On the validity of interpreting functional analyses of inappropriate mealtime behavior using structured criteria

Lisa A. Guerrero

University of Nebraska Medical Center, Munroe Meyer Institute and Rutgers University, Rutgers Biomedical and Health Sciences

Christopher W. Engler

Children's Specialized Hospital

Bethany A. Hansen

University of Nebraska Medical Center, Munroe Meyer Institute

Cathleen C. Piazza

Children's Specialized Hospital and Rutgers University, Rutgers Biomedical and Health Sciences

Visual inspection is the traditional method behavior analysts use to interpret functional-analysis results. Limitations of visual inspection include lack of standardized rules, subjectivity, and inconsistent interrater reliability (Fisch, 1998). To address these limitations, researchers have developed, evaluated, and refined structured criteria to aid interpretation of functional analyses of destructive behavior (Hagopian et al., 1997; Roane et al., 2013; Saini et al., 2018). The current study applied the structured criteria Saini et al. (2018) described to functional analyses of inappropriate mealtime behavior. We assessed its predictive validity and evaluated its efficiency relative to 3 post hoc visual inspection procedures. Validity metrics were lower than those in Saini et al. however, ongoing visual inspection increased the efficiency of functional analyses by more than 30%. We discuss these findings relative to the procedural differences between functional analyses of destructive behavior and inappropriate mealtime behavior.

Key words: avoidant/restrictive food intake disorder (ARFID), functional analysis, inappropriate mealtime behavior

---

Although some have questioned the value of functional analysis (FA) in the treatment of inappropriate mealtime behavior (IMTB) associated with a diagnosis of avoidant/restrictive food intake disorder (ARFID; Hanley, 2012), studies by Kirkwood et al. (2021) and Bachmeyer et al. (2009) illustrate why identifying the reinforcer(s) for IMTB may be important. Kirkwood et al. reviewed FAs of IMTB that identified escape,

but not attention, as a reinforcer. By contrast, Bachmeyer et al. reviewed FAs of IMTB that had identified both escape and attention. Escape extinction alone reduced IMTB to clinically acceptable rates for participants in Kirkwood et al. but not for those in Bachmeyer et al. Both escape extinction and attention extinction were necessary to decrease IMTB to clinically acceptable rates in Bachmeyer et al. Thus, the researchers used FA results to develop treatments that included contingencies necessary to reduce IMTB (Bachmeyer et al.) and did not include contingencies that had no effect on IMTB (Kirkwood et al.).

Despite the potential benefits of FA in treatment development, many behavior analysts do

not conduct FAs due to lack of time and resources (Oliver et al., 2015). Timely treatment may be particularly important for a child with ARFID as the consequences of the disorder can be serious and even potentially life threatening. Furthermore, children with ARFID who do not receive quick access to treatment might require enteral feedings, which requires surgery, and opens the door to several other challenges (i.e., dependency on tube feedings and limited exposure to oral feedings; Phalen, 2013). Identifying processes that lead to a decreased time requirement to conduct an FA of IMTB may be an important direction for research.

Saini et al. (2018) showed that ongoing visual inspection using the structured criteria developed by Roane et al. (2013) could potentially shorten the duration of FAs of destructive behavior by 40%. Saini et al. applied the structured criteria to data from published FAs of destructive behavior and calculated exact agreement and four statistics that other disciplines use to evaluate predictive validity including sensitivity, specificity, positive-predictive value, and negative-predictive value. The criterion measures included post hoc visual inspection, post hoc author interpretation, and the consensus agreement between post hoc visual inspection and post hoc author interpretation. Levels of agreement between Saini et al.'s ongoing visual inspection procedure and the criterion visual inspection procedures were 93%, 80%, and 95%, respectively. This suggests that their ongoing visual inspection procedure was a valid method of interpreting FAs of destructive behavior.

Whether the ongoing visual inspection procedure Saini et al. (2018) described would have the same predictive-validity metrics and would increase the efficiency of FAs in a different clinical context from the original study is unclear. Although the procedures for the FA of destructive behavior and for IMTB share similarities, they also differ in some ways, and these intricacies can perhaps impact the visual analysis of FA outcomes. One salient difference is that the feeder presents a bite or drink and a vocal prompt to "take a bite (drink)" across test and control conditions in most studies that have described FAs of IMTB (e.g., Girolami & Scotti, 2001; Piazza et al., 2003). By contrast, antecedents, such as availability of therapist attention or instruction delivery, differ across test and control conditions of FAs of destructive behavior. Typically, the presence of the bite or drink in the control condition evokes IMTB and this may result in the control series of these FAs having elevated levels of behavior (Bachmeyer et al., 2019) whereas the FAs of destructive behavior are specifically constructed so as to not have the establishing operation present in the control series (Hanley et al., 2003), which typically yields low levels of behavior to serve as a comparison for test conditions. We believe this is a potentially impactful difference between the FA of IMTB and the FA of destructive behavior.

Additionally, the opportunity for the occurrence of IMTB varies across trials between test and control conditions. For example, in the escape condition of an FA of IMTB, the spoon can be within arm's reach for 30 s, but it is removed upon the occurrence of IMTB. However, the spoon stays within arm's reach for 30 s in the control condition, regardless of the child's behaviors; therefore, there are more opportunities to engage in IMTB in the control condition. Meanwhile, the opportunities for a child to engage in destructive behavior during an FA for this target behavior stays the same across conditions. Given these procedural differences, the purpose of the current study was to replicate the visual inspection procedure Saini et al. (2018) described and extend it to FAs of IMTB.

Method

Inclusion Criteria

The research team consisting of a master's level Board Certified Behavior Analyst®s (BCBA®s), two predoctoral psychology interns,

and a research assistant reviewed the records of 121 children admitted to a day treatment program that specialized in the treatment of pediatric feeding disorders from 2006 to 2016 to identify FA data for inclusion. The records were divided such that one member of the research team reviewed a record independently for inclusion in the study. A second member of the research team reviewed the same record after the first member of the research team finished their review to ensure accuracy of inclusion in the study. There were no discrepancies when the second member of the team reviewed the records. The inclusion criteria were: (a) a therapist or trained caregiver had conducted an FA of IMTB using a single-case pairwise research design (Iwata et al., 1994), (b) observers collected frequency data on IMTB during the FA, (c) the FA included at least one test condition and a control condition, (d) the FA included at least three data points for each condition, and (e) the data for the FA were graphed in a Microsoft® Excel® spreadsheet. Exclusion criteria included: (a) a therapist or caregiver had not conducted an FA of IMTB with the patient, or (b) a therapist or caregiver had conducted an FA, but the FA did not meet the inclusion criteria (e.g., fewer than three data points per condition). Nine FAs included a tangible test condition, and we excluded those data due to their small number.

The record review identified 111 children who had data that met the inclusion criteria and 10 children who had data that met the exclusion criteria, resulting in 219 FA data sets that met the inclusion criteria. The number of FAs exceeded the number of participants because some participants had one FA with solids and one with liquids, when both were targets of treatment. A flow chart describing study enlistment is shown in Figure 1.

Figure 1
Study Enlistment Flow Chart
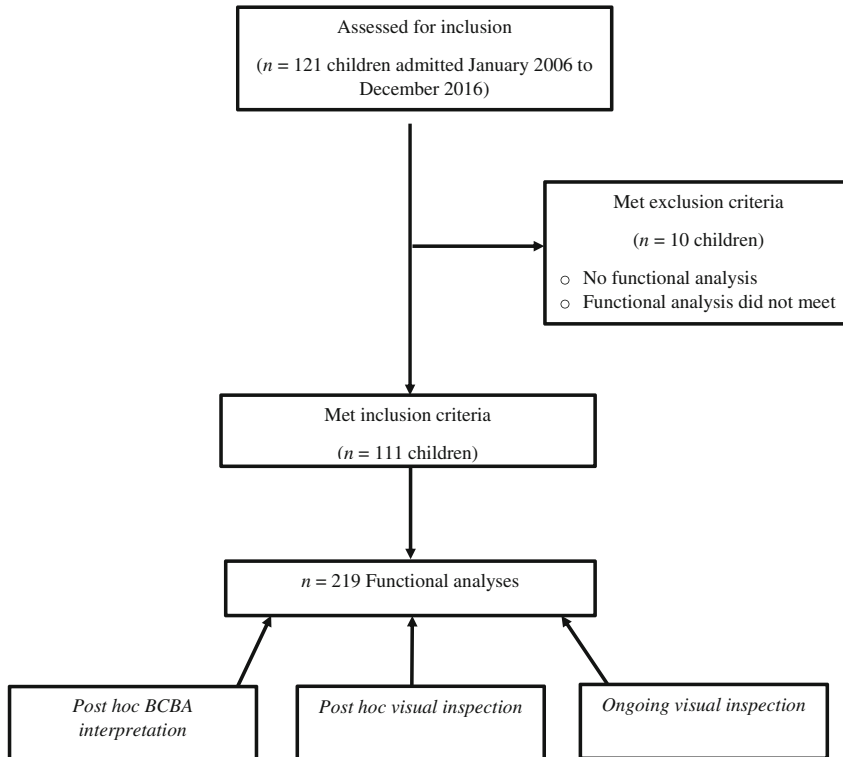
Of the 111 participants, 56% were 0-2 years old and 44% were 2-13 years old, 55% were male, 23% had a developmental diagnosis (i.e., autism spectrum disorder), and 77% had a medical diagnosis (i.e., gastroesophageal reflux disease). A physician and a speech language pathologist who specialized in feeding and swallowing cleared every participant as medically appropriate for the day treatment program and safe for oral feeding (e.g., no risk for aspiration). The research team, supervised by a licensed psychologist with expertise in the assessment and treatment of ARFID, conducted a retrospective medical-record review for the participants to confirm that they met the criteria for ARFID. Table 1 reports detailed demographics.

Functional Analysis Procedure

The FA procedure in the current study followed the procedures described by Kirkwood et al. (2021). The 219 FAs included escape and

Table 1

Participant Demographics (n = 111)

Participant variable	n
Age	
Infant (0-2 years)	62 (56%)
Children (2-13 years)	49 (44%)
Sex	
Male	61 (55%)
Female	50 (45%)
Developmental diagnosis*	
Intellectual or developmental disability	22 (18%)
Autism spectrum disorder	6 (5%)
Medical history*	
Anatomical abnormalities	10 (8%)
Cardiac disorder or dysfunction	7 (6%)
Enzyme deficiency or excess	4 (3%)
Failure to thrive	56 (46%)
Gastrostomy-tube dependence	64 (53%)
Gastroesophageal reflux disease	38 (31%)
Gastrointestinal problems	16 (13%)
Nervous system disorder or dysfunction	2 (2%)
Oral-motor dysfunction	8 (7%)
Prematurity	23 (19%)
Respiratory disorder or dysfunction	28 (23%)
Vision impairments	4 (3%)
Other	35 (29%)

* Participants may have had more than one developmental or medical diagnosis.

attention test conditions and a control condition. Across all pairwise FAs, the order for test and control conditions was randomized and the feeder presented five bites in solids sessions or five drinks in liquid sessions.

Observers used laptop computers with DataPal software, a beta version of BDataPro (Bullock et al., 2017), to record the frequency of IMTB. Observers scored IMTB when the utensil was within arm's reach of the participant, and the participant: (a) turned their head 45° or more in any direction away from the utensil, (b) moved their head away from the utensil 5 cm or more, or (c) blocked or covered their mouth with their hands or an object. DataPal calculated IMTB per minute by dividing the number of IMTB by the duration the utensil was within arm's reach of the participant.

A BCBA or licensed psychologist specializing in the assessment and treatment of ARFID supervised clinical teams who conducted the FAs and collected data. The clinical teams included graduate students, predoctoral interns, and Registered Behavior Technicians®. The clinical team sometimes also included a caregiver who received behavioral skills training on the procedures of an FA of IMTB. In the sample of 219 FAs, three were run by a caregiver. As part of their training, each clinical team member participated in an intensive training on data collection that included the review of written operational definitions and direct observation of sessions. Next, the team member practiced collecting data during real-time sessions and assessed their interobserver agreement with a data collector who had demonstrated acceptable reliability. The team member collected data used for clinical care and research only after their interobserver agreement with the reliable data collector was 80% or more for at least three consecutive FA sessions.

Study Overview

First, BCBAs analyzed FA graphs using traditional visual inspection (Kazdin, 1982), which we

call post hoc BCBA interpretation. Then, the research team used the structured criteria (Roane et al., 2013) to analyze the FA graphs post hoc, which we call post hoc visual inspection. Finally, the research team simulated applying the structured criteria in an ongoing fashion (Saini et al., 2018), which we call ongoing visual inspection. We applied post hoc BCBA interpretation, post hoc visual inspection, and ongoing visual inspection to each pairwise comparison (i.e., escape vs. control, attention vs. control) for every FA graph.

Post Hoc BCBA Interpretation

We asked nine BCBAs who had previously undergone the training and experiential standards noted above for the assessment and treatment of ARFID, including conducting and interpreting FAs of IMTB, to use visual inspection to interpret the FAs. The BCBAs self-reported years of experience ranging from 2-30 years. We divided the nine BCBAs into three groups of three, such that each BCBA group had one senior level BCBA-Doctoral®, one masters level BCBA with several years of specialization, and one graduate level BCBA undergoing specialization of assessment and treatment of ARFID.

The first and second authors randomly divided the 219 de-identified FA graphs into three groups of 73 graphs and created accompanying spreadsheets for each group. Each FA graph had a numerical identifier printed on the graph, and the spreadsheet listed these numerical identifiers. The first and second authors then randomly assigned one set of 73 FA graphs to each group of BCBAs. Each BCBA received a spreadsheet with the list of 73 numbered FA graphs and a copy of each graph assigned to their group.

The first author instructed the BCBAs to visually inspect their assigned FA graphs and indicate on the accompanying spreadsheet next to the numerical identifier for each graph whether the FA data identified an escape-only function, an attention-only function, escape and attention functions, or no function. Although

the current authors divided BCBAs into groups, the first author instructed each BCBA to interpret the FAs assigned to their group independently, without discussing their interpretations with BCBAs in their group or in the other two groups such that other reviewers were blind to the others' conclusions. The BCBAs returned their spreadsheets to the first or second author when they had interpreted their 73 assigned FA graphs and written their interpretation on the spreadsheet.

The first and second authors compared the interpretations of the three BCBAs in each group for each of the 73 FA graphs assigned to their group. We used a consensus-agreement procedure (Hagopian et al., 1997) to interpret each FA graph in which it was concluded that the FA identified escape-only, attention-only, escape and attention, or no function if at least two of three BCBAs in the group agreed that the FA identified the same single function, a multiply controlled function, or no function. For example, if the first and second BCBAs identified an escape function, but the third BCBA identified an escape and attention function, we concluded that the function was escape. The groups produced a majority agreement on every graph.

The first author entered the consensus-agreement interpretation into a Microsoft® Excel® spreadsheet. The spreadsheet counted the number of FAs that identified an escape-only function, an attention-only function, multiply controlled function, or no function, respectively, divided the number by 219, and converted the ratios to percentages.

Post Hoc Visual Inspection

An author of Saini et al. (2018) reviewed a Microsoft® PowerPoint® presentation and gave written instructions to the current study's first author and two research assistants. The materials summarized the formulas and rules for using structured criteria. The first author programmed Microsoft® Excel® spreadsheets to apply the structured criteria in post hoc visual inspection.

A formula in the spreadsheet calculated a mean and a standard deviation for IMTB per minute in the control condition for each pairwise comparison (i.e., one for the escape comparison and one for the attention comparison). The spreadsheet calculated one standard deviation above the mean, which was the upper criterion level. The spreadsheet also calculated one standard deviation below the mean, which was the lower criterion level. We subtracted the number of data points below the lower-criterion line, which is denoted here as l, from the number of data points above the upper-criterion line, which is denoted here as u. Separately, we then divided the number of data points in the test condition, which is denoted here as n, by two (i.e., $u - l \geq n / 2$). If the difference between the number of data points below the lower-criterion line and the number of data points above the upper-criterion line ($u - l$) was greater than or equal to the total number of data points in the test condition divided by two ($n / 2$), and the data did not meet criteria for trends, low magnitude of effects, or multiple maintaining variables (Roane et al., 2013), the tested consequence functioned as a reinforcer for IMTB. If the difference was not greater than or equal to the total number of data points in the test condition divided by two, the tested consequence did not function as a reinforcer for IMTB.

Ongoing Visual Inspection

We applied the structured criteria (Roane et al., 2013) to the first three pairs of data points of a test and control comparison and did not analyze the data for that comparison further if ongoing structured criteria identified a function. If ongoing visual inspection did not identify a function using the first three pairs of data points, we applied the criteria to the first four pairs of data points of the test and control comparison. We repeated the process of adding one pair of data points and applying the criteria until either we identified a function or until we had applied the criteria to the entire pairwise comparison of the test and control conditions.

We also applied the special rules Roane et al. (2013) described for trends.

Consensus Agreement

Saini et al. (2018) used consensus agreement between post hoc visual inspection and post hoc author interpretation as a criterion measure. In the current study, the consensus-agreement measure included FA graphs for which post hoc BCBA interpretation and post hoc visual inspection had the same interpretation and excluded FA graphs for which the two visual inspection procedures did not have the same interpretation. For example, if post hoc BCBA interpretation and post hoc visual inspection identified attention as the reinforcer, we included the data for the consensus-agreement measure. By contrast, if post hoc BCBA interpretation identified attention as the reinforcer but post hoc visual inspection did not identify attention as the reinforcer, we did not include the data in the consensus-agreement measure.

Data Analyses

We used the procedure Saini et al. (2018) described to assess the accuracy of ongoing-visual inspection. The first author created a Microsoft® Excel® spreadsheet for the 219 FA graphs. The spreadsheet had the numerical identifier for each graph and the identified function(s) from post hoc BCBA interpretation, post hoc visual inspection, and ongoing visual inspection. We evaluated the accuracy of ongoing visual inspection separately for escape tests and attention tests because the FAs in the current study were conducted using a pairwise design. Therefore, the first author created a Microsoft® Excel® spreadsheet to analyze escape tests and a spreadsheet to analyze attention tests of each FA. The first author and the research assistants compared interpretations of ongoing visual inspection to: (a) post hoc BCBA interpretation, (b) post hoc visual inspection, and (c) the consensus-agreement measure, in that order.

Exact Agreement

We used exact agreement to make the primary comparisons between ongoing-visual inspection and the three criterion procedures for escape tests and for attention tests. The first author programmed the Microsoft® Excel® spreadsheet to calculate exact agreement for each of the three criterion procedures. Exact agreement assessed whether ongoing-visual inspection and post hoc procedures produced the same interpretation for a given FA graph. The Microsoft® Excel® spreadsheet scored an exact agreement when ongoing visual inspection and the criterion procedure targeted for comparison produced the identical interpretation for the FA; that is, both procedures identified an escape-only function, an attention-only function, escape and attention functions, or no function. The procedures disagreed when: (a) they identified different functions (e.g., escape vs. attention); (b) one identified a function and the other did not identify a function (e.g., escape vs. no function); (c) one identified a single function and the other identified multiple functions (e.g., escape vs. escape and attention); or (d) one identified multiple functions and one identified no function (e.g., escape and attention vs. no function). The spreadsheet then divided the number of exact agreements by the number of FA graphs and converted the ratio to a percentage.

Sensitivity, Specificity, Positive-Predictive Value, Negative-Predictive Value, and Total Agreement

The first author programmed the spreadsheet to calculate true positives, false positives, false negatives, and true negatives for escape tests and attention tests separately. The spreadsheet scored a true positive for the escape test or a true positive for the attention test when ongoing visual inspection and the post hoc procedure identified escape or identified attention, respectively, as reinforcement for IMTB. The spreadsheet scored a false positive for the escape test or a false positive for the attention test when ongoing visual inspection identified escape or identified attention, respectively, as reinforcement for IMTB and the post hoc procedure did not identify escape or did not identify attention, respectively, as reinforcement for IMTB. The spreadsheet scored a false negative for the escape test or a false negative for the attention test when ongoing visual inspection did not identify escape or did not identify attention, respectively, as reinforcement for IMTB and the post hoc procedure identified escape or identified attention, respectively, as reinforcement for IMTB. The spreadsheet scored a true negative for the escape test or a true negative for the attention test when ongoing visual inspection and the post hoc procedure did not identify escape or did not identify attention, respectively, as reinforcement for IMTB.

The first author programmed the spreadsheet to calculate sensitivity, specificity, positive-predictive value, and negative-predictive value to evaluate the predictive validity of ongoing visual inspection relative to the post hoc procedures for escape tests and attention tests separately. Sensitivity was the proportion of FAs for which the visual inspection procedure correctly identified a function, calculated by dividing true positives by the sum of true positives and false negatives and converting the ratio to a percentage. Specificity was the proportion of FAs for which the visual inspection procedure did not identify a function, calculated by dividing true negatives by the sum of true negatives and false positives and converting the ratio to a percentage. Positive-predictive value was the probability that the function(s) identified with ongoing visual inspection was accurate, calculated by dividing true positives by the sum of true positives and false positives and converting the ratio to a percentage. Negative-predictive value was the probability that the function(s) not identified with ongoing visual inspection was accurate, calculated by dividing true negatives by the sum of true negatives and false negatives and converting the ratio to a percentage. The spreadsheet also calculated total agreement by dividing the sum of true positives and true

negatives by the sum of true positives, false positives, true negatives, and false negatives and converting the ratio to a percentage.

Efficiency

We evaluated the efficiency of ongoing visual inspection for escape tests and attention tests separately and analyzed only the FAs for which ongoing visual inspection and the consensus agreement between post hoc BCBA interpretation and post hoc visual inspection produced the same interpretation. The spreadsheet: (a) calculated the mean number of sessions conducted to identify a function for FA data sets using ongoing visual inspection and using consensus agreement, (b) subtracted these means for ongoing visual inspection from those of consensus agreement, (c) divided the difference by the mean number of sessions for consensus agreement, and (d) converted the ratio to a percentage.

Reliability

We assessed interrater agreement between the three BCBAs in each of the three groups for post hoc BCBA interpretation of the 73 FA graphs per group. We calculated interrater agreement for each graph by dividing the number of agreements by the number of agreements and disagreements and converting this ratio to a percentage. We compared the interpretations of the first and second BCBAs, the first and third BCBAs, and the second and third BCBAs for each FA graph. An agreement occurred when the BCBAs made an identical interpretation of the FA graph (i.e., escape-only function, attention-only function, escape and attention functions, or no function). The mean interrater agreement was 87% (range, 33%-100%).

We also assessed reliability for the post hoc visual inspection and for ongoing visual inspection. The first author and a research assistant independently used post hoc visual inspection to interpret 75 FA graphs (34%) and ongoing visual inspection to interpret 75 FA graphs (34%). We calculated interrater agreement as described above, which was 99% for post hoc structured criteria and 99% for ongoing structured criteria.

Results

Exact Agreement

Exact agreement assessed whether ongoing visual inspection produced the same interpretation per FA as the post hoc procedures. Exact agreement between ongoing visual inspection and post hoc BCBA interpretation was 80%, between ongoing visual inspection and post hoc visual inspection was 79%, and between ongoing visual inspection and the consensus agreement of post hoc BCBA interpretation and post hoc visual inspection was 80% (see Table 2).

Sensitivity, Specificity, Positive- and Negative-Predictive Values, and Total Agreement

Table 3 shows the number of true positives, false positives, false negatives, and true negatives between ongoing visual inspection and post hoc BCBA interpretation (first table spanner), post hoc visual inspection (second table spanner), and the consensus agreement between post hoc BCBA interpretation and post hoc visual inspection (third table spanner) for escape tests. Figure 2 shows the percentages for sensitivity, specificity, positive-predictive value, negative-predictive value, and total agreement between ongoing visual inspection and post hoc BCBA interpretation (black bars), post hoc visual inspection (gray bars), and the consensus agreement of post hoc BCBA interpretation and post hoc visual inspection (open bars) for escape tests. In summary, for escape tests there was a sensitivity that ranged 98%-99%, specificity ranged between 82%-84%, positive-predictive value was 95%, negative-predictive value ranged between 93%-98%; and total agreement was 95%.

Table 2

Exact Agreement Between Ongoing Visual Inspection and Post Hoc Procedures

Function	No.			
	Ongoing visual inspection	Post hoc BCBA interpretation	Disagreements	Exact agreement
Escape	60	76	16	79%
Attention	8	4	4	50%
Escape and attention	107	89	18	83%
No function	44	50	6	88%
Total	219	219	44	80%
	Ongoing visual inspection	Post hoc visual inspection	Disagreements	Exact agreement
Escape	60	76	16	79%
Attention	8	7	1	88%
Escape and attention	107	85	22	79%
No function	44	51	7	86%
Total	219	219	46	79%
	Ongoing visual inspection	Consensus agreement	Disagreements	Exact agreement
Escape	53	66	13	80%
Attention	6	2	4	33%
Escape and attention	91	76	15	84%
No function	42	48	6	88%
Total	192	192	38	80%

Table 4 shows the number of true positives, false positives, false negatives, and true negatives between ongoing visual inspection and post hoc BCBA interpretation (first table spanner), post hoc visual inspection (second table spanner), and the consensus agreement between post hoc BCBA interpretation and post hoc visual inspection (third table spanner) for attention tests. Figure 3 shows the percentages for sensitivity, specificity, positive-predictive value, negative-predictive value, and total agreement for ongoing visual inspection compared to post hoc BCBA interpretation (black bars), post hoc visual inspection (gray bars), and the consensus agreement of post hoc BCBA interpretation and post hoc visual inspection (open bars) for attention tests. In summary, for attention tests sensitivity ranged between 94%-100%, specificity ranged between 73%-81%, positive-predictive value ranged between 75%-81%, negative-predictive value ranged between 93%-100%, and total agreement ranged between 83%-90%.

Efficiency of Using Ongoing Visual Inspection

The mean number of total sessions needed to determine whether escape or attention functioned

Table 3

True and False Positives and Negatives for Escape Function

Ongoing-visual inspection	Criterion measure	
	Post hoc BCBA interpretation	
	Function present	Function absent
Function present	157 (a)	11 (b)
Function absent	8 (c)	40 (d)
	Post hoc visual inspection	
	Function present	Function absent
Function present	156 (a)	12 (b)
Function absent	5 (c)	43 (d)
	Consensus agreement	
	Function present	Function absent
Function present	155 (a)	10 (b)
Function absent	5 (c)	40 (d)

Note. True positives are in the cells labeled *a* under 'function present' for ongoing visual inspection and the post hoc procedure. False positives are in the cells labeled *b* under 'function present' for ongoing visual inspection and 'function absent' for the post hoc procedure. False negatives are in the cells labeled *c* under 'function absent' for ongoing visual inspection and 'function present' for the post hoc procedure. True negatives are in the cells labeled *d* under 'function absent' for ongoing visual inspection and the post hoc procedure.

as reinforcement was 10 (range, 3-58) and 11 (range, 3-63), respectively, for ongoing visual inspection and 16 (range, 6-70) and 17 (range,
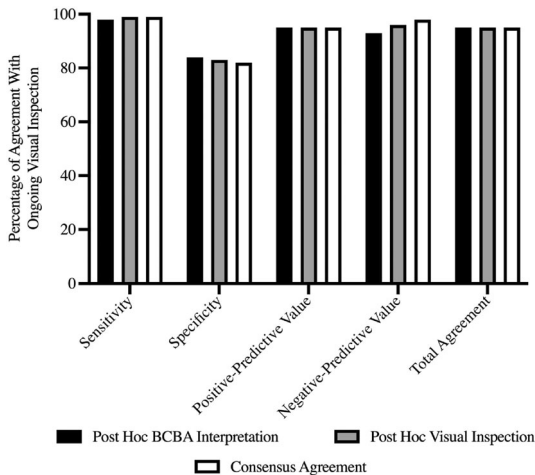
Table 4

True and False Positives and Negatives for Attention Function

Ongoing visual inspection	Criterion measure	
	Post hoc BCBA interpretation	
	Function present	Function absent
Function present	87 (a)	29 (b)
Function absent	6 (c)	80 (d)
	Post hoc visual inspection	
	Function present	Function absent
Function present	91 (a)	25 (b)
Function absent	0 (c)	86 (d)
	Consensus agreement	
	Function present	Function absent
Function present	82 (a)	19 (b)
Function absent	0 (c)	80 (d)

Figure 3

Levels of Agreement Between Ongoing Visual Inspection and Post Hoc Procedures for Attention Tests

Figure 2

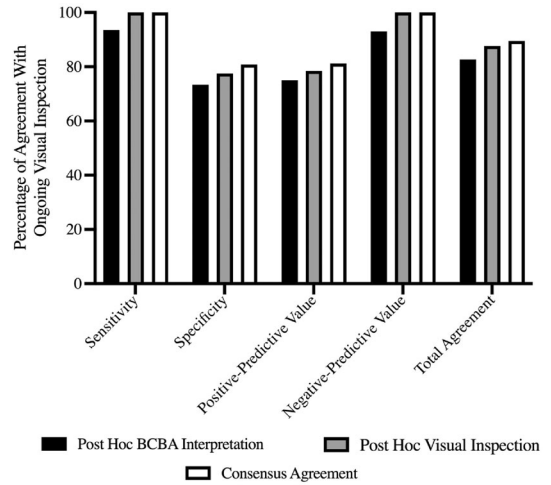*Levels of Agreement Between Ongoing Visual Inspection and Post Hoc Procedures for Escape Tests*

6-70), respectively, for consensus agreement of post hoc BCBA interpretation and post hoc visual inspection. Ongoing visual inspection decreased the number of sessions to interpret the escape and attention tests by 38% and 36%, respectively.

Discussion

In the current study, we simulated applying structured criteria to ongoing FAs of IMTB using the procedure Saini et al. (2018) described. Saini et al. used data from studies on FA of destructive behavior and assessed the predictive validity of the ongoing visual inspection procedure using three criterion measures. One criterion measure was the post hoc interpretations of the authors of the published FA data. Unlike Saini et al., we used FA data from patients admitted to a day treatment program, and we asked BCBAs with specialized training in ARFID who routinely conducted and interpreted FAs of IMTB to interpret the FA graphs. We called this post hoc BCBA interpretation. Like Saini et al., we applied structured criteria to the FA data as the second criterion measure (i.e., post hoc visual inspection) and used the consensus agreement between post hoc BCBA interpretation and post hoc visual inspection as the third criterion measure.

Exact agreement levels in the current study between the ongoing visual-inspection procedure and the post hoc procedures were lower than those of Saini et al. (2018). Exact agreement in Saini et al. was 93% for post hoc visual inspection, 80% for post hoc author interpretation, and 95% for the consensus agreement

between post hoc visual inspection and post hoc author interpretation. Exact agreement in the current study was 80% for post hoc BCBA interpretation, 79% for post hoc visual inspection, and 80% for the consensus agreement between post hoc BCBA interpretation and post hoc visual inspection. The lower end of the exact agreement range in the current study was 79%, which is close to but less than the acceptable level of agreement in applied behavior analysis research. The results suggest that ongoing visual inspection was not as good at identifying the function of IMTB as it was for identifying the function of destructive behavior.

Like Saini et al. (2018), we calculated sensitivity, specificity, positive-predictive value, negative-predictive value, and total agreement. We used these measures to examine why levels of exact agreement were lower in the current study relative to Saini et al. Overall, the validity metrics of ongoing visual inspection for the escape test were acceptable, suggesting that interpretations of the escape test did not contribute substantially to the lower levels of exact agreement. Compared to the three post hoc procedures, (a) sensitivity was above 90%, (b) specificity was above 81%, (c) positive-predictive value was above 90%, and (d) negative-predictive value was above 92%. The results suggest that ongoing visual inspection (a) was likely to identify an escape function when escape was a reinforcer for IMTB, (b) slightly over-identified an escape function when escape was not a reinforcer for IMTB, (c) had a high probability of identifying an escape function accurately, and (d) had a high probability of identifying the absence of an escape function accurately.

The validity metrics of ongoing visual inspection for the attention test were more variable. Compared to the three post hoc procedures, (a) sensitivity was above 94%, (b) specificity was above 73%, (c) positive-predictive value was above 75%, and (d) negative-predictive value was above 93%. The results suggest that ongoing visual inspection (a) was likely to identify an attention function when attention was a

reinforcer for IMTB, (b) over-identified an attention function when attention was not a reinforcer for IMTB, (c) had a lower probability of identifying an attention function accurately, and (d) had a high probability of identifying the absence of an attention function accurately.

We visually inspected the FA graphs for which ongoing visual inspection and post hoc BCBA interpretation disagreed to identify the factors that may have accounted for the lower levels of specificity and positive-predictive value for the attention test. Ongoing visual inspection identified attention as reinforcement for IMTB, but post hoc BCBA interpretation did not for 29 FAs. For these 29 FAs, ongoing visual inspection identified attention as reinforcement after three (18 FAs), four (seven FAs), five (one FA), or six (two FAs) pairs of data points. The general pattern observed when visually inspecting these graphs was that the rate of IMTB (a) was variable and decreased in the attention condition such that rates became comparable with the control condition after ongoing visual inspection identified attention as reinforcement, (b) tended to be variable in the control condition, and (c) was equal or crossed in the attention and control conditions for some sessions but still met the ongoing visual inspection criteria for an attention function. Figure 4 shows

Figure 4

Example of Functional Analysis Graph in Which Ongoing Visual Inspection Identified Attention as a Reinforcer for Inappropriate Mealtime Behavior, but Post Hoc Visual Inspection Did Not
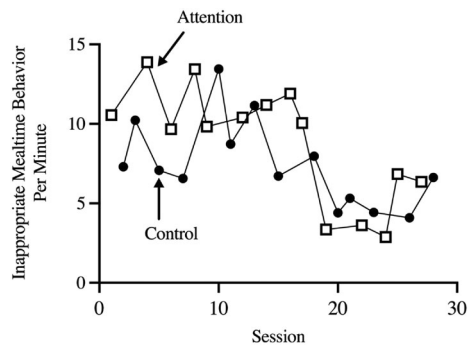
an example in which ongoing visual inspection identified an attention function after three pairs of data points. Post hoc BCBA interpretation did not identify an attention function because the rate of IMTB decreased and was equivalent in the attention and control conditions by the end of the comparison.

Alternatively, some disagreements between ongoing visual inspection and post hoc procedures were due to variability in the control condition. There were six FAs for which post hoc BCBA interpretation identified attention as a reinforcer, but ongoing visual inspection did not. The rate of IMTB was high (i.e., outliers) in one or two initial sessions of the control condition in these FAs. The rate of IMTB decreased in the control condition and increased or remained stable in the attention condition as the phase progressed. The rate of IMTB was consistently higher in the attention relative to the control condition by the end of the phase, resulting in post hoc BCBA interpretation identifying an attention function. By contrast, the outliers in the initial control-condition sessions inflated the standard deviation for the rate of IMTB, which the ongoing visual inspection procedure uses to calculate the upper- and lower-criterion lines. The inflated standard deviation skewed the upper- and lower-criterion lines upward and downward, respectively. The rate of IMTB was consistently higher in the attention relative to the control condition by the end of the phase. The rate was not sufficiently high in the attention condition, however, it met the ongoing visual inspection criteria for number of data points above the upper-criterion line to identify an attention function.

Variable or episodic high rates of IMTB in the control condition may be more likely if the feeder presents a bite or a drink and a vocal prompt to "take a bite (drink)" across test and control conditions (Girolami & Scotti, 2001; Piazza et al., 2003). The presentation of bites or drinks across test and control conditions may have an evocative effect on IMTB or may reduce the discriminability of the conditions.

The discriminability between attention and control conditions may be particularly problematic for some children given additional similarities between the two conditions (beyond bite or drink presentation). The spoon or cup remains in its original position in space if the child does not accept the bite or drink in both conditions. Some children may not discriminate between the 20-30 s of contingent attention the feeder delivers if the child engages in IMTB in the attention condition and the relatively continuous, noncontingent attention the feeder delivers in the control condition. These condition similarities may increase the likelihood that rates of IMTB may be high and variable during the attention–control comparison for some children and may have contributed to the disagreements between the use of structured criteria and visual inspection interpretations for the attention tests. The results of the current study suggest that a conservative approach to a pairwise comparison of attention and control conditions is indicated (i.e., conducting more sessions for more stability in the data) if behavior analysts conduct FAs of IMTB as described above.

The evocative effects of bite presentation in the control condition could be mitigated by using a reversal design as in Piazza et al. (2003) rather than a pairwise comparison. Mean IMTB per minute across participants in the control condition was 0.7 with a standard deviation of 0.52 in Piazza et al. By contrast, IMTB per minute in the control condition in the current study ranged from 0-54. We are unable to report the mean and standard deviation for the sample in this study; however, we hypothesize that it is larger than those in the Piazza et al. study based on the range. Therefore, we wonder if the pairwise design may be associated with higher and more variable rates of IMTB in the control condition relative to a reversal design for some participants. An alternative modification might be to conduct control-condition sessions until the rate of

IMTB stabilizes and then begin the test–control comparison. Stabilizing the rate of IMTB may increase the validity metrics for the attention test of FAs of IMTB.

Another goal of the current study was to examine the efficiency of ongoing visual inspection for identifying the function of IMTB. The results of Saini et al. (2018) suggested that using ongoing visual inspection would have reduced the number of sessions needed to interpret FAs of destructive behavior by 41%. Similarly, using ongoing visual inspection would have reduced the number of sessions needed to interpret escape and attention tests of FAs of IMTB by 38% and 36%, respectively. Training clinicians to use ongoing visual inspection also increases the efficiency of FAs and could increase its use in routine clinical practice as Retzlaff et al. (2020) showed. One limitation is that we did not present session-duration data as a measure of efficiency.

Similarly, we did not include data from tangible tests in the current study given how few data sets we had with a tangible test. There were only nine tangible tests in the FA graphs that met inclusionary criteria for this study. Due to the sample size of tangible tests being so limited, we did not find it to be useful for generalizing conclusions. Therefore, we did not conduct analyses on these tests. Future research should evaluate if the validity metrics for structured criteria with tangible tests are like those of the escape and attention tests. Another limitation of the current study is that we only included FAs conducted with a pairwise design. Future research should test the structured criteria with FAs conducted with other single-case designs, such as the reversal design.

Modifications to ongoing-visual inspection, the FA procedure, the design used to conduct the FA, or a combination may be needed to improve the validity metrics for interpreting the attention test using structured criteria. Further refinement of ongoing visual inspection for use with FAs of IMTB may be indicated given

(a) clinician-reported time restrictions are a barrier to conducting FAs (Hanley et al., 2003), (b) visual inspection is a subjective method of data interpretation (Fisch, 1998), and (c) the inconsistent reliability of visual inspection (Saini et al., 2018). Thus, more time-efficient procedures for completing FAs of IMTB that produce results with high interrater agreement seem warranted.

REFERENCES

Bachmeyer, M. H., Kirkwood, C. A., Criscito, A. B., Mauzy IV, C. R., & Berth, D. P. (2019). A comparison of functional analysis methods of inappropriate mealtime behavior. *Journal of Applied Behavior Analysis*, *52*(3), 603-621. https://doi.org/10.1002/jaba.556

Bachmeyer, M. H., Piazza C. C., Fredrick, L. D., Reed, G. K., Rivas, K. D., & Kadey, H. J. (2009). Functional analysis and treatment of multiply controlled inappropriate mealtime behavior. *Journal of Applied Behavior Analysis*, *42*(3), 641-658. https://doi.org/10.1901/jaba.2009.42-641

Bullock, C. E., Fisher, W. W., & Hagopian, L. P. (2017). Description and validation of a computerized behavioral data program: "BDataPro". *The Behavior Analyst*, *40*(1), 275-285. https://doi.org/10.1007/s40614-016-0079-0

Fisch, G. S. (1998). Visual inspection of data revisited: Do the eyes still have it? *The Behavior Analyst*, *21*(1), 111-123. https://doi.org/10.1007/BF03392786

Girolami, P. A., & Scotti, J. R. (2001). Use of analog functional analysis in assessing the function of mealtime behavior problems. *Education and Training in Mental Retardation and Developmental Disabilities*, *36*(2), 207-223. http://www.jstor.org/stable/23879736

Hagopian, L. P., Fisher, W. W., Thompson, R. H., Owen-DeSchryver, J., Iwata, B. A., & Wacker, D. P. (1997). Toward the development of structured criteria for interpretation of functional analysis data. *Journal of Applied Behavior Analysis*, *30*(2), 313-326. https://doi.org/10.1901/jaba.1997.30-313

Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis in Practice*, *5*(1), 54-72. https://doi.org/10.1007/BF03391818

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis*, *36*(2), 147-185. https://doi.org/10.1901/jaba.2003.36-147

Iwata, B. A., Duncan, B. A., Zarcone, J. R., Lerman, D. C., & Shore, B. A. (1994). A sequential,

test-control methodology for conducting functional analyses of self-injurious behavior. *Behavior Modification*, *18*(3), 289-306. https://doi.org/10.1177%2F01454455940183003

Kazdin, A. E. (1982). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science*, *13*, 33-47. https://psycnet.apa.org/record/1983-11589-001

Kirkwood, C. A., Piazza, C. C., & Peterson, K. A. (2021). A comparison of function- and nonfunction-based extinction treatments for inappropriate mealtime behavior. *Journal of Applied Behavior Analysis*, *54*(3), 928-945. https://doi.org/10.1002/jaba.825

Oliver, A. C., Pratt, L. A., & Normand, M. P. (2015). A survey of functional behavior assessment methods used by behavior analysts in practice. *Journal of Applied Behavior Analysis*, *48*(4), 817-829. https://doi.org/10.1002/jaba.256

Phalen, J. A. (2013). Managing feeding problems and feeding disorders. *Pediatrics in Review*, *34*(12), 549-557. https://doi.org/10.1542/pir.34.12.549

Piazza, C. C., Fisher, W. W., Brown, K. A., Shore, B. A., Patel, M. R., Katz, R. M., Sevin, B. M., Gulotta, C. S., & Blakely-Smith, A. (2003). Functional analysis of inappropriate mealtime behaviors. *Journal of Applied Behavior Analysis*, *36*(2), 187-204. https://doi.org/10.1901/jaba.2003.36-187

Retzlaff, B. J., Phillips, L. A., Fisher, W. W., Hardee, A. M., & Fuhrman, A. M. (2020). Using e-learning modules to teach ongoing-visual inspection of functional analyses. *Journal of Applied Behavior Analysis*, *53*(4), 2126-2138. https://doi.org/10.1002/jaba.719

Roane, H. S., Fisher, W. W., Kelley, M. E., Mevers, J. L., & Bouxsein, K. J. (2013). Using modified visual-inspection criteria to interpret functional analysis outcomes. *Journal of Applied Behavior Analysis*, *46*(1), 130-146. https://doi.org/10.1002/jaba.13

Saini, V., Fisher, W. W., & Retzlaff, B. J. (2018). Predictive validity and efficiency of ongoing visual-inspection criteria for interpreting functional analyses. *Journal of Applied Behavior Analysis*, *51*(2), 303-320. https://doi.org/10.1002/jaba.450