2020

Developing a Brief Behavior Rating Scale for Progress Monitoring of Depression in School Settings

Evan H. Dart

Prerna G. Arora

Tai Collins

Kevin Stark

Clayton R. Cook

See next page for additional authors

Authors

Evan H. Dart, Prerna G. Arora, Tai Collins, Kevin Stark, Clayton R. Cook, Mylien T. Duong, Carolyn A. McCarty, and Beth Doll

Developing a Brief Behavior Rating Scale for Progress Monitoring of Depression in School Settings

Evan H. Dart,[1] Prerna Arora,[2] Tai Collins,[3] Kevin Stark,[4] Clayton R. Cook,[5] Mylien T. Duong,[6] Carolyn A. McCarty,[6] and **Beth Doll** [7]

1 The University of Southern Mississippi, Hattiesburg, MS
2 Pace University, New York, NY
3 University of Cincinnati, Cincinnati, OH
4 The University of Texas at Austin, TX
5 University of Minnesota, Minneapolis, MN
6 University of Washington, Seattle, WA
7 University of Nebraska–Lincoln, NE

Corresponding author — Evan H. Dart, University of Southern Mississippi, 118 College Dr., #5025 Hattiesburg, MS 39406, USA. Email: evan.dart@usm.edu

Abstract

Frequent formative assessment of students' functioning, or progress monitoring, is a critical component of multi-tiered systems of support as data inform data-driven decisions about response to treatment. Progress monitoring tools for students' academic and behavioral functioning are readily available and widely researched; however, despite the documented prevalence of depressive disorders among youth and that schools have been put forth as an ideal location for the delivery of mental health services, there are currently no progress monitoring tools to examine students' response to interventions that target depression. To address this gap, this

study sought to develop a progress monitoring assessment of students' depressive symptoms using an empirically informed model for creating *Brief Behavior Rating Scales* (BBRS). Using this model, a four-item BBRS of depressive symptoms (BBRS-D) was created from the item pools of the *Beck Depression Inventory for Youth* (BDI-Y) and *Children's Depression Inventory* (CDI) administered during a treatment study of depression in female youth; the resulting short scale corresponds well to the full-length assessments (i.e., $r = .65$ and $r = .59$); however, the BBRS-D possessed lower than adequate internal consistency ($\alpha = .50$)

Keywords: progress monitoring, rating scales, emotional/behavioral disorders

Progress monitoring assessments for students' academic and behavioral functioning are readily available and widely researched (Chafouleas, Volpe, Gresham, & Cook, 2010; Fuchs & Fuchs, 2011); however, despite the documented prevalence of depressive disorders among youth (Avenevoli, Swendsen, He, Burstein, & Merikangas, 2015) and schools serving as an ideal location for the delivery of mental health services for youth (Arora, Nastasi, & Leff, 2017; Costello, He, Sampson, Kessler, & Merikangas, 2014), there are currently no progress monitoring tools to assess students' depression over time. Accordingly, this study sought to develop such an assessment to monitor the progress of students'depressive symptoms in schools. Following a review of the literature regarding depression among youth and school-based treatment targeting depression, we discuss the need for school-based progress monitoring tools that can be used to examine student response to interventions that target depression. We conclude with a discussion of the development of our measure, the *Brief Behavior Rating Scale for Depression* (the BBRS-D), as well as implications for usability in schools.

Depression in Youth

Depressive disorders are experienced by a significant number of youth (Center for Behavioral Health Statistics and Quality, 2015). Such disorders are characterized by symptoms such as depressed mood, anhedonia, changes in sleep, appetite, loss of energy, feelings of guilt, and difficulty concentrating (American Psychiatric Association, 2013). Specifically, at any given time, approximately 5% to 8% of youth meet criteria for a depressive disorder, including major depressive disorder

(MDD) or dysthymic disorder (DD; Angold & Costello, 2001; P. Cohen et al., 1993). More recent data suggest that the prevalence of a depressive disorder in adolescents (ages 13–18) is 11.7% (Merikangas et al., 2010). Starting in adolescence, females are diagnosed with depression at a rate of 2 to 1 compared with boys (Galambos, Leadbeater, & Barker, 2004); this disparity persists, and even increases, during adulthood (Hankin & Abramson, 2001). Childhood depression negatively impacts academic achievement; impairs family, peer, and early romantic relationships; and increases risk for further depressive disorders in adulthood (Gould et al., 1998; Lewinsohn, Roberts, Seeley, Rohde, Gotlib, & Hops, 1994). Thus, due to their high prevalence in childhood and negative impact on youth functioning, childhood treatment for depressive disorders is crucial (Kazdin & Weisz, 1998).

School-Based Treatment of Depression

Schools have been posited as the ideal setting for addressing the needs of youth with depression (Rones & Hoagwood, 2000). Specifically, because the delivery of interventions by school psychologists in such settings permits greater access to teacher and parents, as well as opportunity to observe problem behaviors in a naturalistic setting, schools have been highlighted as an excellent environment in which to facilitate the delivery of interventions for depressed youth (Stark, Arora, & Funk, 2011).

The majority of interventions for youth depression are based on and thus incorporate aspects of cognitive behavioral therapy (CBT; Maag, Swearer, & Toland, 2009), which has obtained extensive support for the treatment of youth depression (David-Ferdon & Kaslow, 2008). Various manualized programs for the treatment of youth depression exist, including the ACTION treatment program (Stark, Streusand, Arora, & Patel, 2011), the Adolescent Coping With Depression course (CWD-A; Clarke, Lewinsohn, & Hops, 1990; Ruffalo & Fischer, 2009), and the Primary and Secondary Control Enhancement Training (Weisz, Thurber, Sweeney, Proffitt, & LeGagnoux, 1997). Generally, trials of these interventions, though limited in number, have demonstrated support for their effectiveness in reducing symptoms of depression within the school context (Patel, Stark, Metz, & Banneyer, 2014). In addition, research examining the effectiveness of youth depression

prevention programs, such as the Positive Thoughts and Actions program (PTA; McCarty, Violette, Duong, Cruz, & McCauley, 2013), have also been undertaken, with findings supporting their use with school-aged youth (Stice, Shaw, Bohon, Marti, & Rohde, 2009).

Although school-based service delivery has historically focused on academic and social behavioral concerns (Huang et al., 2005), increased attention has been placed on the role of schools in addressing the needs of depressed youth (Stark, Arora, & Funk, 2011). Specifically, multitiered frameworks of service delivery (e.g., multi-tiered systems of support [MTSS]) have been applied to prevent and address the needs of students with internalizing problems generally, and depression specifically (Carnevale, 2013; Herman, Merrell, Reinke, & Tucker, 2004; McIntosh, Ty, & Miller, 2014). However, these efforts are in their infancy, and additional support to strategically inform the implementation of practices targeting depression within MTSS framework is needed.

Progress Monitoring for Depression in Schools

The implementation of an MTSS framework designed to address youth depression in schools is dependent upon the availability of screening, intervention evaluation, and progress monitoring measures that are used to make treatment decisions (Merrell, 2013). There are multiple publicly available measures that can be used for youth depression screening (e.g., the Student Internalizing Behavior Screener; Cook et al., 2011; Youth Internalizing Problems Screener; Renshaw & Cook, 2018) and treatment evaluation (e.g., *Center for Epidemiologic Studies Depression Scale for Children*; Faulstich, Carey, Ruggiero, Enyart, & Gresham, 1986). Unfortunately, measures for progress monitoring depressive symptoms are notably lacking (von der Embse, Scott, & Kilgus, 2015) and, without such measures, it will be impossible to engage in data-driven decision making within an MTSS framework (Ardoin & Christ, 2009; Hawken, Vincent, & Schumann, 2008). Specifically, traditional rating scales often are too costly and time intensive, thus limiting the feasibility of their use in schools. Furthermore, these traditional rating scales are often not designed for frequent use and are not suitable for measuring small changes in behavior (Christ, Riley-Tillman, & Chafouleas, 2009), thus limiting their potential use as progress

monitoring measures. In a recent systematic review of the literature, though several assessments were identified that might be practically used to monitor internalizing symptoms in children in schools, no empirically supported progress monitoring assessments for school-based treatment of depressive symptoms were found (Dart, Arora, Collins, & Doll, under review). Accordingly, the need for future research to develop or examine the appropriateness, feasibility, and technical adequacy of promising assessments for the purpose of progress monitoring in schools was underscored (Dart et al., under review).

Current Study

In light of the paucity of progress monitoring tools for depression in the context of MTSS in schools, this study sought to develop and validate a measure that can be used in schools to progress monitor students' depressive symptoms. Accordingly, and following Cook and colleagues' (2013) promising four-step model for the development of new BBRSs, this study outlines the systematic development and validation of a BBRS for depressive symptoms—the BBRS-D. In addition, because recent research has suggested that teachers may not be accurate in identifying student internalizing disorders (e.g., Cunningham & Suldo, 2014), this study sought to develop a student-report version of the BBRS-D for use in a school setting.

Method

Gresham and colleagues (2010) and Cook, Volpe, and Delport (2013) outlined a model for developing new BBRSs that involves four distinct steps. This model was based on developments in the clinimetrics literature (de Vet, Terwee, & Bouter, 2003) that provide an empirical precedent for the BBRS development process. The four steps are (a) identify change-sensitive items from standardized rating scales gathered as part of a study evaluating the impact of an intervention, (b) assemble brief and technically adequate rating scales using these items, (c) verify that the rating scales are sensitive to change in response to any treatments that target the domain of interest (e.g., depression), and (d) assess the social validity of the rating scales. This study set out to

accomplish the first two steps by identifying change-sensitive items from extant standardized rating scales of depressive symptoms in children and use those items to assemble the BBRS-D. The first two steps from the model were first implemented by Gresham et al. (2010) to develop a BBRS for social behavior, and their procedures were replicated here.

For the purpose of this study, change-sensitivity is defined as an item's ability to detect change on a specific outcome in the intended direction during or following a course of treatment. Treatment studies are used to identify change-sensitive items because it becomes possible to compare and differentiate intended change (i.e., item score change over the course of treatment) to change from uncontrolled sources or error (e.g., regression toward the mean). In this way, it is possible to identify the items from a measure which are most sensitive to the effects of a treatment.

Participants

Data from a larger intervention study were used to develop the BBRS-D. Specifically, 130 female youth were part of a treatment study examining the effectiveness of the ACTION treatment program (Stark, Streusand, Arora, & Patel, 2011). The treatment study included 148 total participants; however, 18 (12.2%) of those cases were missing data and were removed from this analysis. Participants included youth from the CBT only ($n = 44$), the CBT + parent training ($n = 43$), and control ($n = 43$) conditions. Participants ranged in age from 9 to 14 years ($M = 10.65$ years; $SD = 1.33$) and Grades 4 to 7. Youth race/ethnicity varied, with the majority of participants being White/nonHispanic ($n = 52$) or Hispanic ($n = 50$). Each participant had previously received a diagnosis of MDD ($n = 94$), DD ($n = 26$), Depressive Disorder Not Otherwise Specified (DD-NOS; $n = 19$), or diagnoses of both MDD and DD ($n = 7$). A large portion of youth had at least one comorbid diagnosis ($n = 45$).

Dataset and Intervention

Implementing the BBRS development procedures requires an extant dataset from which BBRS items will be drawn. A dataset from a completed investigation of the ACTION treatment program was available

and met all of the criteria for BBRS-D construction. That is, it contained item-level data from a standardized rating scale assessing depression that was used to evaluate the effects of a treatment for depression in a pre–post fashion. Please note, because the primary purpose of the original investigation was a treatment evaluation, the analyses we conducted represent a repurposing of the dataset.

Measures

Beck Depression Inventory for Youth (BDI-Y). The BDI-Y is a 20-item self-report assessment of symptoms related to depression in children between the ages of 7 and 14 (Beck, Beck, & Jolly, 2001). Items on the BDI-Y are presented as statements to which respondents are asked to respond how applicable each statement is to them. Each statement is worded to indicate symptomatology of depression (e.g., "I think that my life is bad.") and is rated on a 4-point scale (i.e., *never, sometimes, often, always*). In a sample of 859 adolescent females (Stapleton, Sander, & Stark, 2007), the total score of the BDI-Y was found to have adequate internal consistency (i.e., $\alpha = .93$) and correlated highly (i.e., $r = .83$) with the *Children's Depression Inventory* (CDI; Kovacs, 1992).

CDI. The CDI is a 27-item self-report assessment of symptoms related to depression in children between the ages of 7 and 17 (Kovacs, 1992). Items on the CDI are presented as three different but related statements (e.g., "I am sad once in a while." "I am sad many times." and "I am sad all of the time."), and respondents are asked to select which is most true of them. For BBRS-D development, only the most severe statement was used to name the item grouping on the CDI. For example, "I am sad all of the time" would have been selected from the trio presented above if that grouping of items was identified as change sensitive and included on the BBRS-D. In a sample of 147 adolescent females (Smucker, Craighead, Craighead, & Green, 1986), the CDI was found to have adequate internal consistency ($\alpha = .89$). In addition, it shared a moderate correlation ($r = .58$) with the *Center for Epidemiological Studies Depression Scale* (CES-D) in a sample of approximately 1,200 elementary and middle school students (Doerfler, Felner, Rowlison, Raley, & Evans, 1988).

Procedure

Gresham and colleagues (2010) specified a procedure for identifying change-sensitive items from extant rating scales and assembling a BBRS from those items. The procedure involves (a) identifying all items from each rating scale that are sensitive to change; (b) determining the psychometric properties of the initial change-sensitive item pool; and (c) systematically reducing the change-sensitive item pool while retaining adequate psychometric properties to construct the briefest scale possible to enhance feasibility and usability. Each of these three procedures are described in more detail throughout the following sections. All statistical analyses were conducted using Statistical Package for the Social Sciences (IBM Corp. Released, 2012).

Change-Sensitivity Metrics

To identify items from the measures that were sensitive to change, four different metrics of change-sensitivity (Gresham et al., 2010) were used to analyze each of the 47 items (i.e., 20 BDI-Y items and 27 CDI items). Each metric used item scores across participants to determine if statistically significant differences existed between the treatment and control groups from pretreatment to posttreatment. That is, each metric is based in the logic that an item demonstrating positive change (i.e., indicating a reduction in symptoms) from pretreatment to posttreatment for the treatment group, but not the control group, indicates sensitivity to treatment. The four change-sensitivity metrics included an ANOVA, a *t* test, an odds ratio (OR), and a standardized mean difference effect size (SMDES). To be identified as change-sensitive, an item had to meet the criteria on at least two of the four metrics. Although Cook and colleagues (2013) propose a three-metric criterion, a two-metric criterion was chosen because it ensured that items identified by one of the metrics was confirmed by at least one other metric without being too conservative during the item identification process (C. Cook, personal communication, June 11, 2016). Each of these metrics are described in detail below.

ANOVA. The first metric involved conducting a two-way mixed ANOVA for each item where time (i.e., pretreatment and posttreatment) was treated as a within-subjects factor and condition (i.e.,

control group or treatment group) was treated as a between-subjects factor. Changesensitivity was determined by examining the interaction effect between the two factors. That is, if a statistically significant interaction effect indicated that an item score improved to a greater degree between pretreatment and posttreatment for the treatment group when compared with the control group, it was considered change sensitive (Gresham et al., 2010). Because 47 different ANOVAs were conducted, the Bonferroni post hoc correction was used to adjust *p* values when evaluating statistically significant differences. Thus, a *p* value of .00106 (.05 / 47) was adopted for these analyses.

The t test. The *t* test metric involved a two-step process including both an independent-samples *t* test and a pairedsamples *t* test (Gresham et al., 2010). First, an independent-samples *t* test was conducted between the posttreatment scores of the treatment and control group for each item. Any item that produced a statistically significant *T* score was retained for the second step. Next, a dependent-samples *t* test was conducted between the pretreatment and posttreatment scores of the treatment group for any item passing the first *t*-test criterion. Any items producing a statistically significant *T* score at this step were considered change-sensitive. The two significant *t* values signaled that an item was indicative of a treatment effect in individuals receiving treatment and that statistically significant positive change occurred between the pretreatment and posttreatment scores for that item. A *p* value of .00106 (.05 / 47) was adopted for these analyses to adjust for multiple comparisons.

OR. The third metric, an OR, was used to determine the relative odds a member of the treatment group would show favorable change on a single item compared with the control group from pretreatment to posttreatment. To do this, it was necessary to generate a score for each participant across each item that indicated whether that participant demonstrated favorable change on the item between pretreatment and posttreatment. As lower scores on both assessments indicated improved functioning, each participant's pretreatment score was subtracted from their posttreatment score. If the result was a positive number (e.g., 1), this was taken as evidence of favorable change. On the contrary, if the result was a negative number (e.g., –1) or 0, there was no evidence of favorable change. Next, a 2 × 2 table was constructed for

each item whereby favorable change status (i.e., yes or no) served as the rows and group membership (i.e., treatment or control) served as the columns. These tables were used to calculate the odds that a member of the treatment group demonstrated favorable change compared with the control group using a crosstabs analysis within SPSS. Finally, a χ² statistic was calculated on the 2 × 2 table data to determine if the associated OR was statistically significant. An item producing a statistically significant χ² was considered change-sensitive.

SMDES. The fourth, and final, metric consisted of a SMDES to determine the standardized difference between pretreatment and posttreatment change scores for the treatment and control groups. The following formula was used:

$$\text{SMDES} = \frac{(Post_C - Pre_C) - (Post_T - Pre_T)}{SD_{pooled}}$$

First, mean difference scores between pretreatment and posttreatment were obtained for both the treatment and control group by subtracting the mean pretreatment score from the mean posttreatment score for each item. Next, each item's mean difference score from the treatment group was subtracted from its mean difference score from the control group. This mean difference was divided by the pooled (i.e., control and treatment participants) standard deviation of the pretreatment scores for that particular item.

The result was a standardized mean difference for each item that was interpretable using guidelines published by J. Cohen (1992). To be identified as change-sensitive on this metric, an item had to produce an SMDES of .50 or greater. An SMDES indicated that a particular item produced positive change to a greater degree (i.e., a moderate difference in standard deviation units) in the treatment group than in the control group (Gresham et al., 2010).

Psychometric Evaluation

Once we identified the change-sensitive items, the second step in developing the BBRS-D involved an evaluation of the psychometric properties of the item pool. The psychometric evaluation consisted of three analyses examining the item pool's test–retest reliability,

internal consistency, and evidence of validity based on relations to other variables (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing, 1999). All three analyses are described in greater detail in the following sections.

Test–retest reliability. The test–retest reliability of the item pool was determined by calculating a Pearson's *r* correlation between the pretreatment and posttreatment total scores of the item pool for the control group only. Test–retest reliability was calculated using only the control group because they did not receive any treatment; thus, changes between their pretreatment and posttreatment scores are expected to reflect symptom changes due to uncontrolled sources or error. A correlation of .70 was used as a cutoff to determine adequate test–retest reliability (Nunnally & Bernstein, 1994).

Internal consistency. The internal consistency of the item pool was determined by calculating Cronbach's alpha (Cronbach, 1951) for all items included in the pool. The pretreatment item scores for the total sample were used in this analysis. An alpha of .70 was used as the cutoff for adequate internal consistency (Nunnally & Bernstein, 1994).

Validity. Evidence for validity was obtained through correlations between the total score for the item pool and the total scores for the full-length BDI-Y and CDI. Because the BBRS-D contained items from each measure, any items included in the BBRS-D were removed from the BDI-Y and CDI total scores to reduce redundancy and prevent artificial inflation of validity coefficients. The pretreatment total scores for the total sample were used in this analysis. J. Cohen's (1992) criterion for a strong correlation (i.e., $r > .50$) was used as the cutoff for determining adequate validity.

Item Reduction Process

Once the psychometric properties of the item pool were established, the final step in BBRS-D development involved an iterative item reduction process. The item reduction process involves dropping the item with the weakest scores across the four change-sensitivity metrics

and recalculating the psychometric properties of the remaining items. This process continues until either (a) internal consistency fell below .70, (b) temporal stability fell below .70, or (c) the validity coefficient fell below .50.

Results

Item Identification

A total of four of the 47 items (8.5%) met the criterion of being identified as change sensitive according to two or more of the above defined metrics (see Table 1). Fifteen items (31.9%) met criteria on one of the change-sensitivity metrics. The remaining 28 (59.6%) items did not meet the criteria of any of the change-sensitivity metrics. Because the original BDI-Y and CDI items are protected by copyright, all items presented in this article have been modified from their original wording to reflect the symptom they were intended to assess. Item re-wording took place during the manuscript writing process. All analyses were conducted using original item wording and the BBRS-D contains the original item wording from the BDI-Y and CDI. Please contact the authors for a copy of the BBRS-D with original item wording.

Psychometric Properties

Psychometric properties of the four change sensitive items were then calculated (see Table 2). The final four-item the total scores of the BDI-Y (r = .65) and CDI (r = .59). The internal consistency of the assessment was not adequate (α = .50). In addition, the temporal stability (11-week test– retest coefficient) of the assessment was below our cutoff of .70 (r = .56).

Item Reduction

The BBRS development process would typically end with an item reduction process; however, because the initial BBRS-D's internal consistency and test–retest reliability were already lower than the criteria, there was no need to proceed with elimination of items. Doing so would have further reduced the internal consistency and test–retest reliability.

Table 1. Results of the Change-Sensitivity Analyses on items from the BDI-Y and CDI that met at least one change-sensitivity criterion.

	Change-Sensitivity Metrics			
Item	ANOVA (F)	t Tests (t)	OR (χ^2, p)	SMDES
Stomach pain[a]	11.98; $p < .001$	—	5.85 (12.56, $p < .001$)	.61
Sadness[a]	—	5.11; $p < .001$	—	.52
Sleep problems[b]	—	3.48; $p < .001$	—	.51
Tiredness[b]	—	6.06; $p < .001$	—	.57
Poor performance[b]	—	3.53; $p < .001$	—	—
Emptiness[a]	—	4.67; $p < .001$	—	—
Difficulty sleeping[a]	—	5.16; $p < .001$	—	—
Loneliness[a]	—	5.61; $p < .001$	—	—
Unloved[a]	—	4.76; $p < .001$	—	—
Stupidity[a]	—	4.63; $p < .001$	—	—
Feeling sorry[a]	—	3.79; $p < .001$	—	—
Poor performance[a]	—	4.87; $p < .001$	—	—
Sadness[b]	—	—	—	—
Hopelessness[b]	—	—	3.21 (7.69; $p < .006$)	—
Suicidality[b]	—	3.36; $p < .001$	—	—
Social avoidance[b]	—	3.44; $p < .001$	—	—
Physical symptoms[b]	—	3.65; $p < .001$	—	—
Fighting[b]	—	—	3.73 (4.52; $p < .033$)	—
Self blame[b]	—	—	2.63 (3.94; $p < .047$)	—

Table 2. Psychometric Properties of the BBRS-D.

	Reliability		Validity	
BBRS-D	Test–Retest (11 Weeks)	Internal Consistency	BDI-Y	CDI
4 items	.56	.50	.65	.59

Discussion

The purpose of this study was to use an empirically informed process to develop a progress monitoring tool to assess depressive symptoms in response to school-based interventions. We utilized a framework developed by Gresham and colleagues (2010) and Cook and colleagues (2013) to identify change-sensitive items from an extant dataset evaluating the impact of a school-based intervention targeting depression. These change sensitive items were then combined into

a brief (i.e., four items) progress monitoring measure that would be feasible for use in a school setting. This work resulted in a four-item BBRS-D with adequate validity coefficients against full scale measures of depression in youth (i.e., the BYI-D and CDI); however, we could not establish that the reliability (neither internal consistency nor long-term stability) reached a satisfactory level. The development of this brief scale is particularly promising given the prevalence of depressive symptoms in youth and the growing focus on MTSS and school-based mental health (Carnevale, 2013; Herman et al., 2004), as well as the dearth of research on progress monitoring tools for depression in schools.

The BBRS-D is potentially advantageous for a number of reasons. First, preliminary evidence supports the validity of the BBRS-D. Also, much like the BBRS for social skills developed by Gresham and colleagues (2010), the BBRS-D may represent a general outcome measure (GOM) for overall depressive functioning. Specifically, the BBRS-D includes items measuring a variety of symptoms of depression, including sadness, tiredness, and somatic complaints. The identification of items in only one of these domains would have limited the use of the BBRS-D; however, the breadth of symptoms captured by the measure suggests that it may represent the general outcome of depression. The criterion-related validity estimates obtained support this assertion. Third, the four-item BBRS-D would be highly feasible for use in frequent progress monitoring in most school settings, considering that it would likely take a person between 30-s (5 s per item) to a minute (10 s per item) to complete the scale. Considering this, it may be feasible for school staff to complete the BBRS-D for daily administration. In contrast to assessments such as the CDI and BDI-Y, the BBRS-D has a small number of items, requires less administration time, and is composed of items that are sensitive to small changes in behavior.

Limitations and Future Directions

The results of the current study should be interpreted within the context of a number of limitations. First, we identified items as change-sensitive according to only two out of the four statistical metrics instead of a three-metric criterion suggested by Cook and colleagues (2013); however, the exploratory nature of this work and the few items that were ultimately identified lend credence to our decision

to utilize a criterion of two out of four metrics to determine change-sensitivity. In addition, the use of a twometric criterion ensured that a single metric was not solely responsible for identifying an item, allowing us to confirm our findings. Future research should develop empirically derived guidelines for identifying change-sensitive items for the purposes of school-based progress monitoring measure development.

Second, the sample used in our analyses was comprised entirely of female youth. Although depression is much more common in females than males (Galambos et al., 2004), it is unclear if the BBRS-D would be appropriate for progress monitoring depressive symptoms of male students. Differences in response to treatment may make some items more or less sensitive to change depending on an individual's gender. Future research should investigate whether gender and other demographic variables impact the change-sensitivity of the BBRS-D.

Third, we did not evaluate the BBRS-D within the context of a novel treatment setting. As mentioned previously, Cook and colleagues' (2013) four-stage model of BBRS development includes a verification of the assessment's change-sensitivity and an investigation of its social validity as the third and fourth stages, respectively. We did not verify the change-sensitivity of the BBRS-D in a context independent of the development process. That is, these items were only identified as change-sensitive in response to the ACTION treatment program. It is extremely important to verify that these items are intervention independent and are change sensitive in response to other treatments for depression. In addition, we did not investigate the social validity of the BBRS-D as an assessment. Researchers should incorporate the BBRS-D into future treatment studies as an outcome measure to accomplish these goals. Also, the BBRS-D should be administered in different time increments (i.e., daily, weekly, multiple times per week, etc.) to determine its change-sensitivity with different administration intervals and its appropriateness in a variety of settings.

Fourth, combining items from the BDI-Y and CDI into a single measure may necessitate the merging of two different scales of measurement. The BDI-Y uses a 4-point Likerttype scale (i.e., *never, sometimes, often*, and *always*) while the CDI uses a 3-point scale in which one of three statements are selected as true of the respondent. Currently, we do not know what effect any changes have on the BBRS-D and future research should investigate whether these modifications substantially alter its psychometric properties. Research examining item wording

changes across other measures (e.g., Sheridan, Eagle, Cowan, & Mickelson, 2001) have revealed that the psychometric properties are not substantially changed, suggesting this may not be a serious concern; however, the question remains.

Fifth, as we mentioned previously, we were unable to establish the reliability and internal consistency of the four item BBRS-D. One explanation for this may be that the BBRS-D included items that are maximally sensitive to change, making it difficult to demonstrate stability over the 11-week test–retest period. Although our definition of change-sensitivity was based on change in response to treatment, it is possible that these items are very sensitive to other environmental changes, explaining the less than adequate reliability coefficients.

Finally, the process by which items were selected was based purely on empirical data and was not theoretically derived. As such, the process may have led to the identification of items that did not measure the full breadth of depression symptoms, resulting in a measure that would not adequately indicate a variety of depressive symptoms. The process developed by Gresham and colleagues (2010) and Cook and colleagues (2013), and in this study, includes content validation subsequent to the change-sensitivity analyses (i.e., the resulting BBRS is compared to the scales from which it is derived); however, the individual items are not chosen based on a theoretical understanding of depressive symptomology. A related issue is that, due to the lack of data on other types of problem behaviors in the dataset (e.g., externalizing behavior, hyperactivity, anxiety, etc.), we were unable to establish divergent validity of the BBRS-D with scales with which it should not agree. As such, future research should determine whether empirically derived or theoretically derived scales, or some combination of the two processes, result in the most methodologically rigorous but theoretically appropriate depression measures.

Conclusion

The development of the BBRS-D is a preliminary contribution to the literature on school-based mental health, as well as progress monitoring measure development in schools. In the context of school-based mental health, a four-item depression measure would likely be feasible as a progress monitoring tool for female youth within a

MTSS framework. Also, given its brevity, the BBRS-D would be much more applicable for repeated use in schools than longer behavior rating scales. The development work on the BBRS-D is also important in the context of progress monitoring measure development, as this is another application of the model described by Gresham and colleagues (2010) and Cook and colleagues (2013). BBRSs have now been developed for social skills and depression, and the procedures for developing the BBRS can be applied to additional behaviors. This work would result in a suite of BBRS measures for a variety of problem behaviors in schools, which would result in technically adequate and feasible measures to progress monitor a number of target behaviors.

Preliminary data indicate that the BBRS-D holds promise as a progress monitoring tool for depressive symptoms in school-based female youth. This is particularly encouraging given the unique opportunities to treat depression with school-based mental health services. As such, the BBRS-D should continue to be developed and empirically validated, and its use in school-based intervention delivery should be examined. It will also be necessary to develop local or national norms, decision rules, and cut points for the BBRS-D, which would assist in the use of the instrument and interpretation of obtained scores. This work would result in the development of a technically adequate measure that can be used to monitor students' response to intervention and enable those supporting students who are struggling with depression to make timely and meaningful data-driven decisions.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Angold, A., & Costello, E. J. (2001). Epidemiology of depression in children and adolescents. In I. Goodyer (Ed.), *The depressed child and adolescent* (2nd ed. pp. 143–178). Cambridge, UK: Cambridge University Press.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Estimates of standard error when monitoring progress using alternate passage sets. *School Psychology Review*, *38*, 266–283.

Arora, P. G., Nastasi, B., & Leff, S. S. (2017). Rationale for the cultural construction of school mental health programming. *International Journal of School & Educational Psychology*, *5*, 141–151.

Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the national comorbidity survey-adolescent supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*, 37–44.

Beck, J. S., Beck, A. T., & Jolly, J. B. (2001). *Beck Youth Inventories*. San Antonio, TX: Psychological Corporation.

Carnevale, T. D. (2013). Universal adolescent depression prevention programs: A review. *The Journal of School Nursing*, *29*, 181–195.

Center for Behavioral Health Statistics and Quality. (2015). Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health (HHS Publication No. SMA 15-4927, NSDUH Series H-50). Retrieved from http:// www.samhsa.gov/data/

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

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.

Clarke, G. N., Lewinsohn, P. M., & Hops, H. (1990). *Instructor's manual for the adolescent coping with depression course* (The therapist manual and the adolescent workbook may be downloaded for free from the Internet at http:// www.kpchr.cwd.html). Eugene, OR: Castalia.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Cohen, P., Cohen, J., Kasen, S., Velez, C. N., Hartmark, C., Johnson, J., . . . Streuning, E. L. (1993). An epidemiological study of disorders in late childhood and adolescence—I. Ageand gender-specific prevalence. *Journal of Child Psychology and Psychiatry*, *34*, 851–867.

Cook, C. R., Rasetshwane, K. B., Truelson, E., Grant, S., Dart,

A. H., Collins, T. A., & Sprague, J. (2011). Development and validation of the student internalizing behavior screener: Examination of reliability, validity, and classification accuracy. *Assessment for Effective Intervention*, *36*, 71–79. doi:10.1177/1534508410390486

Cook, C. R., Volpe, R. J., & Delport, J. (2013). Systematic progress monitoring of students with emotional and behavioral disorders: The promise of change-sensitive brief behavior rating scales. In H. M. Walker & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders: Applications in schools* (pp. 211–228). New York, NY: Guilford Press.

Costello, E. J., He, J. P., Sampson, N. A., Kessler, R. C., & Merikangas, K. R. (2014). Services for adolescents with psychiatric disorders: 12-month data from the National Comorbidity Survey-Adolescent. *Psychiatric Services*, *65*, 359–366.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Cunningham, J. M., & Suldo, S. M. (2014). Accuracy of teachers in identifying elementary school students who report at-risk levels of anxiety and depression. *School Mental Health*, *6*, 237–250. doi:10.1007/s12310-014-9125-9

Dart, E. H., Arora, P. G., Collins, T. A., & Doll, B. (under review). Progress Monitoring Measures for Internalizing Symptoms: A Systematic Review of the Literature. *School Mental Health*.

David-Ferdon, C., & Kaslow, N. J. (2008). Evidence-based psychosocial treatments for child and adolescent depression. *Journal of Clinical Child & Adolescent Psychology*, *37*, 62– 104. doi:10.1080/15374410701817865 de Vet, H. C., Terwee, C. B., & Bouter, L. M. (2003). Current challenges in clinimetrics. *Journal of Clinical Epidemiology*, *56*, 1137–1141.

Doerfler, L. A., Felner, R. D., Rowlison, R. T., Raley, P. A., & Evans, E. (1988). Depression in children and adolescents: A comparative analysis of the utility and construct validity of two assessment measures. *Journal of Consulting and Clinical Psychology*, *56*, 769–772.

Faulstich, M. E., Carey, M. P., Ruggiero, L., Enyart, P., & Gresham, B. (1986). Assessment of depression in childhood and adolescence: An evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC). *The American Journal of Psychiatry*, *143*, 1024–1027.

Fuchs, L. S., & Fuchs, D. (2011). *Using CBM for progress monitoring in reading*. Washington, DC: National Center on Student Progress Monitoring.

Galambos, N., Leadbeater, B., & Barker, E. (2004). Gender differences in and risk factors for depression in adolescence: A 4-year longitudinal study. *International Journal of Behavioral Development*, *28*, 16–25.

Gresham, F. M., Cook, C. R., Collins, T., Dart, E., Rasetshwane, K., Truelson, E., & Grant, S. (2010). Developing a changesensitive 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.

Gould, M. S., King, R., Greenwald, S., Fisher, P., Schwab-Stone, M., Kramer, R., ... & Shaffer, D. (1998). Psychopathology associated with suicidal ideation and attempts among children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*, 915-923.

Hankin, B. L., & Abramson, L. Y. (2001). Development of gender differences in depression: An elaborated cognitive vulnerability-transactional stress theory. *Psychological Bulletin*, *127*, 773–796.

Hawken, L. S., Vincent, C. G., & Schumann, J. (2008). Response to intervention for social behavior: Challenges and opportunities. *Journal of Emotional and Behavioral Disorders, 16*, 213-225.

Herman, K. C., Merrell, K. W., Reinke, W. M., & Tucker, C. M. (2004). The role of school psychology in preventing depression. *Psychology in the Schools*, *41*, 763–775.

Huang, L., Stroul, B., Friedman, R., Mrazek, P., Friesen, B., Pires, S., & Mayberg, S. (2005). Transforming mental health care for children and their families. *American Psychologist*, *60*, 615–627.

IBM Corp. Released. (2012). IBM SPSS statistics for Windows (Version 21.0). Armonk, NY: IBM Corp.

Kazdin, A. E., & Weisz, J. R. (1998). Identifying and developing empirically supported child and adolescent treatments. *Journal of Consulting and Clinical Psychology*, *66*, 19–36.

Kovacs, M. (1992). *Children's Depression Inventory*. North Tonawanda, NY: Multi-Health Systems.

Lewinsohn, P.M., Roberts, R.E., Seeley, J.R., Rohde, P., Gotlib, I.H., & Hops, H. (1994). Adolescent psychopathology: II. Psychosocial risk factors for depression. *Journal of Abnormal Psychology, 103*, 302-315.

Maag, J. W., Swearer, S. M., & Toland, M. D. (2009). Cognitive-behavioral interventions for depression in children and adolescents: Meta-analysis, promising programs, and implications for school personnel. In M. J. Mayer, J. E. Lochman & R. Van Acker (Eds.), *Cognitive-behavioral interventions for emotional and behavioral disorders: School-based practice* (pp. 235–265). New York, NY: Guilford Press.

McCarty, C. A., Violette, H. D., Duong, M. T., Cruz, R. A., & McCauley, E. (2013). A randomized trial of the positive thoughts and action program for depression among early adolescents. *Journal of Clinical Child & Adolescent Psychology*, *42*, 554–563.

McIntosh, K., Ty, S. V., & Miller, L. D. (2014). Effects of schoolwide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions*, *16*, 209–218.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 980–989.

Merrell, K. W. (2013). *Helping students overcome depression and anxiety: A practical guide*. New York, NY: Guilford Publications.

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory*, *3*, 248–292.

Patel, P. G., Stark, K. D., Metz, K. L., & Banneyer, K. N. (2014). School-based interventions for depression. In M. S. Weist, N. A. Lever, C. P. Bradshaw & J. S. Owens (Eds.), *Handbook of school mental health* (pp. 369–385). New York, NY: Springer.

Renshaw, T. L., & Cook, C. R. (2018). Initial development and validation of the youth internalizing problems screener. *Journal of Psychoeducational Assessment*, *36*, 366–378.

Rones, M., & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychology Review*, *3*, 223–241.

Ruffalo, M., & Fischer, D. (2009). Using an evidence-based CBT group intervention model for adolescents with depressive symptoms: Lessons learned from a school-based adaptation. *Child & Family Social Work*, *14*, 189–197.

Sheridan, S. M., Eagle, J. W., Cowan, R. J., & Mickelson, W. (2001). The effects of conjoint behavioral consultation results of a 4-year investigation. *Journal of School Psychology*, *39,* 361–385.

Smucker, M. R., Craighead, W. E., Craighead, L. W., & Green, B. J. (1986). Normative and reliability data for the Children's Depression Inventory. *Journal of Abnormal Child Psychology, 14,* 25–39.

Stapleton, L. M., Sander, J. B., & Stark, K. D. (2007). Psychometric properties of the Beck Depression Inventory for Youth in a sample of girls. *Psychological Assessment*, *19*, 230–235.

Stark, K. D., Arora, P., & Funk, C. (2011). CBT for youth depression: Implications for training school psychologists to become cognitive behavioral therapists. *Psychology in the Schools*, *48*, 272–282.

Stark, K. D., Streusand, W., Arora, P., & Patel, P. (2011). Treatment of childhood depression: The ACTION treatment program. In P. C. Kendall (Ed.), *Child and adolescent therapy: Cognitive behavioral procedures* (4th ed.; pp. 190–233). New York, NY: Guilford Press.

Stice, E., Shaw, H., Bohon, C., Marti, C. N., & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: Factors that predict magnitude of intervention effects. *Journal of Consulting and Clinical Psychology*, *77*, 486–503. doi:10.1037/a0015168

von der Embse, N. P., Scott, E. C., & Kilgus, S. P. (2015). Sensitivity to change and concurrent validity of direct behavior ratings for academic anxiety. *School Psychology Quarterly*, *30*, 244–259.

Weisz, J. R., Thurber, C., Sweeney, L., Proffitt, V. D., & LeGagnoux, G. L. (1997). Brief treatment of mild-to-moderate child depression using primary and secondary control enhancement training. *Journal of Consulting and Clinical Psychology*, *65*, 703–707.