded that the MAS is

“a reliable scale that can predict how individuals will behave in analogue assessment settings” (p. 113). The validity of MAS outcomes also has been evaluated by comparing FA (Carr & Durand, 1985) and MAS results. Results indicated that the outcomes of FA and MAS assessments matched in all cases.

Subsequent studies have produced mixed findings at best. For example, Zarcone, Rodgers, Iwata, Rourke, and Dorsey (1991) did not replicate the outcomes of Durand and Crimmins (1988a), showing agreement on the source of reinforcement between only 16 of 55 rater pairs and low Pearson correlations. Several additional studies have reported similar outcomes (e.g., Newton & Sturmey, 1991; Paclawskyj *et al.*, 2001; Sigafos, Kerr, & Roberts, 1994; Thompson & Emerson, 1995). Thus, although Durand and Crimmins (1988a) reported encouraging outcomes for the MAS, several subsequent investigations have been unable to replicate their findings.

The Questions about Behavioral Function (QABF; Matson & Vollmer, 1995) is another widely used anecdotal assessment. Several studies indicate that the QABF appears to have good test–retest reliability, interrater agreement, and stability (Paclawskyj, Matson, Rush, Smalls, & Vollmer, 2000), is often able to determine a clear behavioral function (Matson, Bamburg, Cherry, & Paclawskyj, 1999), and has treatment utility (Matson *et al.*, 1999). Paclawskyj *et al.* (2001) performed an evaluation of convergent validity among the QABF, MAS, and FAs and found that FA outcomes agreed with QABF results in 69.2% of cases and with MAS outcomes in 53.8% of cases. A study of key psychometric properties of the MAS and the QABF (Shogren & Rojahn, 2003) showed less than satisfactory interrater agreement, with both scales falling into the *fair* to *good* range. Both assessments were found to be comparable in measuring similar constructs and in terms of reliability.

Although the literature has shown generally low correspondence between anecdotal assess-

ments and FAs, practitioners use anecdotal assessments in clinical settings, schools, and institutional facilities. The current study explored conditions under which anecdotal assessments may provide useful information that can be integrated within a comprehensive functional assessment process and extend the literature by investigating the potential utility of administering anecdotal assessments with multiple respondents. We evaluated the extent of agreement among five respondents for the QABF and MAS, and evaluated correspondence with FA outcomes for a sample of participants from each of four subscale categories (attention, tangible, escape, and sensory) for whom we obtained substantial agreement.

EXPERIMENT 1

Method

Participants and setting. This study was conducted at a large, state-sponsored residential facility for individuals with intellectual disabilities. Assessments were administered in secluded areas of the residential apartments or quiet areas in the vocational training setting.

Residents. Twenty-seven individuals who resided at the facility participated in Experiment 1. Their ages ranged from 27 to 66 years, and all had been diagnosed with intellectual disabilities. Table 1 shows each individual's age and functioning level.

Target behaviors. All participants had a history of problem behavior of sufficient severity to necessitate the development of behavior support plans. The behavioral definitions used in Experiment 1 were developed by the individuals' unit psychologists and were part of each individual's behavior support plan. Twelve participants presented with a single target behavior, and 15 presented with two target behaviors, for a total of 27 participants with 42 target behaviors. The MAS and QABF were completed for each target behavior; thus, for residents who presented with two target behaviors, 20 assessments were completed (five

MAS and five QABF for each target behavior). Due to an error in administration of the assessments, data from only four respondents are reported for Barbara. Table 1 includes a description of residents' target behaviors and definitions.

Respondents. Respondents for Experiment 1 were 113 staff members who had worked regularly with residents as direct-care, vocational, or unit management staff for a minimum of 6 months. Their educational backgrounds were unavailable, because prior to the study the facility had discontinued a longstanding hiring policy that required a high-school education or equivalent for employment. All respondents were employees of the facility at the time of the interviews. Multiple respondents (typically five) were interviewed for each resident's target behavior. This number of respondents was chosen because it seemed reasonable that it would be possible to identify five caregivers who had sufficient histories with participants to provide meaningful responses to the assessment items.

Materials. Materials used in Experiment 1 included writing utensils and two sets of each questionnaire (MAS and QABF). The general information sections of each assessment (name, residence, date, rater, target behavior, etc.) were completed by the interviewer before the assessment. Interviewers read aloud and marked the answers stated on one set of questionnaires while respondents read along with the second set.

MAS. The MAS (Durand & Crimmins, 1988b) is a 16-question assessment with four questions that correspond to each of four subscale categories: escape, sensory, attention, and tangible. Respondents answered questions using a 7-point Likert-type scale; scores indicated the extent to which the rater observed the behavior, from 0 (*never*) to 6 (*always*). Intermediate values allowed respondents to score 1 (*almost never*), 2 (*seldom*), 3 (*half of the time*), 4 (*usually*), or 5 (*almost always*). The four questions that corresponded to each category were summed and ranked by point value. The category with the

Table 1
Residents' Demographic Information, Target Behaviors, and Topographical Descriptions

| Resident | Age | Functioning level | | Target behavior |
|----------|-----|-------------------|------|--|
| Annie | 31 | Profound | PAO | Scratching, hitting, pinching, biting, pushing, or grabbing another person |
| Asa | 51 | Profound | AGP | Biting or overturning furniture, slamming doors |
| | | | SIB | Banging head on hard surfaces, biting hand or finger |
| Barbara | 41 | Severe | SIB | Hand mouthing |
| Carl | 39 | Mild | VDB | Verbally abusive or threatening behavior |
| | | | PAO | Hitting, scratching, grabbing, kicking, or spitting at others |
| Chad | 27 | Profound | SIB | Head striking any object (including people) |
| | | | PAO | Making contact with another person with sufficient force to cause injury |
| Derek | 40 | Profound | PDB | Bucking in wheelchair, hitting, or pushing objects |
| | | | PAO | Pushing, kicking, slapping, or biting others |
| Donnie | 56 | Profound | PDB | Displacing training materials, overturning furniture, throwing objects, stripping beds, and expelling mucus |
| | | | VDB | Brief loud yelling or screaming |
| Garfield | 47 | Profound | PAO | Hitting, biting, pinching, shoving, or pushing others |
| | | | STE | Attempting to steal food items |
| Genna | 57 | Mild | VDB | Yelling, threatening, cursing, or whining |
| Greg | 46 | Profound | Pica | Ingesting nonfood items |
| Jack | 56 | Moderate | VDB | Yelling |
| | | | PAO | Hitting others |
| Jerry | 66 | Profound | PAO | Hitting, kicking, biting, scratching, spitting at, pushing, or throwing objects at others |
| | | | RUM | Bringing up and rechewing stomach contents |
| Peter | 29 | Mild | SIB | Hitting self with hand on any part of body or hitting head or hand against an object |
| Joe | 52 | Profound | PAO | Hitting or wrapping arms around others and bringing them to the ground |
| Jolinda | 55 | Profound | SIB | Biting fingers or hands, head banging, slapping self, picking at skin, sores, or scabs |
| Jon | 34 | Severe | STE | Twirling shirt, tapping on objects, hoarding items, seeking out object to the exclusion of anything else |
| Karen | 47 | Profound | VDB | Yelling, screaming, and crying |
| | | | SIB | Hand biting |
| Kate | 46 | Profound | MO | Mouthing objects |
| Mark | 50 | Profound | SIB | Scratching or rubbing skin, striking self, hitting elbow or body part against hard object or surface |
| Marion | 58 | Profound | Pica | Attempting to ingest nonfood items |
| | | | SIB | Hand mouthing |
| Martin | 48 | Profound | PDB | Slapping tables and walls, throwing materials, stripping, grabbing others' clothes while yelling |
| Mike | 50 | Profound | PAO | Pushing others |
| | | | STE | Taking food items from others |
| Peg | 50 | Severe | VDB | Yelling or continuously talking for 1 min or more about inappropriate topics (e.g., telling on others or blaming others) |
| Rob | 49 | Profound | VDB | Yelling and screaming |
| | | | PDB | Slamming doors, pounding windows, dropping to floor, wheelchair bucking, and public masturbation |
| Ted | 52 | Profound | PAO | Kicking, biting, and hitting |
| Vern | 41 | Profound | Pica | Ingesting nonfood items; insertion of simulated pica items into the mouth |
| Vynita | 48 | Profound | SIB | Picking or scratching sores and scabs |
| | | | STE | Taking food or drink |

Note. PAO = physical aggression to others, STE = stereotypy, SIB = self-injurious behavior, RUM = rumination, PDB = physical disruptive behavior, VDB = verbal disruptive behavior, AGP = aggression toward property, MO = mouthing.

highest value was assumed to represent the maintaining consequence.

QABF. The QABF (Matson & Vollmer, 1995) is a 25-question assessment with five questions assigned to each of five subscales: attention, escape, nonsocial, physical, and

tangible. The scale allowed respondents to use a 4-point Likert-type scale to score how often the client demonstrated the target behavior. Respondents chose from 0 (*never*), 1 (*rarely*), 2 (*some*), and 3 (*often*). Once completed, the assessments were scored for the number of

items endorsed (i.e., if a question was answered) and for the point total assigned to each subscale category. Subscale categories were then ranked according to the score, and the category with the highest point value was assumed to represent the maintaining consequence.

Administration procedures. Twenty-three graduate and undergraduate students (interviewers) were trained to administer the QABF and MAS. Training included reading and discussion of the MAS and QABF manuals, role-playing administration procedures with a senior trainer who provided feedback, observing a senior trainer administer an assessment, and receiving feedback following initial assessment delivery. To assist respondents who may have had difficulty reading the assessment, each respondent was given a copy of the questionnaire to read along as the interviewer read each question aloud and scored answers. Interviewers confirmed at the start of each interview that the respondent had worked with the resident for 6 months or longer and then read the general information, including the definition of the target behavior, before reading assessment questions. Interviewers administered assessments in quiet, secluded areas of residential or vocational buildings to ensure that potential respondents (i.e., other staff members) were not present and could not hear respondents' answers. Interviewers read questions aloud to respondents exactly as written; no additional information or clarification was given. If respondents asked questions, they were told to "answer the best you can." The MAS was administered first, followed by the QABF. After both instruments had been completed, the interviewer thanked the respondent and left the area. If an individual presented with two target behaviors, interviewers returned at another time to administer assessments for the second target behavior.

Respondent agreement evaluation. Two trained graduate students scored each assessment. Resulting scores were compared on a ques-

tion-by-question basis. Agreement in the scoring for both the MAS and QABF was 100%.

Agreement across assessments (within respondents). Agreement across assessments and within respondents was scored if the respondent identified the same maintaining consequence with both instruments. As noted previously, the MAS was organized according to four categories of maintaining variables (sensory, escape, attention, and tangible). The QABF was organized according to five categories of maintaining variables (nonsocial, escape, attention, tangible, and physical). Agreement was scored if a respondent ranked sensory as the maintaining variable on the MAS and either nonsocial or physical as the maintaining variable on the QABF. If a respondent scored two categories as the highest ranking (i.e., there was a tie) on one assessment, both categories were compared with the highest ranking category from the other assessment. For example, agreement was scored if a respondent scored both attention and escape as the highest ranking categories on the QABF (i.e., attention and escape received the same score, which was higher than scores for other categories) and scored attention as the highest ranking category on the MAS.

Agreement across assessments (across respondents). Agreement across assessments (across respondents) was scored if four or five of the five respondents for each resident identified the same maintaining variable for both the MAS and QABF.

Agreement within assessments (across respondents). Agreement within assessments (across respondents) was scored if four or five of the five respondents for each resident identified the same maintaining variable on either the MAS and QABF.

Results and Discussion

Five respondents completed both the MAS and QABF for each target behavior (except for Barbara, as noted previously). Table 2 shows respondent agreement within and across assess-

ments. For the MAS, four or five of the five respondents agreed on a primary maintaining consequence for 52% (22 of 42) of target behaviors. For the QABF, four or five of the five respondents agreed on a primary maintaining consequence for 57% (24 of 42) of target behaviors. Respondents were in agreement across both MAS and QABF for 26% (11 of 42) of target behaviors. Perfect agreement (five of five) occurred with 12 (29%) of MAS respondents, seven (17%) of QABF respondents, and three (7%) respondents across both the MAS and QABF.

Table 3 shows the number of respondents who identified specific categories of maintaining consequences across individuals and target behaviors. Within-assessment ties between primary maintaining consequences account for instances in which the number of identified maintaining consequences was greater than five for a given target behavior. Data in boldface type indicate target behaviors for which four or five respondents agreed on the primary category of maintaining consequence. Data in boldface that extend across both assessments indicate across-respondent agreement for both the MAS and the QABF. These data offer a detailed view of the summary data presented in Table 2 and Figure 1.

Figure 1 shows percentages of primary categories of maintaining variables across respondents for the MAS and QABF (there were 210 responses for each assessment tool). MAS respondents scored sensory as the primary maintaining variable 82 times (39%); a tangible consequence was identified 81 times (38.6%); an attention consequence was identified 40 times (19%); and an escape consequence was identified 31 times (14.8%). Sixteen respondents' MAS scores showed a two-way tie between categories and four showed a three-way tie, resulting in a total number of 234 consequences identified by the 210 respondents (111%).

Of the 210 responses to the QABF, 63 (30%) identified escape as the primary main-

taining consequence; nonsocial reinforcement was identified 62 times (29.5%); a tangible consequence was identified 60 times (28.6%); an attention consequence was identified 38 times (18.1%); and physical reinforcement was identified 14 times (6.7%). QABF respondents scored 26 two-way ties and one three-way tie between primary maintaining variables, resulting in a total number of 237 contingencies identified by the 210 respondents (113%).

Experiment 1 showed agreement across four or five raters in 55% of all assessments administered. Agreement occurred for 22 of 42 target behaviors (52%) in MAS assessments and in 24 of 42 target behaviors (57%) in QABF assessments. The outcomes for the MAS are consistent with the findings reported by Fahrenholz (2004), which showed that 15 of 28 (54%) of MAS assessments showed agreement across at least four of five respondents.

The results of Experiment 1 did not reveal substantial differences in agreement between the QABF and the MAS. The QABF produced slightly greater overall agreement, but was also more likely to show ties between identified categories than the MAS. The QABF produced 27 ties, whereas the MAS produced 20 ties. These results suggest that the overall agreement percentages for the QABF may have been artificially inflated based on an increased tendency for individual raters to identify more than one primary contingency. That is, less differentiation among responses within individual raters (which is not a desirable quality for an assessment instrument) makes it more likely that correspondence across respondents will be observed. In effect, if a common response to the question "Which of these contingencies is responsible for this person's problem behavior?" is "one of two (or three) contingencies" instead of "this one particular contingency," then more opportunities for agreement with other raters will be available. Thus, the instrument that produces less certainty within raters will, logically, produce more agreement across raters. It should be noted that, in some

Table 2
Respondent Agreement Across and Within Assessments

| | Total agreement | 4 of 5 agreement | 5 of 5 agreement |
|--------------|-----------------|------------------|------------------|
| QABF | 24 of 42 (57%) | 17 of 42 (40%) | 7 of 42 (17%) |
| MAS | 22 of 42 (52%) | 10 of 42 (24%) | 12 of 42 (29%) |
| MAS and QABF | 11 of 42 (26%) | 8 of 42 (19%) | 3 of 42 (7%) |

cases, within-rater ties also may occur as a function of multiple maintaining contingencies (i.e., problem behavior may, in fact, be maintained by multiple contingencies of reinforcement). Therefore, it is not possible to determine if higher levels of agreement within the QABF are due to a better ability to identify multiple controlling contingencies or a decreased ability to distinguish among possible maintaining contingencies.

There were some limitations to Experiment 1. First, target behaviors were identified and defined by residents' unit psychologists and support teams prior to the start of the study. Some target behavior definitions included multiple topographies, which may have increased the possibility of maintenance by different consequences. For example, Jolinda's target behavior was SIB, which was defined as biting fingers or hands, head banging, and slapping herself. It is possible that these different topographies of behavior were maintained by different consequences.

Another limitation was that excessive turnover among staff (respondents), reported by the facility to be 64.6% in 2009 (S. Musgrave, personal communication, February 18, 2010), made it impossible to include only respondents who had known individuals for 1 year or longer. Thus, although the administration guide for the MAS suggests that respondents should be acquainted with individuals being assessed for at least 1 year (the QABF administration manual suggests that informants should know individuals for at least 6 months), it was necessary to obtain responses from caregivers who had known individuals for as little as 6 months in the current study. All staff

had worked with the residents for a minimum of 6 months, and some staff had worked with individuals for more than 10 years. It is plausible that differences in scoring and, thus, the ability to identify maintaining variables were at least in part a function of the length of time staff had worked with an individual. The rate at which a target behavior occurred also may have affected staff's ability to accurately identify variables with anecdotal assessments. For example, it may be easier to identify the environmental variable associated with behavior that occurs more frequently in natural settings. Future research should investigate how differences in settings, staff tenure, and rate of problem behavior affect anecdotal assessment outcomes.

Based on the results of Experiment 1, a sample of residents whose anecdotal assessments showed agreement across respondents was selected to participate in Experiment 2, during which we conducted FAs of each participant's problem behavior.

EXPERIMENT 2

Method

Participants, setting, and materials. All sessions were conducted in a clinic for the assessment and treatment of behavior disorders, located on the campus of the residential facility at which Experiment 1 was conducted. Sessions were conducted in one of the clinic's observation rooms. Rooms (3.7 m by 3.7 m) contained a table, two chairs, and materials appropriate for the experimental session. A one-way mirror was installed in one wall of each room for

Table 3
Individual Results, with Respondent Groups Listed Across Primary Maintaining Consequences

| Resident | Behavior | QABF | | | | | MAS | | | |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | | N/S | ATT | TAN | ESC | PH | SEN | ATT | TAN | ESC |
| Annie | PAO | 0 | 0 | 4 | 3 | 1 | 0 | 0 | 5 | 0 |
| Asa | AGP | 2 | 0 | 1 | 3 | 0 | 0 | 2 | 1 | 3 |
| | SIB | 0 | 0 | 1 | 5 | 0 | 1 | 0 | 3 | 3 |
| Barbara | SIB | 0 | 4 | 0 | 0 | 0 | 1 | 4 | 0 | 0 |
| Carl | VDB | 0 | 0 | 1 | 5 | 0 | 1 | 1 | 3 | 0 |
| | PAO | 1 | 1 | 0 | 3 | 0 | 1 | 2 | 3 | 1 |
| Chad | SIB | 0 | 4 | 2 | 1 | 0 | 2 | 0 | 5 | 1 |
| | PAO | 0 | 2 | 3 | 1 | 0 | 0 | 1 | 5 | 0 |
| Derek | PDB | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 5 | 1 |
| | PAO | 0 | 0 | 0 | 4 | 1 | 0 | 0 | 4 | 1 |
| Donnie | PDB | 4 | 1 | 0 | 0 | 0 | 2 | 2 | 2 | 0 |
| | VDB | 3 | 2 | 0 | 0 | 1 | 3 | 1 | 1 | 0 |
| Garfield | PAO | 0 | 0 | 2 | 3 | 0 | 1 | 0 | 2 | 2 |
| | STE | 0 | 0 | 5 | 1 | 0 | 3 | 0 | 2 | 0 |
| Greg | Pica | 4 | 0 | 0 | 1 | 0 | 5 | 0 | 0 | 0 |
| Genna | VDB | 0 | 0 | 1 | 5 | 0 | 0 | 1 | 3 | 3 |
| Jack | VDB | 0 | 1 | 1 | 4 | 0 | 0 | 3 | 2 | 0 |
| | PAO | 0 | 1 | 2 | 4 | 0 | 0 | 4 | 3 | 0 |
| Peter | SIB | 1 | 4 | 0 | 0 | 0 | 4 | 1 | 0 | 0 |
| Jerry | PAO | 0 | 0 | 4 | 1 | 0 | 1 | 0 | 3 | 1 |
| | RUM | 5 | 0 | 1 | 0 | 0 | 5 | 0 | 0 | 0 |
| Joe | PAO | 0 | 3 | 2 | 1 | 0 | 1 | 3 | 0 | 1 |
| Jolinda | SIB | 0 | 2 | 4 | 0 | 0 | 0 | 2 | 4 | 0 |
| | Pica | 2 | 0 | 3 | 0 | 0 | 2 | 2 | 1 | 0 |
| Jon | STE | 1 | 1 | 3 | 1 | 0 | 3 | 1 | 1 | 0 |
| Karen | VDB | 0 | 0 | 1 | 4 | 0 | 3 | 0 | 2 | 4 |
| | SIB | 0 | 0 | 1 | 3 | 1 | 1 | 1 | 2 | 1 |
| Kate | MO | 3 | 1 | 0 | 1 | 1 | 5 | 0 | 0 | 0 |
| Martin | PDB | 4 | 1 | 0 | 0 | 1 | 4 | 1 | 0 | 0 |
| Mark | SCR | 4 | 0 | 0 | 0 | 2 | 5 | 0 | 0 | 1 |
| | SIB | 4 | 0 | 1 | 0 | 0 | 4 | 0 | 1 | 0 |
| Marion | Pica | 2 | 2 | 1 | 0 | 0 | 1 | 0 | 4 | 0 |
| | SIB | 1 | 2 | 2 | 0 | 0 | 1 | 0 | 3 | 1 |
| Mike | PAO | 1 | 0 | 3 | 1 | 0 | 0 | 0 | 5 | 0 |
| | STE | 1 | 0 | 4 | 0 | 0 | 4 | 1 | 0 | 0 |
| Peg | VDB | 2 | 2 | 0 | 1 | 1 | 2 | 1 | 1 | 1 |
| Rob | VDB | 2 | 1 | 1 | 2 | 1 | 0 | 3 | 1 | 2 |
| | PDB | 1 | 1 | 0 | 2 | 1 | 0 | 3 | 1 | 1 |
| Ted | PAO | 1 | 2 | 2 | 0 | 0 | 1 | 0 | 3 | 2 |
| Vern | Pica | 5 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 |
| Vynita | SIB | 5 | 0 | 0 | 0 | 1 | 5 | 0 | 0 | 1 |
| | STE | 3 | 0 | 4 | 0 | 0 | 5 | 0 | 0 | 0 |

Note. On the QABF, N/S = nonsocial; ATT = attention; TAN = tangible; ESC = escape; PH = physical. On the MAS, SEN = sensory; ATT = attention; TAN = tangible; ESC = escape. Data in boldface indicate agreement of four or five respondents. See Table 1 for definitions of behaviors (SCR = scratching).

unobtrusive observation and recording of session data.

For pica assessments, materials included simulated nonfood items that could be safely consumed by participants. Simulated pica items for Greg were all-natural soap (made from edible oils and wax) placed in a soap dish;

mixtures of water, white vinegar, apple cider vinegar, Simply Thick Gel, and food coloring placed in bath gel bottles; a hand sanitizer pump; and a spray bottle, to simulate bath products, cleaning chemicals, and hand sanitizer. Simulated pica items for Vern were rice paper to simulate paper, onion skins to simulate

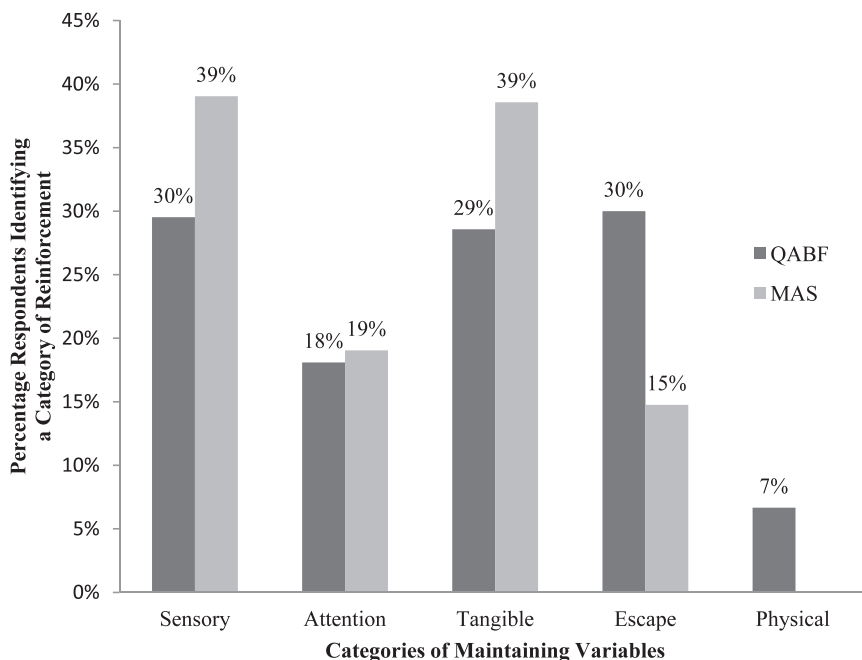


Figure 1. Percentage of respondents identifying particular categories of maintaining consequences on the QABF (black bars) and the MAS (gray bars).

paper and leaves, dried seaweed to simulate leaves, and brown fettuccini to simulate leaves and twigs. These items were continuously available throughout all FA sessions for both participants.

Participants. Eight individuals were selected from among Study 1 participants for whom at least four of the five respondents showed agreement within the QABF and, with the exception of one case (Asa), the MAS (see Table 1 for demographic information). Participants were selected to obtain representation from all subscale categories that showed agreement among at least one group of respondents (nonsocial, attention, tangible, and escape) and based on availability and continuing need for behavioral intervention. Greg (46-year-old man) and Vern (41-year-old man) had been diagnosed with profound mental retardation. Both men engaged in pica, and their anecdotal assessments indicated that pica was maintained by nonsocial reinforce-

ment. Jolinda was a 55-year-old woman who had been diagnosed with profound mental retardation and who exhibited SIB. Annie was a 33-year-old woman who had been diagnosed with profound mental retardation and who engaged in physical aggression to others (PAO). Anecdotal assessment results indicated that target behaviors for Jolinda and Annie were maintained by positive reinforcement in the form of tangible items. Karen was a 47-year-old woman who had been diagnosed with profound mental retardation and who engaged in verbal disruptive behavior (VDB). Asa was a 51-year-old man who had been diagnosed with profound mental retardation and who exhibited SIB. QABF assessment results for Karen and Asa indicated that their problem behaviors were maintained by negative reinforcement in the form of escape from task demands; Asa’s MAS outcomes were inconclusive. Chad was a 27-year-old man who had been diagnosed with profound mental retar-

dition, and Peter was a 29-year-old man who had been diagnosed with mild mental retardation. Peter's FA was conducted prior to the current study, and those data were initially presented by Dracobly and Smith (2012). Both men engaged in SIB that, according to QABF results, was maintained by attention. Chad's MAS results indicated that his problem behavior was maintained by tangible reinforcement, and Peter's MAS results indicated that his problem behavior was maintained by automatic reinforcement. Chad sustained a fall, unrelated to his participation in the study, after participating in only four sessions of his FA. Based on injuries related to the fall and concerns about the severity of his SIB, he was not eligible for further participation in the study and, therefore, it was not possible to evaluate correspondence between his FA and anecdotal assessments.

Target behaviors. The operational definitions used for the FA were based on residents' target behavior definitions found in their formal behavior support plans. Definitions were refined for the FA when necessary based on direct observations of problem behavior and a review of the client's records, including the daily reports of problem behavior across all settings. Greg and Vern's target behavior was *pica*, defined as the insertion of simulated pica items into the mouth. Jolinda's target behavior was *SIB*, defined as biting her fingers or hands, head banging, or slapping herself. Annie's target behavior was *PAO*, defined as scratching, hitting, pinching, biting, pushing, or grabbing another person. Karen's target behavior was *VDB*, defined as yelling, screaming, or crying. Asa's target behavior was *SIB*, defined as banging his head on hard surfaces or biting his hand or finger. Chad's target behavior was *SIB*, defined as his head striking any object (including people). Peter's target behavior was *SIB*, defined as hitting himself with his hand on any part of his body (e.g., head, face, chin, forehead, leg, etc.) or hitting his head or hand against an object. For Peter, *head up*, defined as

no part of his chin or neck touching his chest or shoulders, also was scored.

Functional Analysis

Observation procedures. Trained observers used handheld computers to record target behaviors. Frequency measures were used for pica (Greg and Vern); head banging (Asa, Chad, and Peter); hitting body parts against objects (Peter); hitting others (Annie); biting (Annie); pinching (Annie); hitting self (Peter); head-up (Peter); and scratching, pushing, or grabbing others (Annie). Duration measures were used for yelling, screaming, and crying (Karen); head banging (Jolinda); slapping self (Jolinda); and biting hands or fingers (Asa and Jolinda).

Interobserver agreement. A second observer independently and simultaneously scored 85% of Greg's sessions, 65% of Vern's sessions, 80% of Jolinda's sessions, 60% of Annie's sessions, 64% of Karen's sessions, 63% of Asa's sessions, and 38% of Peter's sessions. Interobserver agreement was calculated by dividing each session into 1-s intervals, summing the number of intervals in which the primary and secondary observers agreed on the occurrence or nonoccurrence of the target behavior, dividing the result by the total number of intervals in the session, and converting the outcome to a percentage. Agreement was calculated slightly differently for Peter due to difficulty in determining the exact second of the onset of head up. Therefore, agreement for Peter's sessions was calculated as above but with a moving 2-s window (e.g., if the primary observer recorded an event at time x , agreement was scored if the secondary observer recorded the same event at time $x - 1$ s, time x , or time $x + 1$ s), dividing the result by the total number of intervals in the session, and converting the result to a percentage. Mean interobserver agreement was 99% (range, 85% to 100%) for Peter, 99% (range, 99% to 100%) for Greg, 99% (range, 97% to 100%) for Vern, 99% (range, 96% to 100%) for Jolinda, 96% (range,

80% to 100%) for Annie, 98% (range, 89% to 100%) for Karen, and 98% (range, 91% to 100%) for Asa.

General procedures. Procedures were similar to those described by Iwata et al. (1982/1994; contact the authors for a detailed description of procedures for each condition). All eight participants were exposed to three test conditions (alone or no interaction, attention, and demand) and a control condition (play), presented in a multielement format. Six of the participants (excluding Greg and Vern) also were exposed to a tangible condition. Each session lasted 10 min. One to six sessions were conducted per day in the following order: alone or no interaction, attention, play, tangible (if relevant), and demand. Sessions were conducted at the same time each day, 3 to 5 days a week, and the number of sessions conducted each day was arranged so as to start with a different session on successive days (i.e., no day ended with a complete cycle through conditions) so that sequencing patterns were unlikely to develop. Graduate students who were trained in facility protocols for management of aggression, protection of human subjects, and cardiopulmonary resuscitation served as therapists. Because of the high intensity of his SIB, an FA of precursor behavior was conducted for Peter (Smith & Churchill, 2002). Experimental contingencies were in effect for his precursor behavior (head up) during the analysis, and no consequences were provided for SIB. The operant function of his SIB was inferred from outcomes of the precursor assessment.

Results and Discussion

Results of each participant's FA are presented in Figures 2 and 3. Table 4 presents a comparison of anecdotal assessment results and FA results.

Annie. Annie's FA results are shown in Figure 2. Following one cycle of conditions during which no responding was observed, PAO occurred at consistently high frequencies in tangible sessions ($M = 32.5$ responses per session)

and at lower frequencies in the no-interaction ($M = 5.5$ responses per session), attention ($M = 4$ responses per session), play ($M = 2.2$ responses per session), and demand ($M = 2$ responses per session) conditions. These outcomes indicated that PAO was maintained by social positive reinforcement in the form of access to tangible items. Outcomes of both the MAS (five of five respondents) and QABF (four of five respondents) corresponded to the FA results.

Greg. Greg's results are shown in Figure 2. Pica occurred exclusively in the alone condition and remained at zero in the attention, play, and demand conditions, suggesting an automatic reinforcement function for his pica. Four of five respondents to the QABF identified a nonsocial function, and five of five respondents to the MAS identified a sensory function, demonstrating perfect correspondence with the results of the FA.

Vern. Vern's FA results are displayed in Figure 2. Pica occurred across all four conditions, but differentiation can be seen from Sessions 12 through 20, with highest levels of pica in the alone condition ($M = 8$ responses per session), followed by attention ($M = 5.1$ responses per session), play ($M = 3.1$ responses per session), and demand ($M = 1$ response per session) conditions. For both the MAS and QABF, five of five respondents indicated a nonsensory or automatic reinforcement function of Vern's pica; thus, anecdotal assessment results corresponded perfectly with results from the FA.

Jolinda. Jolinda's FA results are displayed in Figure 2. She engaged in more SIB during the alone condition ($M = 3.4$ responses per session) than during other conditions; however, she also engaged in SIB during other test conditions (tangible $M = 0.72$ responses per session; demand $M = 0.36$ responses per session; attention $M = 0.18$ responses per session; control $M = 0$ responses per session). No SIB occurred during the final three cycles of assessment. Based on these inconsistent and largely undifferentiated outcomes, the FA does



Figure 2. Functional analysis results for Annie, Greg, Vern, and Jolinda.

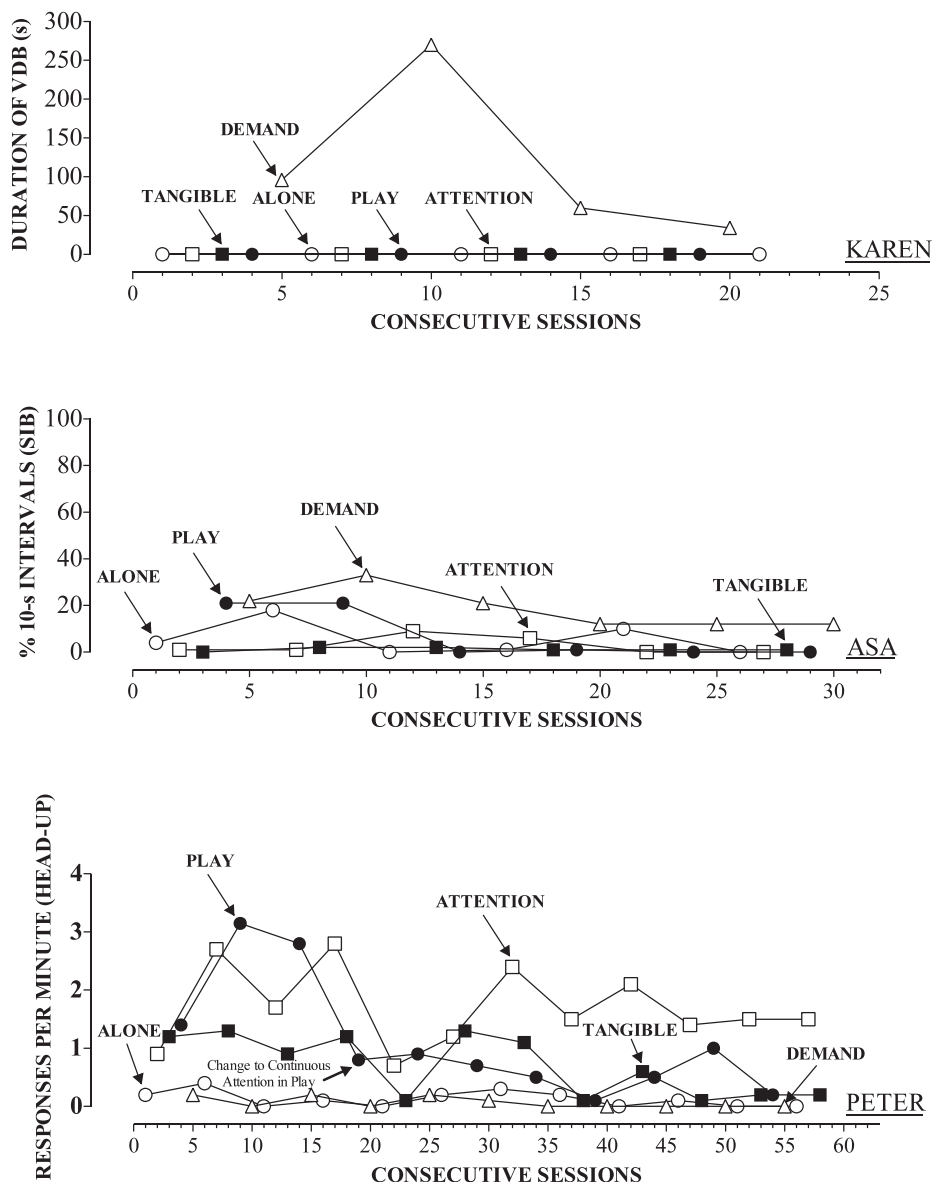


Figure 3. Functional analysis results for Karen, Asa, and Peter. Peter's data are reproduced from Dracobly and Smith (2012). VDB = verbal disruptive behavior.

not provide differential support for any account of Jolinda's SIB. These outcomes show no correspondence with those from both anecdotal assessments, in which four of five respondents identified social positive reinforcement in the form of tangible items as the likely maintaining contingency for Jolinda's SIB.

Karen. Karen's FA results are shown in Figure 3. Verbal disruptive behavior occurred exclusively in the demand condition ($M = 115$ s per session), strongly indicating that her problem behavior was maintained by negative reinforcement in the form of escape from task demands. Results from both the MAS (four of

Table 4
Correspondence Between Anecdotal Assessments and the Functional Analysis

| Resident | Behavior | QABF | | | | | MAS | | | | FA |
|----------|----------|----------|----------|----------|----------|----|----------|-----|----------|----------|------------|
| | | N/S | ATT | TAN | ESC | PH | SEN | ATT | TAN | ESC | |
| Annie | PAO | 0 | 0 | 4 | 3 | 1 | 0 | 0 | 5 | 0 | TAN/agree |
| Asa | SIB | 0 | 0 | 1 | 5 | 0 | 1 | 0 | 3 | 3 | ESC/agree |
| Greg | Pica | 4 | 0 | 0 | 1 | 0 | 5 | 0 | 0 | 0 | N/S/agree |
| Peter | SIB | 1 | 4 | 0 | 0 | 0 | 4 | 1 | 0 | 0 | ATT/agree |
| Jolinda | SIB | 0 | 2 | 4 | 0 | 0 | 0 | 2 | 4 | 0 | Unclear FA |
| Karen | VDB | 0 | 0 | 1 | 4 | 0 | 3 | 0 | 2 | 4 | ESC/agree |
| Vern | Pica | 5 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | N/S/agree |

Note. On the QABF, N/S = nonsocial; ATT = attention; TAN = tangible; ESC = escape; PH = physical. On the MAS, SEN = sensory; ATT = attention; TAN = tangible; ESC = escape. Data in boldface indicate agreement of four or five respondents.

five respondents) and QABF (four of five respondents) corresponded to the FA results.

Asa. Asa's FA results are shown in Figure 3. SIB consistently occurred during a high percentage of intervals in the escape condition ($M = 18.7\%$ of intervals) relative to other test conditions. Lower levels of problem behavior were observed in the play ($M = 7.2\%$ of intervals), alone ($M = 5.5\%$ of intervals), attention ($M = 2.8\%$ of intervals), and tangible ($M = 1.2\%$ of intervals) conditions. These outcomes suggest that his SIB was maintained by negative reinforcement in the form of escape from task demands. Five of five raters on the QABF identified escape as the primary maintaining contingency; however, the MAS did not produce agreement among four of five raters for Asa's SIB. Therefore, only the results of Asa's QABF corresponded with FA results.

Peter. Results of Peter's precursor FA are presented in Figure 3. Precursor behavior consistently occurred at higher rates in the attention condition ($M = 1.7$ responses per minute) than in other test conditions. Lower levels of problem behavior were observed in play ($M = 1.1$ responses per minute), tangible ($M = 0.69$ responses per minute), no-interaction ($M = 0.13$ responses per minute), and demand ($M = 0.06$ responses per minute) conditions. SIB occurred only once during the precursor FA, in the first presentation of the attention condition. Taken together, these

results suggest that his SIB was maintained by positive reinforcement in the form of caregiver attention. QABF results showed that four of five raters identified attention as the primary maintaining consequence; however, MAS results showed four of five raters identified a sensory function for Peter's SIB. Therefore, only the results of Peter's QABF corresponded with FA results.

GENERAL DISCUSSION

The purpose of the current study was twofold. Experiment 1 evaluated within- and across-assessment agreement of the MAS and QABF among five respondents. At least four of five respondents agreed on the function of problem behavior in 57% of QABF assessments and 52% of MAS assessments. Although this level of agreement is clearly not optimal, it may be encouraging if, when agreement is obtained, there is a high probability that the results are valid. Thus, Experiment 2 provided a preliminary evaluation of validity by examining the extent of correspondence between the anecdotal assessments and FAs. Correspondence between the QABF and FA was found for six of seven participants (86%). For the sole case in which correspondence was not observed, the outcomes of the FA were not sufficiently differentiated to determine correspondence. Correspondence between the MAS and FA was found for four

participants, or 66.6%, of cases in which it was possible to determine correspondence. Thus, the QABF showed slightly higher correspondence with analogue assessments than the MAS. Overall, these outcomes suggest that use of multiple respondents with the QABF and MAS produced valid results for approximately half of the cases assessed.

These findings are consistent with those of Paclawskyj et al. (2001), who showed that correspondence can be obtained among the QABF, MAS, and FA. Interestingly, data from one of seven (14%) FAs in the current study produced patterns of responding that did not permit a clear identification of operant function. Thus, although correspondence between the anecdotal assessments and FA could not be assessed, it is possible that Jolinda's anecdotal assessments produced more useful information about the function of her SIB than did the FA. Caregiver reports and clinical observations suggested that her SIB may have been maintained by tangible reinforcement, as was indicated by anecdotal assessment outcomes. In the absence of a clear FA or function-based treatment outcomes, the validity of her results remains unconfirmed.

Effective and efficient treatment of behavior disorders depends, in part, on the identification of maintaining variables for problem behavior. However, consumers of behavioral services, such as schools, families, clinics, or institutions, often have limited time and resources to assess the function of problem behavior. Although research outcomes have shown only limited support for the reliability and validity of anecdotal assessments, they continue to be used extensively in practice, possibly because of perceived improvements in efficiency relative to other assessment methods. The current results indicate that multiple-respondent anecdotal assessments may be useful when they are integrated into a comprehensive approach for the identification of the operant functions of problem behavior and development of function-based interventions. For example, a series

of procedures from descriptive assessment (to identify specific events that tend to occur before and after problem behavior), through multiple-respondent anecdotal assessment (to identify hypothesized contingencies of reinforcement), to brief experimental analysis (e.g., test-control treatment analysis), may be sufficient to identify (through descriptive and anecdotal assessments) and confirm (through brief experimental analysis of function-based treatment) the environmental determinants of problem behavior as well as a potential course of treatment. When the results of multiple-respondent anecdotal assessment do not show agreement, more extensive experimental analysis may be necessary. Thus, although the current results are encouraging, multiple-respondent anecdotal assessment should be used with proper recognition of its limitations, and only within a comprehensive approach to treatment that includes systematic manipulation of relevant variables and direct observation of behavior.

Some limitations of the present study are worth noting. First, contingencies for Peter's FA were placed on a precursor to the target behavior instead of the target behavior evaluated in his anecdotal assessments. Although the indirect nature of both the precursor and anecdotal assessments may limit interpretations of their outcomes, a systematic analysis of function-based treatment for Peter's SIB was conducted after this study, and the outcomes provided further evidence to support the validity of the assessment (Dracobly & Smith, 2012). Second, because Jolinda's FA was undifferentiated, it was not possible to evaluate correspondence between the FA and the anecdotal assessments. Additional procedures, such as altering FA procedures to better approximate conditions that occur in the natural environment or systematically evaluating the effects of function-based treatment, may have provided additional evidence about the operant function of her behavior. Third, both participants whose anecdotal assessments indicated a sensory reinforcement function exhib-

ited pica. It is possible that the operant function of pica is particularly discriminable; therefore, future investigations should make efforts to include participants who exhibit a more topographically diverse variety of behaviors that are thought to be maintained by automatic reinforcement. Fourth, the number of respondents, as well as the criterion for agreement among respondents (four of five), were selected somewhat arbitrarily. No data currently exist to provide an empirical basis for selecting these parameters. Future research might investigate the necessary and sufficient number of respondents and level of agreement to achieve positive outcomes. Fifth, although procedures were in place to insure that caregivers did not interact with or hear each other during administration of the assessments, it is possible that they discussed their responses outside the experimental context. Future researchers might instruct respondents not to discuss their responses with other potential respondents. Finally, treatments based on the operant functions identified by anecdotal assessments and FA were not evaluated; showing effective treatment in natural environments would lend additional external validity to the current findings.

The results of the current study, combined with those from previous investigations, suggest that multiple-respondent anecdotal assessment represents a promising approach to functional assessment. Both the MAS and QABF showed agreement among at least four of five respondents in a little over half of the cases assessed, with the QABF showing slightly higher agreement among respondents. Furthermore, for all cases in which differentiated FA results were compared with anecdotal assessments, correspondence was observed for the QABF. This information could be important to clinicians in settings where resources are limited and it is necessary to assess behavior quickly, economically, and with as little risk to participants as possible. Results from this study build on previous research that has suggested that multiple-respondent anecdotal assessment may

represent an efficient means to identify the likely operant function of problem behavior for many participants (Fahrenholz, 2004; Matson *et al.*, 1999).

REFERENCES

- Applegate, H. J. E., Matson, J. L., & Cherry, K. E. (1999). An evaluation of functional variables affecting severe problem behaviors in adults with mental retardation by using the Questions about Behavioral Function Scale (QABF). *Research in Developmental Disabilities, 20*, 229–237. doi: 10.1016/S0891-4222(99)00005-0
- Asmus, J. M., Vollmer, T. R., & Borrero, J. C. (2002). Functional behavioral assessment: A school based model. *Education and Treatment of Children, 25*, 67–90.
- Bihm, E., Kienlen, T., Ness, E., & Poindexter, A. (1991). Factor structure of the motivation assessment scale for persons with mental retardation. *Psychological Reports, 68*, 1235–1238. doi:10.2466/pr0.1991.68.3c.1235
- Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis, 1*, 175–191. doi:10.1901/jaba.1968.1-175
- Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126. doi:10.1901/jaba.1985.18-111
- Dracobly, J. D., & Smith, R. G. (2012). Progressing from identification and functional analysis of precursor behavior to treatment of self-injurious behavior. *Journal of Applied Behavior Analysis, 45*, 361–374.
- Durand, V. M., & Crimmins, D. B. (1988a). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders, 18*, 99–117.
- Durand, V. M., & Crimmins, D. B. (1988b). *The Motivation Assessment Scale (MAS) administration guide*. Topeka, KS: Monaco & Associates.
- Fahrenholz, A. R. (2004). *Multiple-respondent anecdotal assessments for behavior disorders: An analysis of interrater agreement and correspondence with functional analysis outcomes* (Master's thesis). University of North Texas, Denton.
- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*, 147–185. doi:10.1901/jaba.2003.36-147
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*, 197–209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3–20, 1982) doi:10.1901/jaba.1994.27-197

- Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1990). The experimental (functional) analysis of behavior disorders: Methodology, applications, and limitations. In A. C. Repp & N. N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 301–330). Sycamore, IL: Sycamore.
- Matson, J. L., Bamburg, J. W., Cherry, K. E., & Paclawskyj, T. R. (1999). A validity study of the Questions About Behavioral Function (QABF) Scale: Predicting treatment success for self-injury, aggression and stereotypies. *Research in Developmental Disabilities, 20*, 163–176. doi:10.1016/S0891-4222(98)00039-0
- Matson, J. L., & Vollmer, T. R. (1995). *The Questions about Behavioral Function (QABF) user's guide*. Baton Rouge, LA: Scientific Publishers.
- Neef, N. A., & Peterson, S. M. (2007). Functional behavior assessment. In J. O. Cooper, T. E. Heron, & W. L. Heward (Eds.), *Applied behavior analysis* (2nd ed., pp. 500–524). Upper Saddle River, NJ: Pearson Education.
- Newton, J. T., & Sturmey, P. (1991). The Motivation Assessment Scale: Inter-rater reliability and internal consistency in a British sample. *Journal of Mental Deficiency Research, 35*, 472–474. doi:10.1111/j.1365-2788.1991.tb00429.x
- Paclawskyj, T. R., Matson, J. L., Rush, K. S., Smalls, Y., & Vollmer, T. R. (2000). Questions about Behavioral Function (QABF): A behavioral checklist for functional assessment of aberrant behavior. *Research in Developmental Disabilities, 21*, 223–229. doi:10.1016/S0891-4222(00)00036-6
- Paclawskyj, T. R., Matson, J. L., Rush, K. S., Smalls, Y., & Vollmer, T. R. (2001). Assessment of the convergent validity of the Questions about Behavioral Function (QABF) Scale with analogue functional analysis and the Motivation Assessment Scale. *Journal of Intellectual Disability Research, 45*, 484–494. doi:10.1046/j.1365-2788.2001.00364.x
- Shogren, K. A., & Rojahn, J. (2003). Convergent reliability and validity of the Questions about Behavioral Functions and the Motivation Assessment Scale: A replication study. *Journal of Developmental and Physical Disabilities, 15*, 367–375. doi:10.1023/A:1026314316977
- Sigafoos, J., Kerr, M., & Roberts, D. (1994). Interrater reliability of the Motivation Assessment Scale: Failure to replicate with aggressive behavior. *Research in Developmental Disabilities, 15*, 333–342. doi:10.1016/0891-4222(94)90020-5
- Smith, R. G., & Churchill, R. M. (2002). Identification of environmental determinants of behavior disorders through functional analysis of precursor behaviors. *Journal of Applied Behavior Analysis, 35*, 125–136. doi:10.1901/jaba.2002.35-125
- Sturmey, P. (1995). Analog baselines: A critical review of the methodology. *Research in Developmental Disabilities, 16*, 269–284. doi:10.1016/0891-4222(95)00014-E
- Thompson, S., & Emerson, E. (1995). Inter-informant agreement on the Motivation Assessment Scale: Another failure to replicate. *Mental Handicap Research, 8*, 203–208. doi:10.1111/j.1468-3148.1995.tb00156.x
- Zarcone, J. R., Rodgers, T. A., Iwata, B. A., Rourke, D. A., & Dorsey, M. F. (1991). Reliability analysis of the Motivation Assessment Scale: A failure to replicate. *Research in Developmental Disabilities, 12*, 349–360. doi:10.1016/0891-4222(91)90031-M

Received October 14, 2011

Final acceptance July 2, 2012

Action Editor, Michael Kelley