

RELIABILITY AND VALIDITY OF THE FUNCTIONAL
ANALYSIS SCREENING TOOL

BRIAN A. IWATA

UNIVERSITY OF FLORIDA

ISER G. DELEON

KENNEDY KRIEGER INSTITUTE AND JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE

AND

EILEEN M. ROSCOE

NEW ENGLAND CENTER FOR CHILDREN

The Functional Analysis Screening Tool (FAST) is a 16-item questionnaire about antecedent and consequent events that might be correlated with the occurrence of problem behavior. Items are organized into 4 functional categories based on contingencies that maintain problem behavior. We assessed interrater reliability of the FAST with 196 problem behaviors through independent administration to pairs of raters (Study 1). Mean item-by-item agreement between pairs of raters was 71.5%. Agreement for individual items ranged from 53.3% to 84.5%. Agreement on FAST outcomes, based on comparison of informants' highest totals, was 64.8%. We assessed the validity of the FAST by comparing its outcomes with results of 69 functional analyses (Study 2). The FAST score predicted the condition of the functional analysis in which the highest rate of problem behavior occurred in 44 cases (63.8%). Potential uses of the FAST in the context of a clinical interview, as well as limitations, are discussed.

Key words: assessment, functional analysis, rating scale, problem behavior

Functional analysis (FA) methodology involves observation of behavior under a series of test and control conditions and is regarded as the benchmark standard for assessment of problem behavior in both clinical research and practice (Hanley, Iwata, & McCord, 2003). Prior to conducting an FA, therapists often gather information about the client and problem behavior by interviewing significant others, which may be helpful in designing some aspects of FA conditions. For example, several formats

have been suggested for conducting interviews with caretakers (Grodén, 1989; Iwata, Wong, Riordan, Dorsey, & Lau, 1982; O'Neill, Horner, Albin, Storey, & Sprague, 1990), which included questions about environmental circumstances that might be correlated with the occurrence of problem behavior. Verbal reports about behavior, however, often are unreliable and inaccurate. For example, it has been found that caregiver descriptions of client preferences often do not correspond with empirical (direct) assessment of those preferences (Green, Reid, Canipe, & Gardner, 1991; Green et al., 1988). Given the inability of caregivers to identify stimuli that might serve as reinforcers for *any* behavior, identification of reinforcers that maintain *specific* behaviors would seem to be an even more formidable challenge beyond the scope of an interview.

We thank SungWoo Kahng, Dorothea Lerman, Jana Lindberg, Bridget Shore, Michele Wallace, and April Worsdell, who helped with initial stages of scale development; Nancy Sanders, who coordinated pilot testing; and Kathryn Horton, Kathryn Jann, Griffin Rooker, and Kevin Schlichenmeyer, who assisted with data analysis.

Address correspondence to Brian A. Iwata, 114 Psychology Building, University of Florida, Gainesville, Florida 32611 (e-mail: iwata@ufl.edu).

doi: 10.1002/jaba.31

Nevertheless, a number of verbal report (indirect) methods have been developed to facilitate identification of contingencies that maintain problem behavior. Most consist of a series of questions or statements about circumstances under which behavior may or may not occur, to which an informant answers "yes" or "no," or indicates the extent of agreement on a Likert-type numeric scale. Although repeatedly shown to have poor reliability or validity (see reviews by Barton-Arwood, Wehby, Gunter, & Lane, 2003; Kelley, LaRue, Roane, & Gadaire, 2011; Sigafos, Kerr, Roberts, & Couzens, 1993; Sturmey, 1994), the use of questionnaires as a primary (or sometimes the only) approach to behavioral assessment appears to be widespread among clinicians and educators (Desrochers, Hile, & Williams-Mosely, 1997; Ellingson, Miltenberger, & Long, 1999; Knoster, 2000; Van Acker, Boreson, Gable, & Potterton, 2005). The continued popularity of these methods in spite of their limitations probably can be attributed to three factors. First, rating scales and questionnaires provide a consistent format for conducting an interview. Instead of having to prepare questions prior to each interview and overlooking certain details as a result, the clinician has ready access to the same set of questions. Second, the nature of the task is such that relatively little skill is required to administer the rating scale, making it ideal for use by those whose training in behavior analysis is limited. Finally, the process is extremely efficient, sometimes requiring only 15 min.

Assuming that information about functional characteristics of problem behavior inevitably will be sought during the course of an interview, we attempted to develop a questionnaire whose content was consistent with empirical research on the FA of problem behavior. Our purpose was not to produce a questionnaire that would supplant an FA but, rather, one that might be helpful in structuring a preliminary interview. We began by considering conditions from the experimental FA literature that are known to occasion and

maintain problem behavior. After generating an initially large item pool describing these conditions, we reduced and refined it through pilot testing. The resulting scale, the Functional Analysis Screening Tool (FAST), is described in this report. After we provide information on scale development, we present data from a reliability (interobserver agreement) analysis of the FAST (Study 1) and from a comparison of FAST outcomes with those obtained from FAs (Study 2).

DEVELOPMENT OF THE FAST

Functional Categories, Item Development, and Scale Revision

The FAST was designed to prompt informant verbal reports about conditions under which problem behavior might occur and to organize those reports according to common contingencies that maintain problem behavior: positive and negative reinforcement. We divided these contingencies further based on whether the source of reinforcement was social (delivered by others) or automatic (produced directly by the response), which yielded four functional categories: (a) social-positive reinforcement (access to attention or tangible items), (b) social-negative reinforcement (escape from task demands or other types of social interaction), (c) automatic-positive reinforcement (self-stimulatory behavior), and (d) automatic-negative reinforcement (alleviation of pain or discomfort).

Our selection of these specific categories was based on the following rationales. First, data from several large-scale studies (e.g., Derby *et al.*, 1992; Iwata *et al.*, 1994) indicate that most problem behavior is maintained by contingencies of social-positive, social-negative, or automatic-positive reinforcement. Second, the delivery of tangible items as consequences for problem behavior always occurs in the context of a social interaction (attention), making the distinction between these influences extremely difficult simply based on recall. Therefore, we combined

access to attention and access to tangible items under a single category (social-positive reinforcement) instead of separating them, as is done in some scales such as the Motivation Assessment Scale (MAS; Durand & Crimmins, 1988) and the Questions about Behavioral Function (QABF, Matson & Vollmer, 1995). In a similar way, we combined escape from task demands with other types of social escape and avoidance under the social-negative reinforcement category (neither the MAS nor the QABF makes a distinction between these forms of negative reinforcement). Finally, the automatic-negative reinforcement category is a tenuous one. Although alleviation of discomfort has a high degree of face validity because it seems to be a reinforcer for some behavior (e.g., scratching an itch), its direct influence on problem behavior has been largely hypothetical, based on either inferences from nonhuman work or correlational data (Cataldo & Harris, 1982) rather than experimental data with clinical populations. For example, although problem behavior may appear to be exacerbated by illness, it could be maintained purely by social consequences that become more valuable in the presence of discomfort. O'Reilly (1997) presented data that showed that an individual's SIB during attention and demand conditions of an FA was correlated with the presence of otitis media, which suggested that attention and escape from task demands were reinforcing when illness was present but not when it was absent. Nevertheless, inclusion of this function complements the other three in spite of a general absence of research on its characteristics.

Based on an examination of assessment conditions typically used in FA research on problem behavior (see Hanley et al., 2003, for a description of many of these conditions), we developed lists of events that have been shown to serve as motivating (establishing and abolishing) operations or reinforcing consequences, and in some cases, behaviors that might be members of a response class with problem behavior (e.g.,

noncompliance). The initial scale consisted of 32 questions, with eight questions for each of the four functions.

We subsequently modified the scale based on results of a series of evaluations. First, all members ($N = 9$) of a psychology department at a residential center for persons with intellectual disabilities used the initial scale during their assessments of problem behavior. After a 4-month trial period, they provided detailed written feedback about scale content and format, which we used to make format revisions and to modify wording of specific items. Second, reliability analyses were conducted in three state residential programs for persons with intellectual disabilities. The scale was administered to pairs of direct-care staff who worked closely with individuals who engaged in varied problem behaviors. Following the first administration, reliability (percentage agreement) scores were calculated separately for each of the 32 items and were used as the basis for revision. Items with the lowest reliabilities were reworded or deleted, and a revised scale was administered to another sample of informants. We revised the scale four times in this manner based on data collected for a total of 182 individuals, yielding a final scale that consists of 16 items.

Description of the Scale

Figure 1 shows the current version of the FAST, which consists of three sections. (See Supporting Information for a full-page version.) The first section contains brief instructions; it also is used to record information about the client, problem behavior, and the client-informant relationship. The second section contains 16 questions that focus on antecedent conditions under which problem behavior may or may not occur, consequences that typically follow problem behavior, or correlated behaviors. The informant responds to each question by circling yes or no to indicate that the events described do or do not occur or N/A to indicate either a lack of information or that the question is not applicable

FAST

Functional Analysis Screening Tool

Client: _____ Date: _____

Informant: _____ Interviewer: _____

To the Interviewer: The FAST identifies factors that may influence problem behaviors. Use it only for screening as part of a comprehensive functional analysis of the behavior. Administer the FAST to several individuals who interact with the client frequently. Then use the results to guide direct observation in several different situations to verify suspected behavioral functions and to identify other factors that may influence the problem behavior.

To the Informant: Complete the sections below. Then read each question carefully and answer it by circling "Yes" or "No." If you are uncertain about an answer, circle "N/A."

Informant-Client Relationship

- Indicate your relationship to the person: Parent Instructor
 Therapist/Residential Staff _____ (Other)
- How long have you known the person? Years Months
- Do you interact with the person daily? Yes No
- In what situations do you usually interact with the person?
 Meals Academic training
 Leisure Work or vocational training
 Self-care _____ (Other)

Problem Behavior Information

- Problem behavior (check and describe):
 Aggression _____
 Self-Injury _____
 Stereotypy _____
 Property destruction _____
 Other _____
- Frequency: Hourly Daily Weekly Less often
- Severity: Mild: Disruptive but little risk to property or health
 Moderate: Property damage or minor injury
 Severe: Significant threat to health or safety
- Situations in which the problem behavior is most likely to occur:
Days/Times _____
Settings/Activities _____
Persons present _____
- Situations in which the problem behavior is least likely to occur:
Days/Times _____
Settings/Activities _____
Persons present _____
- What is usually happening to the person right before the problem behavior occurs?

- What usually happens to the person right after the problem behavior occurs?

- Current treatments _____

1. Does the problem behavior occur when the person is not receiving attention or when caregivers are paying attention to someone else?	Yes	No	N/A	
2. Does the problem behavior occur when the person's requests for preferred items or activities are denied or when these are taken away?	Yes	No	N/A	
3. When the problem behavior occurs, do caregivers usually try to calm the person down or involve the person in preferred activities?	Yes	No	N/A	
4. Is the person usually well behaved when (s)he is getting lots of attention or when preferred activities are freely available?	Yes	No	N/A	
5. Does the person usually fuss or resist when (s)he is asked to perform a task or to participate in activities?	Yes	No	N/A	
6. Does the problem behavior occur when the person is asked to perform a task or to participate in activities?	Yes	No	N/A	
7. If the problem behavior occurs while tasks are being presented, is the person usually given a "break" from tasks?	Yes	No	N/A	
8. Is the person usually well behaved when (s)he is not required to do anything?	Yes	No	N/A	
9. Does the problem behavior occur even when no one is nearby or watching?	Yes	No	N/A	
10. Does the person engage in the problem behavior even when leisure activities are available?	Yes	No	N/A	
11. Does the problem behavior appear to be a form of "self-stimulation?"	Yes	No	N/A	
12. Is the problem behavior <u>less</u> likely to occur when sensory stimulating activities are presented?	Yes	No	N/A	
13. Is the problem behavior cyclical, occurring for several days and then stopping?	Yes	No	N/A	
14. Does the person have recurring painful conditions such as ear infections or allergies? If so, list: _____	Yes	No	N/A	
15. Is the problem behavior <u>more</u> likely to occur when the person is ill?	Yes	No	N/A	
16. If the person is experiencing physical problems, and these are treated, does the problem behavior usually go away?	Yes	No	N/A	

Scoring Summary

Circle the number of each question that was answered "Yes" and enter the number of items that were circled in the "Total" column.

Items Circled "Yes"	Total	Potential Source of Reinforcement
1 2 3 4 _____		Social (attention/preferred items)
5 6 7 8 _____		Social (escape from tasks/activities)
9 10 11 12 _____		Automatic (sensory stimulation)
13 14 15 16 _____		Automatic (pain attenuation)

Figure 1. Functional Analysis Screening Tool.

to the client's problem. Items 1 through 4 focus on problem behavior maintained by social-positive reinforcement, which has been shown to occur when access to attention (Item 1) or preferred items (Item 2) is restricted and when these events are delivered as consequences (Item 3), but not usually when free access to positive reinforcement is available (Item 4). Items 5 through 8 focus on problem behavior maintained by social-negative reinforcement, which often is accompanied by noncompliance (Item 5) and is likely to occur in the presence of task or social demands (Item 6) when escape is available (Item 7), but not when demands are absent (Item 8). Items 9 through 12 focus on problem behavior maintained by automatic-positive reinforcement, whose occurrence is unrelated to social interaction (Item 9), may or may not be influenced by the availability of stimulating activities (Items 10 and 12), and usually is maintained by sensory stimulation (Item 11). Items 13 through 16 focus on problem behavior maintained by automatic-negative reinforcement. Pain-attenuating problem behavior, to the extent that it occurs, seems to be cyclical (Item 13), to occur more often in individuals who experience recurring medical problems (Item 14), and when these problems are present (Item 15) rather than absent (Item 16). The third section of the FAST consists of a scoring summary. Items from the second section are grouped according to behavioral function, and results are summarized by circling numbers corresponding to questions for which a "yes" answer was given.

We selected a yes–no format for answers rather than a numerical scale for two reasons. First, we conducted an informal analysis of several rating scales (including a preliminary version of the FAST) and found that the factor that accounted for the largest proportion of variability on individual item agreement was the number of response options: Reliability was inversely correlated with the number of choices. Second, labels that define distinctions among numerical anchors seemed fairly arbitrary, as in Anchor 1 *almost*

never versus Anchor 2 *seldom*, or Anchor 4 *usually* versus Anchor 5 *almost always* on the MAS.

STUDY 1: RELIABILITY ANALYSIS

Method

Subjects, setting, and administration procedure. Data were collected for 151 individuals drawn from three client populations (University of Florida, Kennedy Krieger Institute, and New England Center for Children), all of whom had been diagnosed with an intellectual disability or autism and had been referred for assessment of problem behavior. One hundred individuals were male and 51 were female; their ages ranged from 5 to 53 years ($M = 17.8$ years). Some individuals exhibited more than one behavior problem (see Table 1 for a complete listing), resulting in a total sample of 196 behaviors for which a FAST was completed.

Informants consisted of parents, relatives, teachers, teacher aides, and direct-care staff who were responsible for the care or training of clients. Although this sample was heterogeneous and uncontrolled, it was representative of the population from whom information about clients' problem behavior would be sought. The educational background of informants who were training staff ranged from high school diploma to master's degree but was unknown for family members. Two informants independently

Table 1
Problem Behaviors for Which FAST and FA Data Were Collected

Problem behavior	Study 1 Pairs of FASTs	Study 2 FAs
Aggression	51	21
Elopement	7	0
Inappropriate verbal	21	3
Noncompliance	8	1
Property destruction	24	6
Self-injury	66	31
Stereotypy	17	7
Other (defecation, theft)	2	0
Total	196	69

completed each FAST either on the same day or within 2 to 3 days of each other. Informants were given as much time as needed to complete the questionnaire (typical completion time was 15 to 20 min).

Reliability analysis. Interrater reliability (agreement) between pairs of informants was calculated in three ways. First, the two FASTs for each problem behavior were compared on an item-by-item basis, and a percentage agreement score was calculated by dividing the number of agreements on yes or no answers by 16 (the total number of items). This calculation yielded an overall agreement score for each FAST. (Any item for which either informant answered N/A was excluded from this and all subsequent calculations.) Second, the number of agreements for each separate item was divided by the number of pairs of FASTs, which yielded an agreement score for each of the 16 items. Finally, reliability was assessed for FAST outcomes (the function receiving the most yes responses). The number of agreements on the most frequent yes total for a given function (maximum possible was four) was divided by the number of pairs of FASTs. If one informant's FAST responses produced a tie (e.g., if an informant gave four yes responses to each of two functions), an agreement was scored if either matched the other informant's highest total.

Results and Discussion

Mean overall agreement for the FAST, based on item-by-item comparisons between pairs of informants, was 71.5% (range, 28.6% to 100%). Using the 80% criterion typically considered acceptable for direct-observation measures, reliability of the FAST is moderate at best. Figure 2 shows a frequency distribution of percentage agreement scores for the 196 pairs of FASTs. The majority of agreement scores ($n = 92$) fell between 61% and 80%.

Interpretation of these results is difficult because an acceptable level of agreement has not been established for behavioral rating scales. Moreover, reliability for most published rating

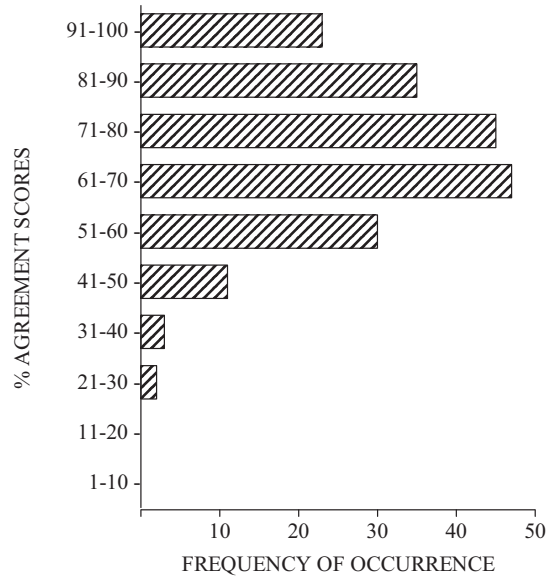


Figure 2. Frequency distribution of percentage agreement scores based on item-by-item comparisons between 196 pairs of informants (Study 1).

scales either has not been reported at all or has been examined via correlational analysis, which does not establish the extent of agreement between any pair of raters and therefore is irrelevant at the level of individual analysis. The most precise measure of reliability is percentage agreement based on item-by-item comparison of informants' scores (see above). Table 2 summarizes percentage agreement data that have been reported (as well as not reported) for published rating scales as a point of comparison. Because numeric Likert ratings used in some scales offer a wider range of response options than yes-no answers, they may yield lower point-by-point agreement. Therefore, we have listed, when available, reliability percentages based on adjacent Likert scores, in which an agreement is scored if one informant's numeric rating falls within ± 1 of the other's. Thus, if the Likert values on a scale comprise a 6-point range, an agreement is a moving range that encompasses

Table 2
Interrater Reliability of Rating Scales Based on Item-by-Item Agreement

Study	<i>N</i>	Agreement
Motivation Assessment Scale (Durand & Crimmins, 1988)		
Conroy et al. (1996)	14	Mean ^a = 56% (range, 25% to 78%)
Duker and Sigafoos (1998)	90	Median ^a = 63.3% (range, 54.4% to 71.1%)
Sigafoos et al. (1994)	18	Mean ^a = 41% (range, 18.8% to 62.5%)
Zarccone et al. (1991)	55	Mean ^a = 48% (range, 0% to 88%)
Questions about Behavioral Function (Matson & Vollmer, 1995)		
Nicholson et al. (2006)	118	Median ^a = 78% (range, 69.5% to 84.8%)
Structured Interview Protocol (Sigafoos et al., 1993)		
Sigafoos et al. (1993)	18	Mean ^b = 43.3% (range, 11.1% to 83.3%)
Data unavailable for:		
Behavior Analytic Questionnaire (Hauck, 1985)		
Contextual Assessment Inventory (McAtee et al., 2004)		
Functional Assessment for Multiple Causality (Matson et al., 2003)		
Functional Assessment Checklist for Teachers and Staff (March et al., 2000)		
GB Motivating Screening Tool (Barrera & Graver, 2009)		
Motivation Analysis Rating Scale (Wieseler et al., 1985)		
Problem Behavior Questionnaire (Lewis et al., 1994)		

^aAdjacent numeric agreement.

^bYes–no agreement.

half the values (except when two informants' scores are at the extreme opposite values), yielding 50% chance agreement that is equivalent to chance agreement for yes–no responses. Relative to percentage agreement scores reported for other scales, the FAST yielded comparable reliability.

Table 3 shows percentage agreement scores for each of the 16 items included in the FAST. Agreement on individual items varied widely: Item 12 (“Is the problem behavior less likely to occur when sensory stimulating activities are available?”) yielded the lowest agreement (53.3%), whereas Item 4 (“Is the person usually well behaved when he or she is getting lots of

attention or when preferred activities are freely available?”) yielded the highest agreement (84.5%). Agreement was not noticeably different for questions within a given functional category; however, answers to questions about antecedent events usually were more reliable ($M = 78.9\%$) than were answers to questions about consequent events ($M = 67.7\%$). This finding was unusual given that the antecedent event that precedes problem behavior often has passed by the time behavior occurs, which is one limitation of event-based descriptive analyses. Perhaps the higher reliability for antecedent events reflected recollection of repeated episodes of problem behavior in an ongoing antecedent context.

Table 4 summarizes results obtained for agreement on FAST outcomes, that is, the extent to which two informants' most frequent yes answers coincided for the same function. Whereas item-by-item agreement reflects the extent to which informants provide the same answers to the same questions, outcome agreement reflects the extent to which informants, regardless of their answers to individual questions, provide yes answers to more questions in the same functional category. Using direct-

Table 3
Percentage Agreement Scores for Individual FAST Items

Item	Agreement	Item	Agreement
1	70.8	9	77.4
2	78.6	10	71.1
3	66.5	11	69.1
4	84.5	12	53.3
5	70.6	13	70.2
6	67.9	14	80.3
7	67.5	15	69.9
8	75.9	16	75.0

Table 4
Summary of Agreements on FAST Outcomes

Type of agreement	Occurrences	Percentage
Agreement, single function	51/76	67.1
Agreement, multiple functions	76/120	63.3
Total agreements	127/196	64.8

observation data as an example, item-by-item agreement is analogous to interval agreement, whereas outcome agreement is more similar to total agreement. This latter type of reliability is not considered acceptable for data gathered through direct observation; nevertheless, it may represent the way in which rating-scale data are aggregated for clinical decision making. As can be seen from the table, outcome agreement for the FAST (64.8%) was actually somewhat lower than item-by-item agreement (71.5%).

STUDY 2: VALIDITY ANALYSIS

Method

Subjects and setting. We conducted FAs for 59 individuals, a subset of those for whom FAST data were collected in Study 1. Some individuals participated in more than one FA for different target behaviors, yielding a total of 69 FAs. Sessions were conducted either in individual therapy rooms or in isolated areas of classrooms. Sessions lasted 10 min, were conducted several times daily (based on scheduling constraints), and were completed within 1 day to 1 week for each individual.

Functional Analysis Procedure

Response measurement and reliability. Target behaviors consisted of those listed in Table 1 and were defined operationally on an individual basis. Trained observers collected data on laptop or handheld computers or on preprinted data sheets. Data were summarized as either rate (responses per minute) or the percentage of continuous 10-s intervals during which behavior occurred. An independent observer recorded data during a mean of 43.1% of sessions (range across subjects,

13.3% to 100%). Interobserver reliability was assessed by dividing session time into consecutive 10-s intervals and comparing observer's records on an interval-by-interval basis. Agreement for rate measures was calculated by dividing the smaller number of responses recorded in each interval by the larger and averaging these fractions across the session. Agreement for interval measures was calculated by dividing the number of intervals with scoring agreements (on the occurrence or nonoccurrence of behavior) by the total number of intervals in the session. Mean agreement across subjects was 96.4% (range, 85.7% to 100%).

FA conditions. Subjects were exposed to four assessment conditions (alone, attention, play, and demand) in a multielement design based on procedures described by Iwata, Dorsey, Slifer, Bauman, and Richman (1982/1994). An ignore condition (a therapist was present but did not interact with the client throughout the session) was substituted for the alone condition if the target behavior was aggression. Therapists consisted of behavior analysts or graduate students assigned to the individual cases, all of whom had extensive experience conducting FAs. Assessment continued until the on-site assessment team determined that the functions of an individual's problem behavior had been identified.

Data interpretation. After completion of all FAs, individual graphs were prepared without identifying information and were shown to a team of five behavior analysts who were highly experienced in conducting FAs and who were blind to the FAST outcomes. The group examined each graph and reached a consensus about the function of problem behavior. These designations were used as the basis for FAST-FA comparisons.

FAST-FA comparison. Under ideal conditions, validity of an instrument such as the FAST would be assessed by determination of the extent to which the FAST outcome (the function category with the highest total) corresponded to the FA condition with the highest response rates.

However, because only moderate agreement between pairs of informants was obtained in Study 1, arbitrary designation of one FAST as the primary score for comparison with FA data presented a problem: The function identified by one FAST might correspond to the function identified by the FA, whereas the function identified by the other FAST might not. As a result, we compared each pair of FASTs to its respective FA such that each FAST–FA comparison could yield complete (both respondents), partial (one respondent), or no agreement and used the corresponding values of 1, .5, and 0 when calculating agreement. Thus, correspondence between one FAST and FA outcome was partially canceled out if correspondence was not obtained between the other FAST and FA outcome. As was the case in calculating reliability for FAST outcomes, if an informant's FAST responses produced a tie for the highest function, an agreement was scored if either matched the outcome of the FA. We conducted a second analysis of the data by comparing FAST and FA outcomes when both informants agreed on the FAST outcome. This analysis answered the question, Given that two informants' FASTs identified the same function for problem behavior, did that function match the FA outcome?

Results and Discussion

Figure 3 shows representative samples of data illustrating correspondence (left) and noncorrespondence (right) between FAST and FA outcomes for each of three functions: social-positive, social-negative, and automatic-positive reinforcement. The automatic-negative reinforcement function does not appear because a typical FA would not yield such an outcome. Each set of FA data showed very clear results, whereas the FAST outcomes were not necessarily as clear, even when they showed correspondence with the FA data. For example, both of Sean's informants gave more yes answers to questions that reflected maintenance by social-positive reinforcement (corresponding to his FA results). However, both

informants also gave only one fewer yes answer for a different function (Informant 1: social-negative reinforcement; Informant 2: automatic-positive reinforcement). Similarly, both of Kim's informants gave more yes answers to questions that reflected maintenance by social-negative reinforcement (corresponding to her FA results). Kim's Informant 1, however, gave only one fewer yes answer for each of two additional functions (social- and automatic-positive reinforcement). Warren's case was one of the few examples in which FAST responses not only corresponded with FA results but also showed a high degree of differentiation among FAST responses. Thus, in most cases of correspondence between FAST and FA outcomes, informant responses showed little differentiation across categories; that is, conclusions about function based on FAST data were determined by a one-question difference.

Sarah's Albert's, and Heather's results all showed the absence of correspondence between FAST and FA outcomes. As was typical in cases of noncorrespondence, one (Sarah) or both (Albert and Heather) of the informants failed to identify the function of problem behavior; in addition, informants again (all three cases) showed little differentiation in responding to questions across different categories. It also is interesting to note that agreement on function between FAST informants did not improve validity. For example, FAST informants agreed on the function of both Albert's and Heather's problem behaviors; this consensus, however, did not correspond with their FA outcomes.

Table 5 summarizes results of all of the FAST–FA comparisons. There were no cases in which the FA data indicated that problem behavior was multiply controlled (maintained by more than one source of reinforcement); we attributed this to the fact that we did not aggregate problem behaviors during assessment (Beavers & Iwata, 2011). Overall correspondence between FAST and FA outcomes was 63.8%, and the highest degree of correspondence was obtained when results of the FA indicated that problem

FAST-FA: Correspondence

FAST-FA: Non-correspondence

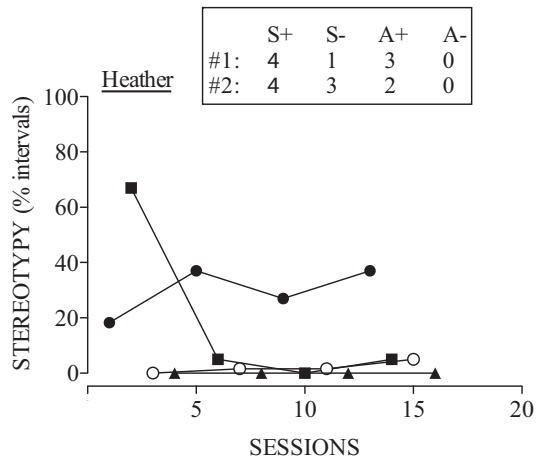
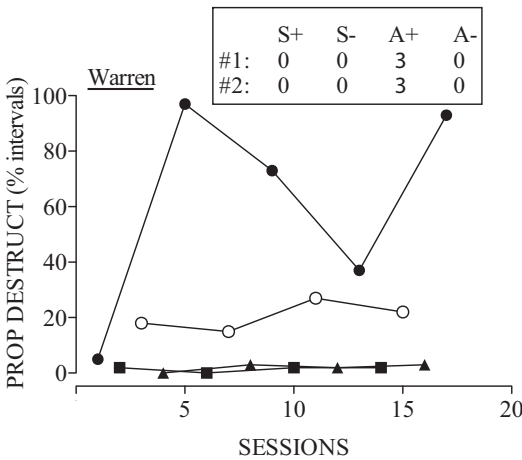
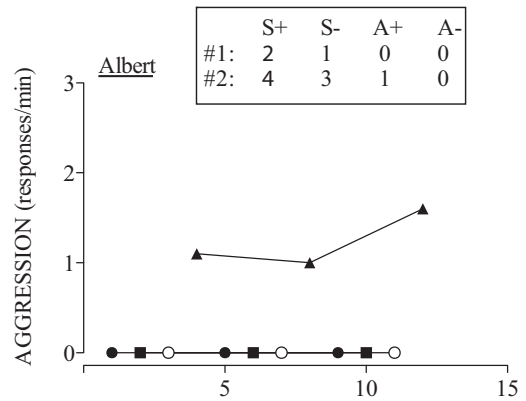
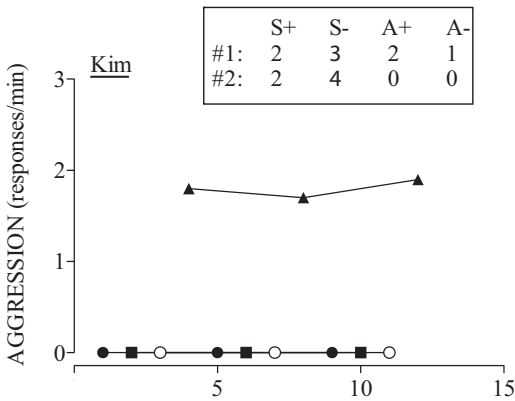
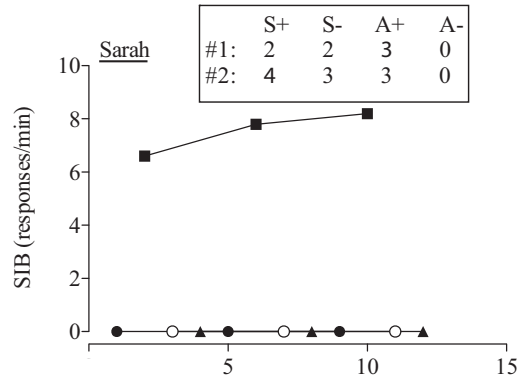
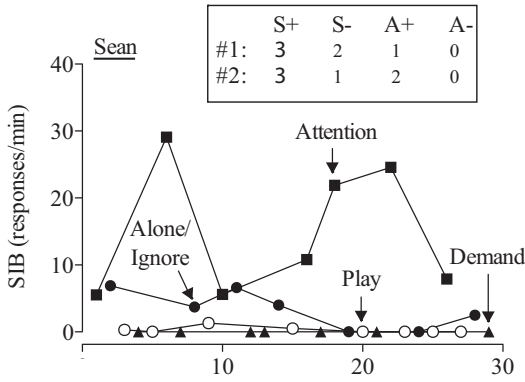


Figure 3. Representative data from 69 FAST-FA comparisons (Study 2). Left and right columns show correspondence and noncorrespondence, respectively, between informants' FAST ratings and FA results. Numbers in boxes indicate informants' yes responses to items in each functional category: social-positive reinforcement (S+), social-negative reinforcement (S-), automatic-positive reinforcement (A+), and automatic-negative reinforcement (A-). Each informant's highest score is shown in boldface.

Table 5
Summary of Agreements Between FAST and FA Outcomes

FA outcome	Cases	Matches to FAST	Percentage
Comparison based on all pairs of FASTs			
Social-positive reinforcement	18	14	77.8
Social-negative reinforcement	25	14	56
Automatic-positive reinforcement	26	16	61.5
Total	69	44	63.8
Comparison based on FASTs for which there was agreement on function			
Social-positive reinforcement	7	7	100
Social-negative reinforcement	11	6	54.6
Automatic-positive reinforcement	6	4	66.7
Total	24	17	70.8

behavior was maintained by social-positive reinforcement. When FAST-FA comparisons were limited to those for which both informants agreed on the FAST function, overall correspondence increased somewhat (70.8%), which was attributable almost entirely to the fact that there was 100% correspondence between FAST-FA outcomes for the social-positive reinforcement category.

The general lack of predictive validity of the FAST was not surprising, given its moderate reliability. Although it is possible that a higher hit rate would have been obtained by comparing only one FAST to each FA, there was no way to determine which of the two available FASTs should have served as the comparison. Given the moderate reliability obtained in Study 1 when typical informants completed the FAST, perhaps the only way to establish validity would be to have the FAST completed by an expert behavior analyst who also happened to know the client well. Of course, such an exercise would not resemble typical conditions of indirect assessment, in which informants may have little or no formal training in behavior analysis. Another possibility would involve administration of the FAST to several respondents and the use of some type of rule for aggregating responses. However, given that the FAST was not found to be highly reliable, the use of averaging may simply have the effect of mixing reliable and unreliable data to an unknown degree.

GENERAL DISCUSSION

We used content analysis from the experimental literature, extensive field testing, and repeated item revision through a series of preliminary reliability analyses to develop a rating scale for gathering information about functional characteristics of problem behavior. Based on the data reported here, our goals (accommodation of multiple behavioral functions, adequate reliability and validity, and ease of administration) were approximated. That is, although the FAST's reliability and validity compared well with what has been reported for other rating scales, it cannot be considered adequate for the purposes of treatment development. Some authors have suggested that behavioral rating scales are viable alternatives to (Durand & Crimmins, 1988) or even improvements over (Matson, Bamberg, Cherry, & Paclawskyj, 1999) FAs; however, a great deal of data (see previous reviews), including those from the present study, indicate otherwise.

Our failure to obtain high reliability and validity could hardly be considered an unexpected result, given the nature of the data generated by the FAST. An informant's task is exceedingly difficult and prone to a number of errors and biases. This problem may be overlooked by interviewers because, by comparison, their task is simple: Either give the rating scale to the informant or read it to the informant and mark

answers. In addition, a checklist can be completed in a matter of minutes, creating the illusion that the task is easily accomplished. However, accurate answers require at least the following: (a) having observed most of the events in question, (b) organizing disparate events into categories contained on the rating scale, (c) estimating conditional probabilities of events to determine their relative influence on behavior, and (d) recalling all of these details while completing the checklist. Thus, accurate completion of checklists, such as those currently being used to identify behavioral function, requires skills that far exceed those of an individual who collects direct-observation data; he or she is required only to detect the occurrence of an ongoing event (rather than many that have occurred in the past), discriminate it from others, and mark it while it is being observed. Moreover, it has been shown that correlations that result even from highly reliable and accurate observational (descriptive) data typically are not good indicators of behavioral function (Thompson & Borrero, 2011). As a result, it is unclear how unsystematic recollection of numerous past events is likely to be any better.

What, then, might be the value of using instruments such as the FAST during the course of assessment? First, to the extent that structured questionnaires provide a uniform format for gathering information, reduce the amount of time needed to prepare or conduct an interview, and have known reliability and validity (even though both may be low), they might improve the consistency and efficiency of clinical interviews. Second, information obtained from questionnaires might serve as the basis for follow-up interviews and observations, which might facilitate identification of idiosyncratic antecedent or consequent events to be included in an FA. For example, the first section of the FAST contains a series of open-ended questions that we did not include in any analysis as part of this study. Perhaps answers to those questions could be compared with answers to the FAST questions as

a check for inconsistencies and as a basis for clarifying questions or observations. Finally, when multiple informants strongly concur that an individual's problem behavior has a specific function, verification by way of an FA might be more efficient through elimination of unnecessary test conditions. For example, we found (Study 2) that when two informants' FAST scores agreed that an individual's problem behavior was maintained by social-positive reinforcement, the results were confirmed in every subsequent FA. Smith, Smith, Dracobly, and Pace (2012) also reported good correspondence between rating scale and FA outcomes when there was consensus among four of five informants on the MAS or QABF. By extension, if several knowledgeable informants agree that an individual's problem behavior is maintained by attention, then perhaps a single function test (Iwata & Dozier, 2008) that consists of contingent attention (test) and noncontingent attention (control) conditions might suffice to verify the function of problem behavior. Whether such a process actually decreases the amount of time required to conduct an FA is unknown because the benefit of efficiency (fewer FA conditions) must be weighed against the cost of inaccuracy (missing a function because it was not included in the FA).

Future research on function-based rating scales also might examine their adaptation for particular response topographies. For example, scales such as the MAS, QABF, and FAST encompass multiple functions and thus seem applicable to problem behavior that might be maintained by varied sources of reinforcement. However, results from several studies suggest that stereotypy is most likely to be maintained by automatic reinforcement (Querin *et al.*, 2013), whereas aggression is most likely to be maintained by social reinforcement (Marcus, Vollmer, Swanson, Roane, & Ringdahl, 2001). Thus, structured interviews that focus on these target behaviors might include additional items related to specific functions while items related to unlikely functions are eliminated.

A distinct danger in formalizing the role of verbal report in the assessment of behavior is its adoption for reasons other than those intended regardless of what the data show because verbal reports are easily obtained. Therefore, we reemphasize the fact that the FAST is not an approximation to an FA of problem behavior; it is simply one way to gather information during an interview. Nevertheless, because the clinical interview is a common context for posing all sorts of questions about behavior, efforts to improve its consistency may be helpful.

REFERENCES

- Barrera, F. J., & Graver, E. E. (2009). A comparison of behavioral functions in community and facility settings. *Journal on Developmental Disabilities, 15*, 30–34.
- Barton-Arwood, S. M., Wehby, J. H., Gunter, P. L., & Lane, K. L. (2003). Functional behavior assessment rating scales: Intrarater reliability with students with emotional or behavioral disorders. *Behavioral Disorders, 28*, 386–400.
- Beavers, G. A., & Iwata, B. A. (2011). Prevalence of multiply controlled problem behavior. *Journal of Applied Behavior Analysis, 44*, 593–597. doi: 10.1901/jaba.2011.44-593
- Cataldo, M. E., & Harris, J. (1982). The biological basis for self-injury in the mentally retarded. *Analysis and Intervention in Developmental Disabilities, 2*, 21–39. doi: 10.1016/0270-4684(82)90004-0
- Conroy, M. A., Fox, J. J., Bucklin, A., & Good, W. (1996). An analysis of the reliability and stability of the Motivation Assessment Scale in assessing the challenging behaviors of persons with developmental disabilities. *Education & Training in Mental Retardation and Developmental Disabilities, 31*, 243–250.
- Derby, K. M., Wacker, D. P., Sasso, G., Steege, M., Northup, J., Cigrand, K., & Asmus, J. (1992). Brief functional assessment techniques to evaluate aberrant behavior in an outpatient setting: A summary of 79 cases. *Journal of Applied Behavior Analysis, 25*, 713–721. doi: 10.1901/jaba.1992.25-713
- Desrochers, M. N., Hile, M. G., & Williams-Mosely, T. L. (1997). Survey of functional assessment procedures used with individuals who display mental retardation and severe problem behaviors. *American Journal on Mental Retardation, 101*, 535–546.
- Duker, P. C., & Sigafoos, J. (1998). The Motivation Assessment Scale: Reliability and construct validity across three topographies of behavior. *Research in Developmental Disabilities, 19*, 131–141. doi: 10.1016/S0891-4222(97)00047-4
- Durand, V. M., & Crimmins, D. B. (1988). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders, 18*, 99–117. doi: 10.1007/BF02211821
- Ellingson, S. A., Miltenberger, R. G., & Long, E. S. (1999). A survey of the use of functional assessment procedures in agencies serving individuals with developmental disabilities. *Behavioral Interventions, 14*, 187–198. doi: 10.1002/(SICI)1099-078X(199910/12)14:4<187:AID-BIN38>3.0.CO;2-A
- Green, C. W., Reid, D. H., Canipe, V. S., & Gardner, S. (1991). A comprehensive evaluation of reinforcer identification processes for persons with profound multiple handicaps. *Journal of Applied Behavior Analysis, 24*, 537–552. doi: 10.1901/jaba.1991.24-537
- Green, C. W., Reid, D. H., White, L. K., Halford, R. C., Brittain, D. P., & Gardner, S. M. (1988). Identifying reinforcers for persons with profound handicaps: Staff opinion versus systematic assessment of preferences. *Journal of Applied Behavior Analysis, 21*, 31–43. doi: 10.1901/jaba.1988.21-31
- Groden, G. (1989). A guide for conducting a comprehensive behavioral analysis of a target behavior. *Journal of Behavior Therapy and Experimental Psychiatry, 20*, 163–169. doi: 10.1016/0005-7916(89)90049-9
- 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
- Hauck, F. (1985). Development of a behavior-analytic questionnaire precising four functions of self-injurious behavior in the mentally retarded. *International Journal of Rehabilitation Research, 8*, 350–352. doi: 10.1097/00004356-198509000-00014
- 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. doi: 10.1901/jaba.1994.27-197 (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3–20, 1982)
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice, 1*, 3–9.
- Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., ... Willis, K. D. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis, 27*, 215–240. doi: 10.1901/jaba.1994.27-215
- Iwata, B. A., Wong, S. E., Riordan, M. M., Dorsey, M. F., & Lau, M. M. (1982). Assessment and training of clinical interviewing skills: Analogue analysis and field replication. *Journal of Applied Behavior Analysis, 15*, 191–203. doi: 10.1901/jaba.1982.15-191
- Kelley, M. E., LaRue, R., Roane, H. S., & Gadaire, D. M. (2011). Indirect behavioral assessments: Interviews and rating scales. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 182–190). New York, NY: Guilford.

- Knoster, T. P. (2000). Practical application of functional behavioral assessment in schools. *Journal of the Association for Persons with Severe Handicaps*, *25*, 201–211. doi: 10.2511/rpsd.25.4.201
- Lewis, T. J., Scott, T. M., & Sugai, G. (2004). The problem behavior questionnaire: A teacher-based instrument to develop functional hypotheses of problem behavior in general education classrooms. *Assessment for Effective Intervention*, *19*, 103–115. doi: 10.1177/073724779401900207
- March, R. E., Horner, R. H., Lewis-Palmer, T., Brown, D., Crone, D., Todd, A. W., & Carr, E. (2000). *Functional Assessment Checklist: Teachers and Staff (FACTS)*. Eugene, OR: Educational and Community Supports.
- Marcus, B. A., Vollmer, T. R., Swanson, V., Roane, H. S., & Ringdahl, J. E. (2001). An experimental analysis of aggression. *Behavior Modification*, *25*, 189–213. doi: 10.1177/0145445501252002
- Matson, J. L., Bamberg, J. W., Cherry, K. E., & Paclawskyj, T. R. (1999). Validity study on the Questions About Behavioral Function (QABF): Predicting treatment success for self-injury, aggression, and stereotypes. *Research in Developmental Disabilities*, *24*, 163–176. doi: 10.1016/S0891-4222(98)00039-0
- Matson, J. L., Kuhn, D. E., Dixon, D. R., Mayville, S. B., Laud, R. B., Cooper, C. L., ... Matson, M. L. (2003). The development and factor structure of the Functional Assessment for Multiple Causality (FACT). *Research in Developmental Disabilities*, *24*, 485–495. doi: 10.1016/j.ridd.2003.07.001
- Matson, J. L., & Vollmer, T. R. (1995). *User's guide: Questions About Behavioral Function (QABF)*. Baton Rouge, LA: Scientific Publishers.
- McAtee, M., Carr, E. G., & Schulte, C. (2004). A contextual assessment inventory for problem behavior: Initial development. *Journal of Positive Behavior Interventions*, *6*, 148–165. doi: 10.1177/10983007040060030301
- Nicholson, J., Konstantinidi, E., & Furniss, F. (2006). On some psychometric properties of the Questions about Behavioral Function (QABF) scale. *Research in Developmental Disabilities*, *27*, 337–352. doi: 10.1016/j.ridd.2005.04.001
- O'Neill, R. E., Horner, R. E., Albin, R. W., Storey, K., & Sprague, J. R. (1990). *Functional analysis of problem behavior: A practical assessment guide*. Sycamore, IL: Sycamore.
- O'Reilly, M. F. (1997). Functional analysis of episodic self-injury correlated with recurrent otitis media. *Journal of Applied Behavior Analysis*, *30*, 165–167. doi: 10.1901/jaba.1997.30-165
- Querim, A. C., Iwata, B. A., Roscoe, E. M., Schlichenmeyer, K. J., Virues Ortega, J., & Hurl, K. E. (2013). Functional analysis screening for problem behavior maintained by automatic reinforcement. *Journal of Applied Behavior Analysis*, *46*, 47–60.
- 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
- Sigafoos, J., Kerr, M., Roberts, D., & Couzens, D. (1993). Reliability of structured interviews for the assessment of challenging behavior. *Behaviour Change*, *10*, 47–50.
- Smith, C. M., Smith, R. G., Dracobly, J. D., & Pace, A. P. (2012). Multiple-respondent anecdotal assessments: An analysis of interrater agreement and correspondence with analogue assessment outcomes. *Journal of Applied Behavior Analysis*, *45*, 779–795. doi: 10.1901/jaba.2012.45-779
- Sturme, P. (1994). Assessing the functions of aberrant behaviors: A review of psychometric instruments. *Journal of Autism and Developmental Disorders*, *24*, 293–304. doi: 10.1007/BF02172228
- Thompson, R. H., & Borrero, J. C. (2011). Direct observation. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 191–205). New York, NY: Guilford.
- Van Acker, R., Boreson, L., Gable, R. A., & Potterton, T. (2005). Are we on the right course? Lessons learned about current FBA/BIP practices in schools. *Journal of Behavioral Education*, *14*, 35–56. doi: 10.1007/s10864-005-0960-5
- Wieseler, N. A., Hanson, R. H., Chamberlain, T. P., & Thompson, T. (1985). Functional taxonomy of stereotypic and self-injurious behavior. *Mental Retardation*, *23*, 230–234.
- Zarcone, J. R., Rodgers, T. A., Iwata, B. A., Rourke, D., & 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 September 24, 2012

Final acceptance November 26, 2012

Action Editor, Richard Smith

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1. Functional Analysis Screening Tool (FAST) Full Page Version