Citation: Volpe, R. J., Chaffee, R. K.,Yeung, T. S., & Briesch, A. M. (2019). *School Mental Health*. https://doi.org/10.1007/s12310-019-09338-w

Initial Development of Multi-item Direct Behavior Rating Measures of Academic Enablers

Robert J. Volpe, Ph.D.

r.volpe@northeastern.edu

Ruth K. Chaffee, Ph.D. chaffee.r@husky.neu.edu

Tat Shing Yeung, MS, MA yeung.ta@husky.neu.edu

Amy M. Briesch, Ph.D. a.briesch@northeastern.edu

Northeastern University

Preparation of this article was supported through funding by the Institute for Education Sciences, U.S. Department of Education (R324A150071). Opinions expressed herein do not necessarily reflect the position of the U.S. Department of Education, and such endorsements should not be inferred. Correspondence concerning this article should be addressed to Robert J. Volpe at 404 International Village, Northeastern University, Boston MA 02115; email: r.volpe@northeastern.edu

Abstract

Although there is a growing body of evidence to support the use of direct behavior rating (DBR) as a formative behavioral assessment method in school-based problem-solving models, this work has centered largely on the assessment of attending and problem behaviors and on the use of single-item DBR scales. The primary purpose of the current study was to report on the development and initial validation of teacher-completed multi-item DBR scales (DBR-MIS) designed to assess three constructs that represent behaviors that are widely considered to support student academic achievement (Academic Engagement, Interpersonal Skills, and Study Skills). Development of the scales involved a content validity study utilizing feedback from a panel of consumers (teachers, parents, and school administrators) and a panel of researchers with expertise in school-based behavioral assessment. Results of an exploratory factor analysis of ratings completed by teachers in Grades K through 3 ($N = 307$) supported a one-factor solution for each of the aforementioned constructs. Favorable internal consistency was found for each scale.

Content Validation of Direct Behavior Rating Multi-Item Academic Enabler Scales Using Exploratory Factor Analysis

Over the past several decades school-based assessment has shifted from a primarily reactive and summative orientation to a proactive, problem-solving framework in which early identification (screening) and repeated formative measurement (progress monitoring) are emphasized (Tilly, 2008). This transition has been spurred in part by changes in federal policy, such as the reauthorization of the Individuals with Disabilities Act (IDEA; 2004) and more recently the Every Student Succeeds Act (ESSA; 2015). In recent years the percentage of school districts reporting implementation of multi-tiered systems of support (MTSS) for students with social-emotional and behavioral (SEB) problems has risen dramatically (Spectrum K12 School Solutions, 2011; Zirkel, 2011). In addition to effective screening assessments for early identification, valid and feasible progress monitoring assessments are a vital component of effective MTSS (National Center on Response to Intervention, 2010; Sugai & Horner, 2009). On both systems- and individual- levels, progress monitoring assessments allow school professionals to evaluate the efficacy of interventions and enable data-based decision making concerning supports for specific students.

The Need for Formative Measures of Student Social Behavior

Briesch and Volpe (2007) provide a conceptual framework for evaluating progress monitoring assessments that balances considerations of psychometric adequacy and feasibility. Ideally, such formative measures should accurately measure the intended behavior or construct, be sensitive to changes in student behavior, and be feasible for school professionals to implement in terms of training, time, and cost. Unfortunately, the current knowledge base regarding behavioral progress monitoring measures is limited relative to those available for academic

progress monitoring (Chafouleas, Volpe, Gresham, & Cook, 2010). Whereas the National Center on Intensive Intervention (NCII; n.d.) Academic Progress Monitoring Tools Chart summarizes the psychometric evidence for 25 tools with multiple subtests assessing various academic skills and grade levels, only six tools are listed in the Behavioral Progress Monitoring Tools Chart. The chart includes two behavior rating scales (Behavior Intervention Monitoring Assessment System [BIMAS; McDougal, Bardos, & Meier, 2011] and the Behavioral Assessment System for Children-2 Progress Monitor [BASC-2 Progress Monitor; Reynolds & Kamphaus, 2009]), observational assessment of academic engagement (i.e., momentary time sampling for academic engagement) and two Direct Behavior Rating (DBR) measures (i.e., academically engaged, disruptive behavior). Unfortunately, each of these currently available progress monitoring tools demonstrates limitations in regard to either feasibility, psychometric adequacy, or both.

Although rating scales are one of the most commonly used tools by school psychologists (Shapiro & Heick, 2004), most have been constructed for use in diagnostic assessments and as such were designed to assess symptoms of common child psychiatric disorders as opposed to socially valid and malleable indicators of student academic and social functioning (Pelham, Fabiano, & Massetti, 2005). Rating scales developed primarily for such summative assessments may require extensive time for completion, are restricted in regard to the range of behavioral constructs assessed, and may be inappropriate for measuring short-term changes in student behavior in response to classroom intervention (Volpe & Gadow, 2010). Finally, none of the rating scales identified by NCII provide “convincing” criterion-related validity and treatment sensitivity evidence. Without evidence supporting desirable psychometric characteristics or feasible implementation, extant rating scales would seem to be of limited utility for progress monitoring purposes in MTSS models.

Another method identified by NCII is observational assessment of academic engagement via momentary time sampling (MTS). Evidence indicates that MTS generates reliable measurement of student engagement behavior (Briesch, Chafouleas, & Riley-Tillman, 2010; Hintze & Matthews, 2004; Wood, Hojnoski, Laracy, & Olson, 2015; Volpe, DiPerna, Hintze, & Shapiro, 2005). However, to achieve acceptable levels of dependability for progress monitoring (e.g., .70) requires trained observers to conduct two 30-minute, three 15-minute, or four to five 10-minute observations on each targeted student (Briesch, Volpe, & Ferguson, 2014; Ferguson, Briesch, Volpe, & Daniels, 2012). Furthermore, MTS data are highly context-specific, which can limit the generalizability of resultant data (Chafouleas et al., 2010). Although MTS for academic engagement demonstrates favorable psychometric characteristics, the time required for training accurate observers and the time needed to conduct observations limit its utility as a progress monitoring measure in MTSS models (Volpe & Gadow, 2010; Volpe, Gadow, Blom-Hoffman, & Feinberg, 2009).

A third class of progress monitoring assessment identified by NCII, DBR, has been described as a hybrid tool combining elements of systematic direct observation with those of behavior rating scales (Chafouleas, 2011). DBR typically are completed by classroom teachers, and unlike systematic direct observation do not require trained external observers. DBR may minimize inference and retrospective judgments by directly rating observable, operationally defined behaviors in close temporal proximity and in naturalistic settings (Chafouleas, Riley-Tillman, & Christ, 2009; Chafouleas, Riley-Tillman, & Sugai, 2007; Christ, Riley-Tillman, & Chafouleas, 2009). Two primary methods of DBR have been investigated. Single-Item Scales (DBR-SIS) typically measure a global construct (e.g., academic engagement) using a single item. In contrast,Multi-Item Scales (DBR-MIS) are comprised of several items (typically between

three and five items) that assess specific behaviors, the scores of which can be summed to obtain a composite score measuring the construct of interest (e.g., Volpe & Briesch, 2012). The majority of DBR research has focused on DBR-SI) ratings of Academic Engagement, Respectful, and Disruptive Behavior (Briesch, Chafouleas, & Riley-Tillman, 2016; Chafouleas, 2011). Studies have demonstrated the reliability and dependability (Chafouleas et al., 2010; Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007), validity (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008), and treatment sensitivity (Chafouleas, Sanetti, Kilgus, & Maggin, 2012; Fabiano, Pyle, Kelty, & Parham, 2017) of DBR-SIS assessing Academic Engagement and Disruptive Behavior.

The use of DBR-SIS of Academic Engagement and Disruptive for progress monitoring assessment offers many advantages (e.g., validity, reliabilitiy, feasibility); however the restriction of SEB progress monitoring assessment to only two constructs is very narrow. Although the relationship between academically engaged and disruptive student behavior and academic achievement is well documented (e.g., Atkins, Hoagwood, Kutash, & Seidman, 2010; Reyes, Brackett, Rivers, White, & Salovey, 2012), a broader set of constructs have been identified as having a positive impact on academic achievement (e.g., DiPerna & Elliott, 1999) and represent appropriate targets for classroom interventions (e.g., DiPerna & Elliott, 2002; Volpe et al., 2006). Thus, an expanded set of progress monitoring tools that assess a wider range of such positive behaviors is needed.

Academic Enablers

In the current study, we report the intial development of DBR-MIS to assess an expanded set of academic enabler constructs. As defined by DiPerna and Elliott (1999), academic competence is “a multi-dimensional construct composed of the skills, attitudes, and behaviors of

a learner that contribute to teachers' judgments of academic performance" (p. 208). Academic skills alone are therefore insufficient for academic success; rather, students must also demonstrate adequate levels of engagement and possess adequate interpersonal skills, study skills, and academic motivation (DiPerna et al., 2002). Academic engagement reflects the degree to which a student sustains attention and participates in the curriculum. Although the construct encompasses more than simply attending behavior, the relationship between inattention and academic outcomes is clear (e.g., Greenwood, 1996; Massetti et al., 2008), even when controlling for cognitive and other behavioral variables (Rabiner & Coie, 2000).

The construct of interpersonal skills encompasses the ability to effectively communicate and cooperate with both peers and adults. Interpersonal skills have been shown to have significant direct and indirect effects on student achievement (Henricsson & Rydell, 2006; Malecki & Elliott, 2002; Wentzel, 1998). According to the models proposed by DiPerna and Elliott (2002), student interpersonal skills interact with prior achievement and predict motivation, which in turn has significant effects on two direct predictors of current achievement: study skills and engagement.

Finally, study skills refer to behaviors that facilitate the process of learning new information, including taking notes, checking work, following directions, and organizational behaviors such as keeping one's materials/desk organized and remembering materials and assignments (DiPerna & Elliott, 2000). Investigations have shown that study skills are a significant, direct predictor of academic achievement across grade levels and are common targets for academic intervention (Gettinger & Seibert, 2002; Harvey & Chickie-Wolfe, 2007; Jenkins & Demaray, 2015; Volpe, DuPaul, DiPerna, & Jitendra, 2006).

Several longitudinal studies have demonstrated the importance of academic enablers in predicting academic achievement for children and adolescents (e.g., DiPerna & Elliott, 1999; DiPerna, Volpe, & Elliott, 2002; DiPerna, Volpe, & Elliott, 2005; Reynolds & Walberg, 1992; Volpe et al., 2006). Expanding formative behavioral assessment to include academic enablers is also critical as such behaviors are frequent targets of school-based interventions for students with learning and behavior problems (e.g., Barkley, 2016; Fabiano et al., 2010).

Purpose of Study

Heretofore, DBR assessment of academic enablers has been limited to the construct of academic enagement (e.g., Chafouleas et al., 2010; Volpe & Briesch, 2012, 2015). Given than academic enablers are both critical to academic success and appropriate targets of classroom intervention, we sought to expand the breadth of formative assessment of academic enablers. The purpose of current study was to develop and evaluate teacher-completed DBR-MIS scales to assess the academic enabler constructs of Academic Engagement, Interpersonal Skills, and Study Skills in early elementary school students. First, the content validity of an initial pool of candidate items for each scale was evaluated by a panel of consumers (e.g., school professionals and parents) and a panel of researchers with expertise in school-based behavioral assessment. The construct validity of the three proposed scales was then subsequently evaluated through exploratory factor analysis (EFA). Finally, we examined the internal consistency of each novel scale.

Method

Item Development

Item content for the Academic Engagement, Interpersonal Skills, and Study Skills scales was developed through a four-stage process. In the first stage, items were drafted based on three

sources of information: a) a national teacher survey of common referral concerns (Briesch et al., 2013), b) a review of extant rating scale and observation measures assessing the constructs of interest, and c) a review of frequently used targets for Daily Report Card interventions (Owens et al., 2012).

In the second stage, a Consumer Advisory Panel (CAP) consisting of four kindergarten through grade 3 (K-3) general education classroom teachers, one special education teacher, one elementary school principal, two school psychologists, and four parents of children in K-3 provided feedback on the aforementioned items. Members of the CAP were sent a link to provide feedback on items using Qualtrics software (Qualtrics, Provo, UT). Members of the CAP rated each candidate item using a 5-point Likert-type scale (0 = *Strongly Disagree*; 4 = *Strongly Agree*) on three criteria: (a) observability (they would be able to see it in a typical classroom setting), (b) malleability (it would be a suitable target for classroom intervention), and (c) social validity (if it improved, it would be helpful to the student in question or the classroom environment at large). Members of the CAP also provided feedback regarding clarity in the wording of items and were asked to list any additional behavioral targets that should be considered in the assessment of the constucts of interest. Both qualitative and quantitative feedback from the CAP was used to inform modifications to the composition of existing items before moving on to the next stage of development.

In the third stage, a Scientific Advisory Panel (SAP) comprised of five scholars with expertise in school-based behavioral assessment and scale development, the constructs of interest, and statistical methods reviewed and rated the items on a 5-point Likert scale (0 = *Strongly Disagree*; 4 = *Strongly Agree*) according to the following criteria: (a) construct

relatedness, (b) observability, (c) sensitivity to treatment, and (d) social validity. The SAP also provided feedback on wording of items, content coverage and any other suggestions for revision.

In the fourth stage, resultant items were administered to a large sample of elementary school teachers. Teacher ratings for each item were submitted to separate EFA for each scale in order to empirically identify items most representative of each of the three academic enabler constructs. Initial item retention decisions were based on the results of the EFA and previous ratings from the CAP and SAP. Subsequently, the SAP conducted a final review to determine which items should be retained for each scale.

Exploratory Factor Analysis

Participants. Participants were general- and special-education teachers from 35 school public school districts across 13 states. A total of 307 K-3 teachers each completed ratings for one randomly selected student in their class. Teachers were primarily female (95.8%) with a wide range of teaching experience (see Table 1). The student sample was comprised of 187 (60.9%) males and 120 (39.1%) females, and students were nearly evenly distributed across grade levels (see Table 2). The composition of students by race and ethnicity was as follows: 67.1% White, 13% Black, 15.0% Hispanic, 3.3% Asian, 1.0% American/Alaska Native, and 5.2% Unknown. Approximately 35% of students were receiving special education services at the time of data collection.

Measures. Teachers were instructed to rate the behavior of target students over the past 5 school days using a 7-point frequency scale ranging from *Never = 0* to *Almost Always = 6*. Items for each academic enabler scale were developed through the 4-stage process described above.

Data collection procedures. Each participating K-3 teacher rated one randomly identified student from his/her class. Teachers were assigned a randomly generated number

between 1 and 20 and selected the corresponding student from his/her alphabetical class roster. The random selection of the student was designed to prevent any score variance restriction. Researchers created and adhered to a standardized protocol to ensure similar procedures during each administration. Participants were sent a link to an online portal to complete the ratings, which the authors created using Qualtrics software (Qualtrics, Provo, UT). This format prevented missing data by requiring all items to be rated for submission. Participants completed ratings for only one student to ensure the independence of ratings.

Data analysis. Exploratory factor analysis is inherently theory-driven and can be used as a confirmatory approach (McDonald, 1999). As such, EFA was performed separately for each of the three scales to identify items that would serve as strongest indicators of the relevant latent construct identified *a priori.* As items were developed to measure their corresponding construct, all items within each DBR-MIS scale were expected to load substantially on a single factor. Principal axis factoring (PAF) was used as the factor extraction method because it does not assume multivariate normality (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Parallel analysis (PA; Horn, 1965) was conducted to inform factor retention decisions within each scale. PA permutated multiple correlation matrices of raw data and computed eigenvalues. A factor with an eigenvalue higher than the 95th percentile of all eigenvalues was retained (Velicer, Eaton, & Fava, 2000). In cases where more than one factor was identified within a scale, direct oblimin, an oblique rotation method, was adopted due to the expected correlations between factors (Osborne & Costello, 2009).

Results

Item Development

Following initial item generation from relevant research (Briesch et al., 2013; Owens et al., 2012) and extant measures (e.g, Academic Competence Evalutaion Scales [ACES], Behavior Rating Inventory of Executive Function [BRIEF]), observation codes (e.g., Achenbach System of Empirically Based Assessment: Direct Observation Form), the CAP provided feedback on an initial pool of 41 potential DBR-MIS items (13 for Academic Enagement, 20 for Interpersonal Skills and 8 for Study Skills). Ranges of mean CAP ratings varied across constructs and across the three dimensions rated by the CAP. For Academic Engagement, ratings of observability ranged from 2.50 to 3.91, ratings of malleability ranged from 1.50 to 3.08, and ratings of social validity ranged from 1.83 to 3.25. For Interpersonal Skills, ratings ranged from 2.25 to 3.58 (observability), from 2.17 to 3.42 (malleability), and from 2.42 to 3.42 (social validity), respectively. For Study Skills CAP ratings ranged from 2.25 to 3.55, from 2.00 to 3.33, and from 2.08 to 3.58 for observability, malleability and social validity, respectively. In response to feedback from the CAP we added an additional interpersonal skills item (*appropriately maintains interactions with peers*).

Next, the SAP provided feedback on all items. Ranges of mean SAP ratings for each item across raters varied across constructs and across the four dimensions. For Academic Engagement, ratings of criterion relatedness and observability both ranged from 2.8 to 4.0, ratings of malleability ranged from 2.0 to 4.0 and ratings of social validity ranged from 2.8 to 4.0. For Interpersonal Skills, ratings of criterion-relatedness ranged from 2.6 to 4.0, ratings of observability ranged from 2.0 to 3.8, ratings of malleability ranged from 2.4 to 3.8, and ratings of social validity ranged from 2.8 to 3.8. Ratings for Study Skills items were consistently higher (criterion relatedness range = 3.4-4.0; observability range = 3.2-4.0; malleability range 3.6-4.0, and social validity range = 3.6-4.0).

A total of five items were deleted because mean ratings from the CAP, the SAP, or both were below 3. Of these deleted items, three were designed to measure interpersonal skills (*appreciates others, shows empathy, and approporiately maintains interactions with peers*) and two were designed to measure academic engagement (*sticks with tasks until complete*, and *shows enthusiasm for lesson*). An additional item (*runs out of time before assignments are complete*) was deleted because it was considered to be redundant with another item (*finishes work on time*). A total of 35 items were submitted to EFA.

Exploratory Factor Analysis

As indicated above, data collection procedures required teachers to rate every item of the three novel scales in order to submit their responses. Descriptive statistics indicated that the distribution of some items on each scale were skewed or kurtotic. All items were within normal limits (i.e., ratio $\leq \pm 3$) for skewness on the Academic Engagement (range = -2.47 - 2.15) and Interpersonal Skills (range = -2.20 - 0.79) scales but not on the Study Skills scale (range = -5.19 - 3.21) whereas some items were kurtotic on all scales (Academic Engagement range = -4.28 - -2.31; Interpersonal Skills range = -4.29 - -2.42; Study Skills range = -4.23 - -2.33). Given the variable directions and severity in skewness and kurtosis for items within each scale, data transformations were not performed as they would have exacerbated skewness and kurtosis for some items.

Inter-item Pearson correlation coefficients within each scale were sufficiently high to conduct EFA, but not so high as to indicate the potential for multicollinearity (i.e., *r*s were generally < .90). Bartlett's Test of Sphericity was significant ($p < .001$) for all scales, which indicated the correlation matrix was not an identity matrix for each scale, and the following Kaiser-Meyer-Olkin Measure of Sampling Adequacy values were obtained: Academic

Engagement = .95, Interpersonal Skills = .97, and Study Skills = .93. Decisions on item reduction were made on the basis of multiple considerations beyond factor loadings including ratings from the CAP and SAP.

Academic Engagement

The initial Academic Engagment scale consisted of 10 items. The range of inter-item correlations was .58-.84 (see Table 3) with base rates from 20.5 to 52.8% (see Table 4). Parallel Analysis indicated a one-factor solution, which explained 73.41% of the observed variance with the overall mean factor loading of the solution was .86 (range = .75-.92). It should be noted that all factor loadings fell within the excellent range (> .71; Comrey & Lee, 1992). A total of five items were selected for retention and the remaining items were removed (see Table 4). Coefficient alpha for the 5-item scale Academic Engagement scale was .94.

Interpersonal Skills

The initial Interpersonal Skills scale consisted of 17 items. The range of inter-item correlations was .545 to .885 (see Table 5) with base rates from 18.2 to 42.7% (see Table 6). Parallel Analysis indicated a two-factor solution. However, the second factor contributed little additional variance (4.78%) and was not conceptually clear. Thus, a single factor solution was forced. This factor explained 71.46% of the observed variance. The overall mean factor loading for items in the one factor model was .84 (range = .75 - .91). A total of five items were retained as shown in Table 6. Coefficient alpha for the 5-item Interpersonal Skills scale was .93.

Study Skills

The initial Study Skills scale consisted of 8 items. The range of inter-item correlations was .465-.834 (see Table 7) with base rates from 21.5 to 58.3% (see Table 8). Parallel Analysis indicated a one-factor solution, which explained 67.10% of the observed variance. The overall

mean factor loading of the solution was .811 (range = .614-.914; see Table 8). A total of 5 items were retained for this scale. The factor loading of one item (*completes homework*) fell below the excellent range of magnitude (i.e. .61), but was still considered acceptable (Stevens, 1992; Tabachnick & Fidell, 2007). Coefficient alpha for the 5-item Study Skills scale was .898.

Correlations between composite scores generated from summing retained items on each scale were high. Correlations between Academic Engagement and Interpersonal Skills and Study Skills were .71 and .90 respectively and the correlation between Interpersonal Skills and Study Skills was .71.

Discussion

In this article, we report the intial development of DBR-MIS designed to measure the three academic enabler constructs of Academic Engagement, Interpersonal Skills, and Study Skills. The current study builds upon extant work in the formative assessment of student behavior in several important ways. First, it broadens the assessment of academic enabler constructs beyond Academic Engagement (e.g., Chafouleas et al., 2010; Volpe & Briesch, 2012, 2015) to include the formative assessment of both Interpersonal Skills and Study Skills. Second, the study employed multiple methods to generate an appropriate pool of items for each construct and is among the first studies to apply EFA methodology to DBR assessment.

Based on the methodology employed by Volpe and Briesch (2012), we collected data from stakeholders (teachers, parents, school administrators and research scientists) regarding criterion-relatedness, observability, malleability and social validity to ensure items generated to measure each academic enabler construct were appropriate for formative assessment in MTSS models. We believe that such screening of items is critically important in the development of formative measures of student behavior and should serve as a model for content validity studies

in DBR assessment. That is, it is not enough to know that specific behaviors are related to a particular construct. In the context of formative assessment in MTSS models, wherein ratings are completed often (e.g., daily or weekly) to evaluate a student's response to intervention, it is also important to know that a) the behavior can be readily observed during the relevant observation interval in the setting of interest (e.g., during classroom instruction on a typical school day; Volpe, McConaughy, & Hintze, 2009), b) it is a behavior that would be expected to change in response to evidence-based intervention (see Gresham et al., 2010), and c) if it did change, it would be of benefit to the student (Wolf, 1978).

We conducted a series of EFAs to identify items that would be the strongest indicators of each academic enabler construct. Results indicated one-factor solutions for both the Academic Engagement and Study Skills constructs. The EFA for Interpersonal Skills items initially generated a two-factor solution, but because the second factor contributed little incremental variance and the grouping of items across the two factors was not conceptually clear, we examined a single-factor model. Across EFAs, items with relatively low base rates and/or low factor loadings in comparison to other items in the same scale were removed, which resulted in 5-item scales for each of these constructs. The single factor models for each of the three academic enabler scales accounted for substantial amounts of variance (between 67.1 and 73.4%) in teacher ratings and each DBR-MIS demonstrated adequate internal consistency (alphas between .898 and .939).

Ours is the first study to examine the factor structure of items designed specifically for progress monitoring of academic enablers and for rating observations conducted over relatively short observation intervals (days or a single week). However, in a recent study, Anthony and DiPerna (2018) conducted an confirmatory factor analysis (CFA) of a short form of the ACES

that included scales similar to those examined in the current study (i.e., Academic Engagement, Interpersonal Skills, and Study Skills). Like many traditional rating scales, directions for the ACES do not specifiy a specific time frame from which informants should base their ratings. Maximally efficient items were drawn from the original teacher-completed ACES (DiPerna & Elliott, 2000) based on item response theory (Anthony & DiPerna, 2017). Results of the CFA were consistent with results of prior studies of the full-length version of the ACES supporting a single tiered structure with correlated factors. This was consistent with the models examined in the current study. However, correlations between the three enabler scales developed in the current study were high (between .71 and .90). Correlations between scales on our newly developed DBR are notably higher than those found for corresponding scales on the ACES in prior studies (e.g., DiPerna et al., 2002, 2005). Although the content of the DBR evaluated in the current study was screened to correspond to the constructs assessed by the ACES, these larger associations between measures assessing related constructs may be the results of our focus on only the most observable and malleable behaviors and ratings conducted based on relatively short observation intervals. Most notably, the correlation between the Academic Engagement and Study Skills scales was much higher than expected. Although one would expect students with high levels of engagement to also demonstrate the effective study habits assessed by the Study Skills scale, to the extent that particular classroom interventions might result in differential intervention response across these two groups of behaviors, it may be useful to maintain these two different scales for the time being.

Limitations

Although the findings of the current study yielded favorable psychometric characterisitcs for the novel academic enabler scales concerning construct validity and internal consistency,

these findings should be considered in light of several limitations. First, although we successfully recruited teachers from 13 states and students rated by these teachers were selected at random from class lists, White students and males were over-represented (U.S. Department of Education, 2018). As such, the results of the current study may not generalize to samples with a higher ratio of students from ethnically diverse backgrounds or samples with a larger proportion of female students. Although the demographics of our participating teachers are fairly consistent with those of teachers in the United States (U.S. Department of Education, 2018), it is not clear to what extent these findings would generalize to samples with a higher proportion of male teachers. Although the size of the sample would be considered adequate for our EFA, due to the composition of the sample in regard to demographic characteristics, it is not well-suited to tests of measurement invariance (e.g., Meade & Bauer, 2007). As such, the aforementioned limitations could not be addressed in the present study.

Future Directions

The current study represents an initial step in a larger scale process of scale development and evaluation. Additional studies currently are being conducted to address the dependability, treatment sensitivity, concurrent and discriminant validity, and acceptability of these novel measures. In the next phase, the dependability of these DBR-MIS will be evaluated through a series of generalizability and dependability studies to identify optimal methods for obtaining dependable progress monitoring data. Treatment sensitivity will be examined by evaluating the extent to which the scales detect changes in student behavior in response to evidence-based classroom intervention (i.e., Daily Behavior Report Card; Volpe & Fabiano, 2013). Future studies also will examine the concurrent validity of these novel scales, using established traditional rating scales and systematic direct observation as criterion measures. In sum, these

studies are designed to provide the evidence needed to utilize this broadened set of tools for the formative assessment of academic enablers in school-based problem-solving models.

References

Anthony, C. J., & DiPerna, J. C. (2017). Identifying sets of maximally efficient items from the Academic Competence Evaluation Scales—Teacher Form. *School Psychology Quarterly*, *32*, 552-559. doi: 10.1037/spq0000205

Anthony, C. J., & DiPerna, J. C. (2018). Piloting a Short Form of the Academic Competence Evaluation Scales. *School Mental Health*, 1-8. doi: 10.1007/s12310-018-9254-7

Atkins, M., Hoagwood, K., Kutash, K., & Seidman, E. (2010). Toward the integration of education and mental health in schools. *Administration and Policy in Mental Health, 37*, 40-47. doi: 10.1007/s10488-010-0299-7

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review, 39*, 408-421.

Briesch, A. M., Ferguson, T. D., Volpe, R. J., & Briesch, J. M. (2013). Examining teachers' perceptions of social-emotional and behavioral referral concerns. *Remedial and Special Education*, *34*, 249-256.

Briesch, A. M., & Volpe, R. J. (2007). Important considerations in the selection of progress monitoring measures for classroom behaviors. *School Psychology Forum, 1,* 59-74.

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., Briesch, A.M., Riley-Tillman, T.C., Christ, T.C., Black, A.C., & Kilgus, S.P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive

behavior of middle school students. *Journal of School Psychology, 48*, 219-246. doi:10.1016/j.jsp.2010.02.001

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem-solving models: Current status and future directions. *School Psychology Review, 39*, 343-349.

Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. A. (2007). Generalizability and dependability of direct behavior ratings to assess social behavior of preschoolers. *School Psychology Review*, *36*, 63-79.

Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct behavior rating (DBR) an emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention, 34*, 195-200.

Chafouleas, S., Riley-Tillman, T. C., & Sugai, G. M. (2007). *School-based behavioral assessment: Informing intervention and instruction*. New York, NY: Guilford Press.

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, *34*, 201-213.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale,NJ: Lawrence Erlbaum Associates.

DiPerna, J. C., & Elliott, S. N. (1999). The development and validation of the Academic Competence Evaluation Scales. *Journal of Psychoeducational Assessment, 17*, 207-225. doi: 10.1177/073428299901700302

DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluation Scales*. San Antonio, TX: The Psychological Corporation.

DiPerna, J. C., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievement: An introduction to the miniseries. *School Psychology Review, 31*, 293-297.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2005). A model of academic enablers and mathematics achievement in the elementary grades. *Journal of School Psychology, 43,* 379-392. doi: 10.1016/j.jsp.2005.09.002.

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2002). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review, 31*, 298-312.

Every Student Succeeds Act (ESSA) of 2015, Pub. L. 114-95, § 114. (2016).

Fabiano, G. A., Vujnovic, R., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Yu, J., . . . Volker, M. (2010). Enhancing the effectiveness of special education programming for children with ADHD using a daily report card. *School Psychology Review, 39*, 219–239.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi: 1082-989X/99/S3.00

Ferguson, T. D., Briesch, A. M., Volpe, R. J., & Daniels, B. (2012). The influence of observation length on the dependability of data. *School Psychology Quarterly, 27*, 187-197. doi: 10.1037/spq0000005

Gettinger, M., & Seibert, J. K. (2002). Contributions of study skills to academic competence. *School Psychology Review*, *31*, 350-365.

Greenwood, C. R. (1996). The case for performance-based instructional models. *School Psychology Quarterly*, *11*, 283-296. doi: 10.1037/h0088935

Gresham, F. M., Cook, C. R., Collins, T., Dart, E., Rasetshwane, K., Truelson, E., & Grant, S. (2010). Developing a change-sensitive brief behavior rating scale as a progress

monitoring tool for social behavior: An example using the Social Skills Rating System—Teacher Form. *School Psychology Review, 39*, 364 –379.

Harvey, V. S., & Chickie-Wolfe, L. A. (2007). *Fostering independent learning: Practical strategies to promote student success*. New York, NY: Guilford Press.

Henricsson, L., & Rydell, A. (2006). Children with behaviour problems: The influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. *Infant and Child Development, 15*, 347-366. doi: 10.1002/icd.448

Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33*, 258-270.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.

Individuals with Disabilities Education Act (IDEA) of 2004, Pub. L. 108-446. 118 U.S.C. § 2647

Jenkins, L. N., & Demaray, M. K. (2015). An investigation of relations among academic enablers and reading outcomes. *Psychology in the Schools*, *52*, 379-389.doi: 10.1002/pits.21830

Malecki, C. K., & Elliot, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17*, 1-23. doi: 10.1521/scpq.17.1.1.19902

McDougal, J. L., Bardos, A. N., & Meier, S. T. (2011). *Behavior Intervention Monitoring Assessment System Technical Manual.* Toronto, Canada: Multi-Health Sytems.

Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4–6 years of age. *Journal of Abnormal Child Psychology*, *36*, 399-410. doi: 10.1007/s10802-007-9186-4

McDonald, R. P. (2013). *Test theory: A unified treatment.* Mahwah, NJ: Erlbaum.

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling, 14,* 611-635. doi: 10.1080/10705510701575461

National Center on Intensive Intervention (n.d.) *Behavioral progress monitoring tools*. Retrieved June 6, 2018, from https://charts.intensiveintervention.org/chart/behavioral-progress-monitoring-tools

National Center on Response to Intervention. (2010, March). Essential components of RTI-A closer look at response to intervention. Washington, DC: U.S. Department of Education, Office of Special Education Programs, Author.

Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review, 12*, 131-146.

Owens, J. S., Holdaway, A. S., Zoromski, A. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., & Murphy, C. E. (2012). Incremental benefits of a daily report card intervention over time for youth with disruptive behavior. *Behavior Therapy*, *43*, 848-861. doi: 10.1016/j.beth.2012.02.002

Pelham, Jr, W. E., Fabiano, G. A., & Massetti, G. M. (2005). Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 449-476. doi: 10.1207/s15374424jccp3403_5

Rabiner, D., & Coie, J. D. (2000). Early attention problems and children's reading achievement: A longitudinal investigation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 859-867. doi: 10.1097/00004583-200007000-00014

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700-712. doi: 10.1037/a0027268

Reynolds, C. R., & Kamphaus, R. W. (2009). *BASC-2 Progress Monitor*. Minneapolis, MN: NCS Pearson.

Reynolds, A. J., & Walberg, H. J. (1992). A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education*, *23*, 306-328. doi: 10.2307/749308

Riley-Tillman, T.C., Chafouleas, S.M., Sassu, K.A., Chanese, J.A.M., & Glazer, A.D. (2008). Examining the agreement of Direct Behavior Ratings and Systematic Direct Observation for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*, 136-143. doi:10.1177/1098300707312542

Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional problems. *Psychology in the Schools*, *41*, 551-561. doi: 10.1002/pits.10176

Spectrum K12 School Solutions. (2011). *Response to intervention adoption survey 2011*. http://www.globalscholar.com/2011RTI

Stevens, J. (1992). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17,* 223-237. doi: 10.1080/09362830903235375

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Fifth Edition. New York: Pearson Education Inc.

Tilly, D. (2008). The evolution of school psychology to science based practice. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp. 17-36). Bethesda, MD: National Association of School Psychologists.

U.S. Department of Education. (2018). *National Center for Education Statistics, Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey," 2017-2018.* Washington, DC: National Center for Education Statistics.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and Solutions in Human Assessment* (pp. 41-71). Boston, MA: Springer.

Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review*, *41*, 246-261.

Volpe, R. J., & Briesch, A. M. (2015). Multi-item Direct Behavior Ratings: Dependability of two levels of assessment specificity. *School Psychology Quarterly, 30*, 431-442. doi: 10.1037/spq0000115

Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, *34*, 454-474.

Volpe, R. J., DuPaul, G. J., DiPerna, J. C., & Jitendra, A. K. (2006). Attention deficit hyperactivity disorder and scholastic achievement: A model of mediation via academic enablers. *School Psychology Review*, *35*, 47-61.

Volpe, R. J. & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of assessment and intervention*. New York, NY: The Guilford Press.

Volpe, R. J., & Gadow, K. D. (2010). Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review*, *39*, 350-363.

Volpe, R. J., Gadow, K. D., Blom-Hoffman, J., & Feinberg, A. B. (2009). Factor-analytic and individualized approaches to constructing brief measures of ADHD behaviors. *Journal of Emotional and Behavioral Disorders*, *17*, 118-128. doi: 10.1177/1063426608323370

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the Direct Observation Form. *School Psychology Review*, *38*, 382-401.

Wolf, M. M. (1978). Social validity: The case for subjective measurement of how behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203-214. doi: 10.1901/jaba.1978.11-203

Wood, B. K., Hojnoski, R. L., Laracy, S. D., & Olson, C. L. (2015). Comparison of observational methods and their relation to ratings of engagement in young children.

Topics in Early Childhood Special Education, *35*, 211-222. doi: 10.1177/0271121414565911

Zirkel, P. A. (2011). State laws and guidelines for RTI: Additional implementation features. *Communiqué*, *39*(7), 30-32.

Compliance with Ethical Standards

Preparation of this article was supported by funding provided by the Institute of Education Sciences, U.S. Department of Education (R324A150071). The authors declare that they have no conflicts of interest. All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was deemed exempt from obtaining signed written consent from participants in accordance with HHS regulations at 45 CFR 46.117(c). This article does not contain any studies with animals performed by any of the authors.

Table 1.

Teacher Demographics

	Frequency	Percentage of Sample
Gender		
Male	13	4.2%
Female	294	95.8%
Years of Experience		
0-5	73	23.8%
6-10	61	19.9%
11-20	107	34.9%
21-30	48	15.6%
> 30	18	5.9%

Table 2.

Student Demographics

	Frequency	Percentage of Sample
Gender		
Male	187	60.9%
Female	120	39.1%
Race		
White	206	67.1%
Black	40	13.0%
Latino	46	15.0%
Asian	10	3.3%
American/Alaska Native	3	1.0%
Hawaiian/Pacific Islander	1	0.3%
Unknown	16	5.2%
Age (Years)		
5	26	8.5%
6	67	21.8%
7	71	23.1%
8	82	26.7%
9	55	17.9%
10	5	1.6%
Grade		
Kindergarten	82	26.7%
First	64	20.8%
Second	79	25.7%
Third	82	26.7%
Special Education Status		
Receiving Special Education	109	35.5%
Not Receiving Special Education	196	63.8%
Unsure	2	0.7%

Table 3.

Motivation/Engagement Inter-Item Correlations

Item[a]	1	2	3	4	5	6	7	8	9	10
1. Interest in task	–									
2. Independent	.72	–								
3. Avoids being distracted	.69	.76	–							
4. Starts promptly	.76	.77	.79	–						
5. Works hard	.81	.73	.75	.79	–					
6. Raises hand	.65	.68	.68	.70	.69	–				
7. Adequate progress	.74	.75	.72	.82	.75	.60	–			
8. On time	.74	.74	.68	.82	.71	.58	.84	–		
9. Actively participates	.77	.67	.61	.73	.71	.63	.69	.73	–	
10. On task	.77	.79	.86	.85	.81	.65	.79	.77	.68	–

Note. [a] = abbreviated description of item.

Table 4.

Engagement Item Means, Standard Deviations, Base Rates, and Factor Loadings

Item[a]	Mean	SD	Base Rate	Factor Loading
1. Interest in task*	3.77	1.45	20.5%	.862
2. Independent	3.54	1.79	31.6%	.859
3. Avoids being distracted*	2.61	1.76	52.8%	.848
4. Starts promptly	3.34	1.76	37.8%	.919
5. Works hard*	3.91	1.54	23.1%	.877
6. Raises hand*	3.73	1.67	24.4%	.749
7. Adequate progress*	3.65	1.64	30.3%	.871
8. On time	3.54	1.80	35.2%	.860
9. Actively participates	3.84	1.59	25.4%	.800
10. On task	3.30	1.62	35.8%	.912

Note. [a] = abbreviated description of item. * = item removed

Table 5.

Interpersonal Skills Inter-Item Correlations

Item[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Shows interest	–																
2. Positive comments	.76	–															
3. Compromises	.69	.67	–														
4. Expresses feelings	.63	.65	.69	–													
5. Helps	.77	.78	.68	.67	–												
6. Initiates interactions	.72	.72	.69	.73	.80	–											
7. Respectful	.67	.69	.73	.67	.71	.69	–										
8. Cooperates	.73	.72	.78	.71	.74	.77	.84	–									
9. Listens	.78	.10	.70	.64	.72	.71	.70	.74	–								
10. Contributes	.73	.66	.62	.63	.75	.76	.62	.72	.70	–							
11. Shares	.64	.63	.67	.64	.73	.73	.65	.71	.66	.68	–						
12. Polite	.67	.68	.73	.67	.68	.70	.89	.81	.68	.62	.66	–					
13. Plays well	.68	.68	.72	.71	.74	.80	.80	.85	.70	.68	.69	.79	–				
14. Kind	.68	.72	.68	.68	.73	.71	.85	.81	.69	.62	.67	.86	.81	–			
15. Asks for help	.63	.62	.54	.69	.71	.68	.62	.64	.65	.70	.59	.60	.62	.60	–		
16. Works well	.71	.69	.75	.69	.74	.77	.82	.87	.72	.72	.69	.80	.86	.79	.63	–	
17. Takes turns	.74	.71	.75	.68	.73	.79	.77	.83	.78	.71	.75	.77	.82	.75	.68	.80	–

Note. [a] = abbreviated description of item

Table 6.

Interpersonal Skills Item Means, Standard Deviations, Base Rates, and Factor Loadings

Item[a]	Mean	SD	Base Rate	Factor Loading
1. Shows interest*	3.64	1.505	27.4%	.828
2. Positive comments*	3.44	1.588	32.6%	.818
3. Compromises	3.00	1.730	42.7%	.820
4. Expresses feelings	3.39	1.682	33.6%	.792
5. Helps*	3.63	1.582	26.4%	.864
6. Initiates interactions	3.66	1.601	27.0%	.871
7. Respectful	4.06	1.520	19.2%	.871
8. Cooperates	3.76	1.568	26.4%	.912
9. Listens*	3.68	1.476	26.7%	.832
10. Contributes*	3.64	1.579	26.4%	.804
11. Shares*	3.91	1.534	21.8%	.793
12. Polite*	4.10	1.514	18.9%	.860
13. Plays well*	3.94	1.598	22.1%	.888
14. Kind*	4.14	1.523	18.2%	.863
15. Asks for help*	3.63	1.632	26.4%	.746
16. Works well*	3.80	1.604	25.1%	.895
17. Takes turns*	3.87	1.523	21.5%	.893

Note. [a] = abbreviated description of item. * = item removed

Table 7.

Study Skills Inter-Item Correlations

Item[a]	1	2	3	4	5	6	7	8
1. Checks work	–							
2. Accurate	.69	–						
3. Well organized	.79	.71	–					
4. Prepared	.76	.69	.82	–				
5. Writes down	.61	.47	.58	.61	–			
6. Keeps track	.74	.61	.85	.83	.59	–		
7. Follows instructions	.78	.72	.79	.81	.56	.78	–	
8. Completes	.51	.49	.52	.58	.50	.57	.47	–

Note. [a] = abbreviated description of item.

Table 8.

Study Skill Item Means, Standard Deviations, Base Rates, and Factor Loadings

Item[a]	Mean	SD	Base Rate	Factor Loading
1. Checks work*	2.48	1.77	58.3%	.866
2. Accurate	3.66	1.56	26.1%	.764
3. Well organized	3.08	1.87	42.3%	.907
4. Prepared*	3.55	1.72	29.0%	.914
5. Writes down*	2.39	2.11	53.7%	.666
6. Keeps track	3.33	1.87	35.8%	.883
7. Follows instructions	3.67	1.56	26.1%	.880
8. Completes	4.13	1.96	21.5%	.614

Note. [a] = abbreviated description of item. * = item removed