Intervention Selection Profile-Function: An Examination of Decisional Accuracy Relative to Traditional FBA Data

Katie Eklund[1], Stephen Kilgus[1], Miranda Zahn[1], Nathaniel von der Embse[2], Jessica Willenbrink[1], Elizabeth Davis[1], & Teagan Twombly[1]

University of Wisconsin-Madison[1]

University of South Florida[2]

School Psychology Quarterly

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180515 to the University of Wisconsin-Madison. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Correspondence regarding this article should be addressed to Katie Eklund at the University of Wisconsin-Madison, Department of Educational Psychology, Madison, WI 53706; email: katie.eklund@wisc.edu

Article word count: 7579 words

Abstract

The purpose of this study was to examine the accuracy of function-based decisions made in consideration of scores from the Intervention Selection Profile (ISP)-Function, a tool founded upon direct behavior rating methodology. The ISP-Function is designed to be a brief measure, given the need for efficient and low resource assessments in schools. Data from a previous investigation were used to create data reports for each of 34 elementary students with a history of exhibiting disruptive behavior in the classroom. The first report summarized ISP-Function data that the student's classroom teacher collected. The second report was representative of more typical functional behavior assessment (FBA), summarizing data collected via a functional assessment interview with the teacher, as well as systematic direct observation data. Nine school psychologists conducted blind reviews of these reports and derived decisions regarding the function of each student's behavior (e.g., adult attention or escape/avoidance). Gwet's agreement coefficients were statistically significant and suggested Fair to Almost Perfect correspondence between ISP-Function and FBA reports. Limitations and implications for practice are discussed herein.

Keywords: Functional behavior assessment, behavior intervention, Tier 2

Impact Statement

School psychologists and behavior consultants are often tasked with collecting functional behavioral assessment (FBA) data for children demonstrating disruptive behavior. The current study examines the utility of the Intervention Selection Profile (ISP)-Function, a brief assessment tool designed to efficiently collect data regarding the function of student behavior. Results demonstrate fair to almost perfect correspondence between the ISP-Function and traditional FBA reports.

Intervention Selection Profile-Function: An Examination of Decisional Accuracy Relative to Traditional FBA Data

Functional behavior assessment (FBA) is defined as a comprehensive multi-method, multi-informant process through which practitioners collect information to determine which environmental variables occasion (antecedents) and maintain (consequences) problem behaviors for a given student (O'Neill et al., 2015). A large body of research supports the utility of FBA data in designing effective interventions (Bruni et al., 2017; Gage et al., 2012). There are two primary categories of FBA methods: experimental and correlational. Experimental FBA procedures include structural or functional analyses of student behavior (Barnhill, 2005), and involve the manipulation of a student's environment to determine the conditions that occasion and maintain the problematic behavior. Various manipulations are made to the environment based on the perceived antecedent and consequences that are maintaining the student's behavior. By evoking the behavior through controlled experimentation, the school staff member, trained in behavior analytics, can be more certain of the conditions upon which a student will display a problematic behavior.

Experimental FBA procedures are often difficult to conduct in a school setting. For this reason, correlational methods such as systematic direct observation, interviews, and ratings scales are more commonly used (O'Neill et al., 1997). Systematic direct observation (SDO) is a correlational method for gathering functional behavior data that is direct and objective. SDO typically requires a trained observer to be present within relevant structured and unstructured settings (e.g., classroom and playground) to collect data across multiple sessions. This observational process yields highly objective and low inference data because they are collected during the time and setting when the behavior in question is displayed. Interviews and rating

scales then involve stakeholders (e.g., students, teachers, parents) providing information about circumstances and environmental stimuli commonly associated with the problem behavior. Both of these methods tend to afford more subjective and higher inference data (relative to SDO) given that data collection occurs at a time that is removed from the behavior in question by days, weeks, or months. Nevertheless, research has demonstrated their capacity to inform effective interventions (McIntosh et al., 2008).

Within multi-tiered systems of support intended to promote student social-emotional and behavioral functioning (e.g., positive behavioral interventions and supports [PBIS]), FBA is often reserved for students referred for intensive intervention at the Tier 3 level. However, there are increasing calls for the completion of brief FBAs at Tier 2 to support the delivery of targeted interventions (Reinke et al., 2013). Indeed, evidence suggests that Tier 2 interventions are more effective when modified or specifically selected to align with the function of a student's problem behavior (Boyd & Anderson, 2013; Kilgus et al., 2016). However, given the non-trivial number of students receiving Tier 2 services (e.g., 10-15% of a school population), some have raised concerns regarding the use of comprehensive, time-consuming FBA procedures at this level (Cheney et al., 2008). Accordingly, some scholars have suggested that given the relatively low stakes nature of Tier 2 decisions compared to Tier 3 (e.g., special education eligibility determinations), as well as the need for efficient and low resource assessment, brief FBAs conducted at Tier 2 could be founded upon the use of a single indirect assessment tool, such as semi-structured interviews or rating scales (Dunlap & Kern, 2018; McIntosh et al., 2008).

Intervention Selection Profile–Function

One such indirect assessment tool that was designed for use at Tier 2 is the *Intervention Selection Profile–Function* (ISP-Function). The ISP-Function is founded upon direct behavior

rating (DBR) methodology (Chafouleas, 2012) and involves five ratings to assess (1) the extent to which a student engages in an operationally defined problem behavior, and (2) the frequency with which that behavior is met with four consequences: escape/avoidance, adult attention, peer attention, and access to tangibles or activities. Each rating is completed using an 11-point unipolar graphic rating scale, which ranges from 0-10 and includes descriptive anchors at its beginning (0%, Never), middle (50%, Sometimes), and end (100%, Always).

To use the ISP-Function, the teacher selects relevant "rating periods," defined as times during which a student commonly engages in problem behavior (e.g., daily large-group math instruction from 10:00–10:45am). The teacher then completes the ISP-Function during these times across multiple time points to yield a stream of data for each of the five items. Each ISP-Function administration involves the teacher briefly and periodically observing the student throughout the rating period, involving a brief visual "check" of the student every 30-60 seconds while engaging in typical classroom instruction. The teacher then rates the five ISP-Function items immediately following the rating period. The ISP-Function can be completed consecutively, allowing for multiple observations in a single day (Kilgus, et al., 2019). This means that even in the case where 4-6 observations are completed, data collection can be completed within 1-2 days. The time a teacher spends considering a student's behavior is considered similar to the commitment associated with SDO administration. Furthermore, the time spent actively rating the student's behavior is considered less than that associated with other rating scales and semi-structured interviews, which involve far more specific evaluations of student behavior and related stimuli.

Once data have been collected across multiple time points, mean of scores for each individual item are derived. Mean problem behavior scores are examined to determine the

prevalence of problem behavior in evaluating the necessity of intervention. Mean consequence scores are interpreted in a manner consistent with conditional probability statistics, indicating the strength of the functional relation between a problem behavior and various consequences (Eckert et al., 2005).

Consequences included in the ISP-Function (i.e., escape/avoidance, adult attention, peer attention, and access to tangibles or activities) are consistent with similar measures such as the Motivation Assessment Scale (MAS; Durand & Crimmins, 1988) and the Functional Assessment Screening Tool (FAST; Iwata & DeLeon, 1995). Other measures, such as the Functional Assessment Checklist: Teachers and Staff (FACTS; March et al., 2000) have taken a step further to differentiate between adult and peer attention. The importance of this differentiation has also been established in experimental research (Campbell & Anderson, 2008).

Multiple studies to date have supported the development and initial validation of the ISP-Function (Kilgus et al., 2017; Kilgus et al., 2019). An initial study examined the accuracy of ISP-Function ratings completed by participants viewing videos of students within a classroom setting. Findings suggested that following a brief training that included opportunities to practice using the ISP-Function with performance feedback, participants were able to generate accurate ISP-Function scores that were highly consistent with systematic direct observation (SDO) data (Kilgus et al., 2017).

A subsequent study examined the reliability, validity, and accuracy of in-vivo ISP-Function ratings (Kilgus et al., 2019). Participants included 34 elementary school students and their classroom teachers. Each student had been identified via universal screening as commonly engaging in disruptive behavior. Each classroom teacher rated their participating student using the ISP-Function across three rating periods within the same setting. A research assistant

collected SDO data on the same five behavioral and consequence targets using partial-interval behavior-consequence (B-C) recording methodology. Three notable findings were documented. First, results indicated the three ISP-Function data points were sufficiently reliable (≥.70) for the adult attention and peer attention items, while the remaining items did not achieve adequate reliability across the three data points. Follow-up analyses suggested that 8-18 data points would be required to achieve reliabilities of .70 for escape/avoidance and access to items/activities. Second, findings supported the concurrent validity of disruptive behavior, adult attention, and peer attention items, which were moderately to highly correlated (r > .30 and .50) with SDO data. Correlations were in the low range for escape/avoidance and access to items/activities; however, these findings may have reflected floor effects in the data due to the low occurrence of consequences within the sample. Third, the ISP-Function yielded accurate estimates, with mean ISP-Function scores falling within only 0.33 to 1.81 points of mean SDO scores (on the 0–10 ISP-Function scale).

Next Steps in ISP-Function Research

Taken together, previous ISP-Function studies have yielded initial support for the tool and its promise as an efficient method of collecting functional assessment data. When viewed through an argument-based approach to instrument validation (Kane, 2013), existing evidence supports the *interpretation* of ISP-Function scores relative to their intended observational targets. Moving forward, there is a need for evidence to support the *use* of the ISP-Function as a functional assessment tool. Consistent with prior research (e.g., Iwata et al., 2013), such evidence might indicate whether practitioners who examine and interpret ISP-Function data arrive at accurate conclusions regarding the likely function of a student's problem behavior. Such evidence is integral to the central validity issue of whether the ISP-Function has the potential to

result in intended positive consequences (Messick, 1995); that is, the production of information that promotes positive Tier 2 intervention outcomes.

Previous research has yielded similar evidence in the context of DBR tools intended for use in progress monitoring (e.g., Christ et al., 2014). Riley-Tillman et al. (2008) examined DBR decisional accuracy, or the extent to which school psychologists reached similar data-based decisions when examining either DBR or SDO progress monitoring data. Two different studies were conducted. In Study 1, 92 school psychologists reviewed either DBR or SDO data reports, each of which included time series data depicted via a line graph. Half of the participants were randomly assigned to review SDO data while the other half reviewed DBR data. Participants in each condition reviewed two reports. Data within the reports were simulated and specifically manipulated so that Report 1 depicted an effective intervention, while Report 2 depicted an ineffective intervention. Report 1 was identical across the DBR and SDO conditions, providing the same case information and identical graphed data. The sole difference between the reports was the measure through which data were collected (i.e., DBR or SDO). Following their review of data, participants indicated whether the intervention should continue with no change, continue with changes, or discontinue and be replaced with another intervention. Study 2 (n = 99) was nearly identical in terms of procedures; however, in this study, data were actual outcome data collected for two cases from a prior study. Participants in one group reviewed the SDO data from these cases, while the other group reviewed the DBR data. Though all four datasets (2 case*2 measures) were representative of a moderate intervention effect, none of the data were identical given their non-simulated nature. Results of chi-square analyses suggested that across both studies, the two groups reached similar intervention-related decisions when examining paired data reports, thereby supporting decisional accuracy.

In summary, previous research has demonstrated preliminary support for the ISP-Function; however, additional research is needed to demonstrate the accuracy of conclusions derived in consideration of these data. Such research might examine whether school psychologists interpret ISP-Function data collected for a student in a manner consistent with more traditional FBA data collected for that same student. School psychologists have specialized training and knowledge in conducting assessments and selecting evidence-based interventions designed to address student behavioral concerns (NASP, 2020). Examining how they interpret data reports and use this information to make intervention decisions is critical when evaluating the utility of a new measure. Thus, the current study is designed to address an important gap in the ISP-Function and broad Tier 2 FBA literature.

Purpose of the Study

The purpose of this investigation was to employ methods that were similar to previous FBA and DBR research (e.g., Iwata et al., 2013; Riley-Tillman et al., 2008) in examining ISP-Function decisional accuracy. Practicing school psychologists examined two sets of blinded data reports pertaining to the same group of 34 students referred for Tier 2 support. The first set of reports summarized ISP-Function data. The second set of reports were more characteristic of a standard FBA, summarizing findings from (1) a semi-structured functional assessment interview with a classroom teacher, and (2) B-C recording data collected via partial-interval SDO within classroom settings. The following research question was posed: to what extent do decisions made in consideration of ISP-Function data agree with decisions made in consideration of standard FBA data? In accordance with prior research (Riley-Tillman et al., 2008), it was hypothesized that school psychologists would derive similar conclusions regarding the function of a student's behavior when examining either data report.

Method

Participants

Nine practicing school psychologists participated in the current study, with an average of 8.3 years of experience. Seven participants were female and two were male. Eight participants were White and one was Black. Participating school psychologists came from six different states, including Arizona, California, Colorado, Connecticut, Maryland, and Wisconsin. Six participants had a doctoral degree and three participants earned a professional degree (e.g., education specialist [Ed.S.]). Each participant was required to be a licensed or credentialed school psychologist as defined by their state department of education. Institutional Review Board approval was obtained for all study procedures.

Data Source

Data reports examined in the current study were generated from a previously published study (Kilgus et al., 2019), which examined the reliability, validity, and accuracy of ISP-Function scores. As noted above, participants in that study included 34 teacher-student dyads from eight elementary (K-5) schools in the Midwest and Southeast. Student participants exhibited moderate levels of disruptive behavior to be addressed via Tier 2 targeted interventions. In Phase 1, research assistants met with teachers to identify the time and setting within which each student's disruptive behavior was most common. Phase 2 then began with research assistants training teacher participants to collect ISP-Function data. Upon training completion, teachers collected ISP-Function data across three occasions within the time and setting identified through Phase 1. Research assistants collected SDO during these same three occasions, thus permitting the direct comparability of ISP-Function and SDO data. During Phase 3, research assistants met with teachers to complete the *Functional Assessment Checklist for*

Teachers and Staff (FACTS; March et al., 2000), a semi-structured functional assessment interview. The order of Phases 2 and 3 were counterbalanced across dyads to remove any order effects in study findings. Both the FACTS and SDO procedure used in this study are described in more detail below (see the Introduction for additional information related to the ISP-Function).

Measures and Materials

Functional Assessment Interview

The FACTS is a semi-structured functional assessment interview that is designed to acquire information regarding student problem behaviors, as well as the antecedents and consequences associated with them. Part A of the FACTS is used to identify a student's strengths, problem behaviors, and the settings within which problem behaviors are most likely to occur. Part B is then used to gather more detailed information regarding the problem behavior, including its topography, frequency, duration, and intensity. Additional information is attained regarding the antecedents of the problem behavior. Specifically, teachers are prompted to consider both potential setting events for the behavior (e.g., conflict at home or academic failure), as well as environmental features that predict its occurrence (e.g., reprimand or difficult tasks). Teachers are then asked to consider which consequences commonly follow behavior, including things that are obtained (e.g., peer attention or preferred activities) and things that are escaped or avoided (e.g., difficult tasks or adult attention). Finally, the interviewer derives a summary of behavior based upon the teacher's responses, which represents a hypothesis regarding the antecedents that occasion and consequences that maintain the problem behavior. The teacher then rates their confidence in the accuracy of this summary using a 6-point scale (1 = "Not very confident" to 6 = "Very confident"). Research has supported FACTS test-retest

reliability (.77), interrater reliability (.50-.88), inter-observer agreement (1.00), convergent validity (e.g., relative to functional analysis), and (e) treatment utility (McIntosh et al., 2008).

Systematic Direct Observation

Systematic Direct Observation (SDO) is commonly considered the gold standard of behavior assessment methods, with a rich psychometric base and history of use within the literature (Chafouleas et al., 2012). SDO is often used as part of comprehensive FBA assessments, inclusive of indirect descriptive tools (e.g., interviews), direct descriptive tools (e.g., SDO), and functional analyses (Shriver et al., 2001). The SDO tool used in the prior study was a behavior-consequence (BC) partial interval recording. An observer recorded whether the student exhibited disruptive behavior at any point during each 20-sec interval for 10 minutes. If disruptive behavior was noted, the observer also recorded whether one or more of the following four consequences occurred following that behavior: adult attention, peer attention, escape/avoidance, and access to items/activities. The operational definitions associated with disruptive behavior and the four consequences were identical to those included on the ISP-Function. Two sets of scores were generated following each observation. First, disruptive behavior scores represented the percentage of intervals that students engaged in disruptive behavior during the observation. Second, consequence scores represented conditional probabilities, defined as percentage of disruptive behavior intervals in which each consequence was observed. A second observer was present for 14% of all observations for the purposes of examining interobserver agreement, defined as the percentage of observed intervals coded identically. The mean IOA across observations was 87% for disruptive behavior (range = 70 – 97%) and 97% for consequences (range = 80–100%).

Data Reports

Two data reports were generated for each student. The first report was considered a standard FBA report, inclusive of FACTS and SDO data. The report began with a description of the FACTS and a table summarizing findings from each section of the teacher interview. Next, a description of the SDO procedure was provided, along with a description of how to interpret resulting scores. Multiple figures and tables were then presented, including (a) a line graph summarizing disruptive behavior scores across the three observations, with *x*-axis = observation sessions and *y*-axis = percentage of intervals scores; (b) a table presenting scores for each of the four consequences across observations; and (c) a bar graph summarizing the mean of scores within each consequence across observations, with *x*-axis = consequences and *y*-axis = mean of consequence scores. No interpretation of scores was provided in the report for any of the aforementioned data, so as to not influence participant data-based judgments regarding the likely function of each student's behavior.

The second report generated for each student summarized ISP-Function data. The report began with a description of the ISP-Function tool. Similar to the aforementioned presentation of SDO data, ISP-Function findings were presented via a series of tables and graphs: (a) a line graph summarizing disruptive behavior ratings across observations, with *x*-axis = observation sessions and *y*-axis = ISP-Function disruptive behavior ratings; (b) a table presenting ISP-Function ratings for each of the four consequences across observations; and (c) a bar graph summarizing the mean of ratings within each consequence across observations, with *x*-axis = consequences and *y*-axis = mean of consequence ratings. Again, no interpretation was provided within the report for the ISP-Function findings. To note, each report was de-identified and assigned a unique code, such that it would be impossible for a participant to determine which FBA and ISP-Function reports corresponded to the same student.

Procedures

The first, second, and fourth authors recruited nine school psychologist participants via their professional networks. Once consent was attained, participants took part in an initial orientation meeting, which was led by the first and second authors and delivered virtually via teleconference software. The brief 30-min orientation began with an overview of the study purpose. Information was then provided regarding the measures administered as part of the prior study, along with information regarding how data resulting from these measures should be interpreted. All participants reported familiarity with these measures; however, the information was still provided to ensure commonality of understanding across participants. Next, trainers reviewed the two sets of reports participants would be receiving and reviewing. Mock reports were shown and information was provided regarding the nature of scores included in the reports. Each report was randomly assigned to each participant.

Trainers then reviewed the Qualtrics electronic survey participants would complete to record their decisions for each report. To complete the survey, participants provided their unique participant number, the code associated with the report they were currently reviewing, and their decision regarding the likely function of the student's problem behavior. In making this decision, participants could select none or all of the following options: adult attention, peer attention, escape/avoidance, access to items/activities, it is unclear, or not applicable: insufficient disruptive behavior was observed. Through the initial training, participants were instructed there were a number of scenarios where the "it is unclear" option would be appropriate, including when (a) a teacher had difficulty identifying relevant consequences during the interview (or rather identified too many consequences), or (b) SDO or ISP-Function data suggested few consequences of disruptive behavior were observed across the observations. Participants were

instructed the “not applicable” option could be appropriate when SDO or ISP-Function data revealed little to no disruptive behavior was observed across the observations, making it impossible to identify common consequences of that behavior. Given their limited frequency, the decision was made to collapse these final two options on the survey into a “no consequence” option for analytic purposes.

Next, participants were randomly assigned to one of three groups, each of which included three participants. Two groups were assigned 11 FBA and 11 ISP-Function reports to review, and the remaining group reviewed 12 student reports. Two of the groups reviewed all FBA reports first, followed by ISP-Function reports. The remaining group followed the opposite order. The order of reports was randomly determined to decrease the likelihood that participants could match one student’s FBA report to their corresponding ISP-Function report. All reports were delivered electronically to participants as portable document format (PDF) files. Once reports were assigned and delivered, participants worked independently to review and derive function-based decisions for each report. Trainers suggested to participants that they take notes while reviewing reports to summarize and synthesize their thinking. After each report was reviewed, participants reviewed their notes and then recorded their decision on the Qualtrics survey before moving on to the next report. Participants were given two weeks to review all reports and submit their decisions.

Once all decisions were recorded, the research team downloaded all data from Qualtrics and summarized agreement for each report across the three participants who evaluated the report. Each group then reconvened in a one-hour follow-up meeting held via teleconference software. During this meeting, the second author reviewed participant decisions for each report, one report at a time. Participants were given the opportunity to provide a rationale for their decision

founded upon their consideration of the data in that report. Once all participants had a chance to speak and the conversation reached a natural conclusion, participants were given the opportunity to revise their decision for that report. Across all reports reviewed, 29% of decisions were revised during this final stage. This approach was founded upon consensus building procedures that have been used in the behavior assessment literature (Jaffery et al., 2015). It was also intended to reflect the FBA decision making process commonly used in schools, wherein teams convene to evaluate data to inform function- and intervention-related decisions (Scott et al., 2005). The meeting then concluded once all reports had been evaluated. Participants were then provided a small monetary compensation for their time and effort.

Data Analysis Plan

For each ISP-Function and FBA report, scores were calculated for each consequence such that the score represented the number of raters who identified the consequence as a likely function of a student's behavior given the data included in the report. Scores ranged between 0 and 3, as three school psychologists reviewed each individual student report. A score of 3 suggested perfect agreement among the school psychologists regarding the consequence being a likely function of a student's behavior. A score of 0 also indicated perfect agreement, with all three school psychologists indicating a consequence was likely not a function of a student's behavior. A score of 1 or 2 indicates some degree of disagreement among the three school psychologists. This method of score calculation was used because it provides a spectrum of certainty that the function is present in any case. If the score is 0, there is a relatively high level of certainty that the function is not present, as scores of 1 or 2 show weak or moderate evidence of the presence of the function, respectively. A score of 3 demonstrates a high level of

confidence that the function is likely impactful in the case. Descriptive statistics are shown in Table 1.

The accuracy of ISP-Function decisions relative to FBA-based decisions was evaluated using Gwet's agreement coefficient (AC) and associated 95% confidence intervals and *p*-values. Like the more common Cohen's kappa, Gwet's AC represents the proportion of agreement between two measures corrected for chance. However, Gwet's agreement coefficient represents an alternative to kappa that is particularly appropriate when behavioral base rates are low, as kappa is biased in this condition (Viera & Garrett, 2005). Gwet's AC may provide a more accurate representation of agreement for the current data, as there were several scores of 0 across ISP-Function and FBA reports. AC values range between 0 and 1, with higher values representing greater agreement. Magnitude of agreement was interpreted using the Interval Membership Probability (IMP) method applied to the benchmark intervals outlined in previous research (Gwet, 2014). Per the Landis-Koch benchmark intervals, an AC value less than .00 is considered Poor, .00 –.20 Slight, .21 –.40 Fair, .41 –.60 Moderate, .61 –.80 Substantial, and ≥ .80 Almost Perfect (Klein, 2018; Landis & Koch, 1977).

Rather than using traditional null hypothesis testing such that measure independence is rejected if the *p*-value associated with each AC statistic is less than .05, we acknowledge that the AC statistic is a point estimate with a margin of error. To reduce the possibility of artificially inflating the agreement by only reporting point estimates and associated benchmarks, we use a probabilistic approach to identifying the benchmark membership of each agreement statistic. To take this approach, we use the IMP procedure recommended by Gwet (2014), which first calls for computation of the AC statistic and associated standard error. Then, we used the irrCAC statistical package in R to calculate the cumulative probability of the AC value falling within

each benchmark interval. This process includes three steps: (a) computing the probability for the coefficient to fall into each benchmark interval, (b) computing the cumulative probability that a coefficient would occur if the true score was at a particular benchmark, and (c) selecting the benchmark interval that has a cumulative probability of 95% or higher, in following with the convention for maintaining a Type I error rate (α) of 5% or lower. The benchmark intervals reported below represent the interval in which there is a 95% or higher probability that the true AC value is within that interval (e.g., Fair) or those below it (e.g., Slight or Poor).

Results

Based on ISP-Function scores, 61.7% (*n* = 21; groups ranged between 5-8 each) of vignettes were rated as maintained by adult attention, 44.1% (*n* = 15; groups 4-6) were rated as peer attention, 23.5% (n = 8; groups 1-5) as escape/avoidance, 2.9% (*n* = 1; groups 0-1) as access to items and activities. The sum of ratings does not add to 100% because raters agreed that 23.5% (*n* = 8; groups 2-3) vignettes had no clear consequence and 47.1% (*n* = 16; groups 3-8) vignettes were rated as having more than one function. Based on FBA reports, 76.5% (*n* = 26; groups 7-10) were rated as adult attention, 61.7% (*n* = 21; groups 5-9) were rated as peer attention, 11.7% (*n* = 4; groups 0-3) escape/avoidance, 2.9% (*n* = 1; groups 0-1) as access to items and activities, 11.7% (*n* = 4; groups 0-3) were rated as no clear consequence, and 58.8% (*n* = 20; groups 4-9) reports were rated as having more than one function. The most likely functions to co-occur across both ISP-Function and SDO were adult and peer attention.

Analyses were intended to address a single research question regarding the extent to which decisions made in consideration of ISP-Function data agree with decisions made in consideration of standard FBA data. The deterministic point estimates of Gwet's AC are presented in Table 2. All AC values were statistically significant per traditional hypothesis

testing. The AC point estimate fell in the Almost Perfect interval for the access to items and activities function (AC = .940 [.855–1.00], p <.001). For peer attention (AC = .679 [.490–.868], p <.001), escape/avoidance (AC = .643 [.441–.844], p <.001), and no consequence (AC = .671 [.479–.862], p <.001), the AC value represented Substantial agreement. Finally, the AC value for adult attention fell in the Moderate agreement interval (AC = .502 [.284–.720], p <.001).

When the IMP strategy was employed, we found that there was a 99.7% chance that the AC statistic for adult attention fell in the Fair interval or lower. Therefore, the agreement between the ISP-Function and the FBA Reports is at least Fair. For peer attention, there was a 99.8% chance that the AC value was in the Moderate range or those below it. There was a 99.3% probability that the true value for escape/avoidance falls in the Moderate interval or those below it. There was a 99.9% chance that access to items and activities fell in the Almost Perfect interval. For no clear consequence, there was a 99.8% probability that the AC value is in the Moderate interval or below. The full results, including the cumulative probabilities for each benchmark interval, are included in the Supplemental Table 1.

Discussion

The purpose of this study was to examine ISP-Function decisional accuracy; more specifically, the extent to which school psychologists reached similar conclusions when examining reports specific to ISP-Function data or more traditional FBA data. A series of functional assessment measures were completed for 34 elementary students who commonly exhibit disruptive behavior, including data from the FACTS, SDO, and the ISP-Function. Data from these measures were summarized into two reports for each student, including one report that summarized ISP-Function findings, and another report that summarized FACTS and SDO findings in accordance with a more standard FBA report. Each of three teams of school

psychologists reviewed the reports for 11 or 12 participating students. Reports were blinded, such that school psychologists could not determine which two reports corresponded to the same student. Data resulting from this review indicated the number of school psychologists who identified each particular consequence as a function when reviewing each report. Analyses then examined the extent to which a similar number of school psychologists identified a function for the same students across the two reports.

The Gwet's AC statistic, with its sensitivity to low behavioral base rates, offered a promising depiction of the correspondence between reports, as all coefficients were statistically significant; all values fell in the Moderate to Almost Perfect intervals based on point estimates. Specifically, values for adult attention fell in the Moderate interval; in the Substantial interval for peer attention, escape/avoidance, and no clear consequence; and in the Almost Perfect interval for access to items and activities. When using the IMP method, there was a 95% or higher chance that each agreement statistic fell in the Fair range or below. With the relatively conservative $\alpha = .05$ criterion, Adult Attention fell in the Fair interval. Peer attention, escape/avoidance, and no clear consequence fell in the Moderate interval. Finally, access to items and activities fell in the Almost Perfect range.

Further consideration of results across consequences reveals an interesting pattern of findings. There was consistent support for the alignment of decisions between ISP-Function and FBA reports on adult attention and peer attention. Specifically, when reviewing either report, school psychologists frequently agreed that these two consequences were likely functions of student disruptive behavior. This finding aligns with previous studies, which have documented stronger psychometric support for these ISP-Function targets (e.g., Kilgus et al., 2017, 2019). This finding also aligns with what one might expect of attention-based consequences. Our

experience suggests adult and peer attention are common consequences of problem behavior at the elementary level. Adult and peer attention are also frequently present in a conspicuous manner that is readily apparent to others (e.g., high fives, laughter, conversation). Accordingly, one could expect many teachers to have substantial experience with adult and peer attention that is contingent upon problem behavior. Furthermore, teachers are likely to possess the capacity to recognize such activity when it is indeed present using the ISP-Function or other similar tools. Additionally, school psychologists frequently agreed that escape/avoidance and access to items/activities were not likely functions of student disruptive behavior. In addition, they were frequently in agreement that at least some function was present, as the "no clear consequence" option was frequently ruled out. With that said, given the lack of students exhibiting disruptive behavior maintained by either escape/avoidance or access to items/activities within the reviewed reports, coupled with only three data points collected for each student, it is difficult to determine how accurate school psychologists were in evaluating the presence or absence of such behavior.

Limitations

Multiple limitations to this study should be noted. First, a limited number of school psychologists reviewed each report. Though unlikely to impact the power of our analyses, adding additional school psychologists would nevertheless have provided a more robust and generalizable test of ISP-Function decisional accuracy. Second, the majority of students in the study were exhibiting attention-maintained behavior. Criterion FBA findings confirmed that few students were exhibiting disruptive behavior that was maintained by either escape/avoidance or access to items/activities. Future research with larger and more diverse student samples is needed to evaluate school psychologists' accuracy in evaluating a wider range of behavioral functions. Such research should also employ procedures consistent with recommendations from prior ISP-

Function research, collecting a larger number of ISP-Function data points per student (e.g., 8-18; Kilgus et al., 2019).

Third, though characteristic of FBAs commonly conducted within school settings, the FBAs conducted for the purpose of this investigation were founded upon "correlational" methods alone (i.e., functional assessment interviews and SDO). Although these methods can afford information about the extent to which behaviors and consequences are related, they cannot support more causal claims regarding this relationship (Carr, 1994). Such information would only be derived via the use of experimental methods, including functional analysis (Gresham et al., 2001). Future research should examine functional analyses to increase the confidence in criterion decisions against which ISP-Function data are compared.

Fourth, the current findings are specific to ISP-Function performance in evaluating the moderately intense disruptive behaviors typically exhibited by students supported at Tier 2, as well as a limited number of consequences associated with them. Additional research is needed to determine how the ISP-Function performs in evaluating alternative problem behaviors and a wider range of consequences associated with them. Furthermore, while this study examined the accuracy of school psychologist decisions regarding behavioral function, additional research should examine the accuracy of decisions regarding the behavior itself, as these decisions are important aspects of the FBA process. Such studies might have school psychologists derive conclusions regarding the intensity of the behavior and the appropriateness of Tier 2 supports, as well as whether the stability and level of problem behavior is sufficient to support conclusions regarding the functions of that behavior. Fifth, the ISP-Function is somewhat limited in that it affords findings specific to problem behavior and its consequences. The tool does not yield other information commonly found in FBAs, including that related the antecedents of problem

behavior. Moving forward, researchers might examine whether this flexible tool might be modified and expanded to collect additional forms of FBA-related information.

Implications for Practice

Findings from this study suggested that when reviewing ISP-Function data, school psychologists are likely to draw conclusions that are similar to those they would draw when considering more traditional FBA data. When considered in relation to previous ISP-Function studies (Kilgus et al., 2017, 2019), the current findings provide initial evidence for the use of the ISP-Function within low-stakes decisions at Tier 2. Of course, future research remains necessary to strengthen this support and increase educator confidence regarding the defensibility of ISP-Function data and conclusions that would result from their use. Pending additional evidence, the manner in which the ISP-Function might eventually be used could take two forms (Kuchle et al., 2015; Majeika et al., 2020).

First, ISP-Function data could be collected once a student has been identified as requiring support but prior to selecting an appropriate intervention. Specifically, data would be collected within the same setting across multiple days to yield a reliable stream of data. Data would then be graphed and interpreted to yield conclusions regarding the likely function of a student's behavior. These findings could be used to identify an intervention aligned with the function of the student's behavior. Second, ISP-Function data could also be collected for students who are unresponsive to an initial course of standard protocol intervention (e.g., Check In/Check Out) that is delivered in a similar manner for all students receiving Tier 2 support. ISP-Function data would then be used to identify an alternative intervention that is more aligned with the function of a student's behavior, using a more differentiated approach to delivering Tier 2 interventions. This latter approach is likely to be more efficient and feasible for many schools, as it limits the

collection of functional assessment data to only a subset of students referred for Tier 2 intervention.

Conclusion

Whichever approach schools choose to take, ISP-Function data could provide schools with beneficial information regarding the function of a student's behavior to guide intervention selection. Results demonstrate fair to almost perfect correspondence between the ISP-Function and traditional FBA reports. As school psychologists and behavior consultants are often tasked with collecting FBA data for children demonstrating disruptive behavior, the current study demonstrates how the ISP-Function could serve as a more efficient tool regarding the function of student behavior.

References

Barnhill, G. P. (2005). Functional behavioral assessment in schools. *Intervention in School and Clinic*, *40*(3), 131-143. https://doi.org/10.1177/10534512050400030101

Boyd, R. J., & Anderson, C. M. (2013). Breaks are better: A tier II social behavior intervention. *Journal of Behavioral Education*, *22*(4), 348-365. https://doi.org/10.1007/s10864-013-9184-2

Bruni, T. P., Drevon, D., Hixson, M., Wyse, R., Corcoran, S., & Fursa, S. (2017). The effect of functional behavior assessment on school-based interventions: A meta-analysis of single-case research. *Psychology in the Schools*, *54*(4), 351-369. https://doi.org/10.1002/pits.22007

Campbell, A., & Anderson, C. M. (2008). Enhancing effects of check-in/check-out with function-based support. *Behavioral Disorders*, *33*(4), 233-245. https://doi.org/10.1177/019874290803300404

Carr, E. G. (1994). Emerging themes in the functional analysis of problem behavior. *Journal of Applied Behavior Analysis*, *27*(2), 393-399. https://doi.org/10.1901/jaba.1994.27-393

Chafouleas, S. M. (2011). Direct Behavior Rating: A review of the issues and research in its development. *Education and Treatment of Children*, *34*(4), 575-591.

Chafouleas, S. M., Kilgus, S. P., Riley-Tillman, T. C., Jaffery, R., & Harrison, S. (2012). Preliminary evaluation of various training components on accuracy of Direct Behavior Ratings. *Journal of School Psychology*, *50*(3), 317-334. https://doi.org/10.1016/j.jsp.2011.11.007

Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *The Journal of Special Education, 42*(2), 108-126.

Christ, T. J., Nelson, P. M., Van Norman, E. R., Chafouleas, S. M., & Riley-Tillman, T. C. (2014). Direct Behavior Rating: An evaluation of time-series interpretations as consequential validity. *School Psychology Quarterly*, *29*(2), 157-170. https://psycnet.apa.org/doi/10.1037/spq0000029

Dunlap, G., & Kern, L. (2018). Perspectives on functional (behavioral) assessment. *Behavioral Disorders, 43*(2), 316–321. https://doi.org/10.1177/0198742917746633

Durand, V. M., & Crimmins, D. B. (1988). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders*, *18*(1), 99-117. https://doi.org/10.1007/BF02211821

Eckert, T. L., Martens, B. K., & DiGennaro, F. D. (2005). Describing antecedent-behavior-consequence relations using conditional probabilities and the general operant contingency space: A preliminary investigation. *School Psychology Review*, *34*(4), 520-528.

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders*, *37*(2), 55-77. https://doi.org/10.1177%2F019874291203700201

Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review*, *30*(2), 156-172. https://doi.org/10.1080/02796015.2001.12086106

Gwet, K. L. (2014) *Handbook of inter-rater reliability* (4th ed.). Advanced Analytics, LLC.

Iwata, B. A., & DeLeon, I. G. (1995). *Functional assessment screening tool* (FAST). University of Florida.

Iwata, B. A., DeLeon, I. G., & Roscoe, E. M. (2013). Reliability and validity of the functional analysis screening tool. *Journal of Applied Behavior Analysis*, *46*(1), 271-284. https://doi.org/10.1002/jaba.31

Jaffery, R., Johnson, A. H., Bowler, M. C., Riley-Tillman, T. C., Chafouleas, S. M., & Harrison, S. E. (2015). Using consensus building procedures with expert raters to establish comparison scores of behavior for direct behavior rating. *Assessment for Effective Intervention*, *40*(4), 195-204. https://doi.org/10.1177%2F1534508415569527

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kilgus, S. P., Fallon, L. M., & Feinberg, A. B. (2016). Function-based modification of check-in/check-out to influence escape-maintained behavior. *Journal of Applied School Psychology*, *32*(1), 24-45. https://doi.org/10.1080/15377903.2015.1084965

Kilgus, S. P., Kazmerski, J. S., Taylor, C. N., & von der Embse, N. P. (2017). Use of direct behavior ratings to collect functional assessment data. *School Psychology Quarterly*, *32*(2), 240-253. https://psycnet.apa.org/doi/10.1037/spq0000156

Kilgus, S. P., von der Embse, N. P., Eklund, K., Izumi, J., Peet, C., Meyer, L., & Taylor, C. N. (2019). Reliability, validity, and accuracy of the Intervention Selection Profile-Function: A brief functional assessment tool. *School Psychology*, *34*(5), 531-540. https://doi/org/10.1037/spq0000325

Klein, D. (2018). Implementing a general framework for assessing inter-rater agreement in Stata. *The Stata Journal, 18,* 871–901. http://doi.org/10.1177/1536867X1801800408

Kuchle, L., Zumeta Edmonds, R., Danielson, L. C., Peterson, A., & Riley-Tillman, T. C. (2015). The next big idea: A framework for integrated academic and behavioral intensive intervention. *Learning Disabilities Research & Practice*, *30*(4), 150-158. https://doi.org/10.1111/ldrp.12084

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174. https://doi.org/10.2307/2529310

Majeika, C. E., Bruhn, A. L., Sterrett, B. I., & McDaniel, S. (2020). Reengineering Tier 2 interventions for responsive decision making: An adaptive intervention process. *Journal of Applied School Psychology, 36*(2), 111-132. https://doi.org/10.1080/15377903.2020.1714855

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)*. Educational and Community Supports.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

McIntosh, K., Borgmeier, C., Anderson, C. M., Horner, R. H., Rodriguez, B. J., & Tobin, T. J. (2008). Technical adequacy of the Functional Assessment Checklist: Teachers and Staff (FACTS) FBA interview measure. *Journal of Positive Behavior Interventions, 10*(1), 33-45. https://doi.org/10.1177/1098300707311619

National Association of School Psychologists (2020). *The professional standards of the National Association of School Psychologists.* www.nasponline.org

O'Neill, G. W., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K, & Newton, J. S. (1997). *Functional assessment and program development for problem behavior: A practical handbook.* Brooks/Cole.

O'Neill, R. E., Albin, R. W., Store, K., Horner, R. H., & Sprague, J. R. (2015). *Functional assessment and program development for problem behavior: A practical handbook* (3rd ed.). Cengage Learning.

Reinke, W. M., Stormont, M., Clare, A., Latimore, T., & Herman, K. C. (2013). Differentiating tier 2 social behavioral interventions according to function of behavior. *Journal of Applied School Psychology*, *29*(2), 148-166.

Riley-Tillman, T. C., Chafouleas, S. M., Briesch, A. M., & Eckert, T. L. (2008). Daily behavior report cards and systematic direct observation: An investigation of the acceptability, reported training and use, and decision reliability among school psychologists. *Journal of Behavioral Education, 17*(4), 313-327. https://doi.org/10.1007/s10864-008-9070-5

Scott, T. M., Liaupsin, C., Nelson, C. M., & McIntyre, J. (2005). Team-based functional behavior assessment as a proactive public school process: A descriptive analysis of current barriers. *Journal of Behavioral Education*, *14*(1), 57-71. https://doi.org/10.1007/s10864-005-0961-4

Shriver, M. D., Anderson, C. M., & Proctor, B. (2001). Evaluating the validity of functional behavior assessment. *School Psychology Review*, *30*(2), 180-192.

Steege, M. W., Pratt, J. L., Wickerd, G., Guare, R., & Watson, T. S. (2019). *Conducting school-based functional behavioral assessments: A practitioner's guide.* Guilford Publications.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363.

Table 1

Descriptive Statistics of the Identified Function of Behavior by Method

Consequence	Report	*M*	*SD*	Skew	Kurtosis
AA	FBA	2.21	1.3	-1.02	-0.91
	ISP	1.68	1.45	-0.20	-1.95
PA	FBA	1.71	1.47	-0.28	-1.93
	ISP	1.41	1.52	0.11	-2.04
EA	FBA	0.21	0.59	3.34	12.01
	ISP	0.68	1.25	1.24	-0.43
AIA	FBA	0.06	0.34	5.33	27.17
	ISP	0.09	0.51	5.33	27.17
NA	FBA	0.32	0.91	2.37	3.90
	ISP	0.76	1.3	1.06	-0.86

Note. FBA = Functional Behavior Assessment, ISP = ISP-Function; AA = adult attention, PA = peer attention, EA = escape/avoidance, and AIA = access to items and activities, NA = No clear consequence; *M* = mean score for a given function taken as the average of all vignette scores for that function with a range of 0-3 (0 represents no experts identifying the function for a given vignette, 3 represents all experts identifying the function for a given vignette)

Table 2

ISP-Function Decisional Accuracy Relative to FBA Findings using Gwet's AC1

Target	Point Estimate Interval	Coefficient	Standard error	*p*	95% CI		IMP Interval	IMP Cumulative Probability
					Lower Bound	Upper Bound		
Adult Attention	Moderate	0.502	0.107	<.001*	0.284	0.720	Fair	.997
Peer Attention	Substantial	0.679	0.093	<.001*	0.490	0.868	Moderate	.998
Escape / Avoidance	Substantial	0.643	0.099	<.001*	0.441	0.844	Moderate	.993
Access to Items and Activities	Almost Perfect	0.940	0.042	<.001*	0.855	1.00	Almost Perfect	.999
No clear Consequence	Substantial	0.671	0.094	<.001*	0.479	0.862	Moderate	.998

Note. IMP = Interval Membership Probability

*p < .05