Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2019, Vol. 27(2) 110–118

DOI: 10.1177/1063426617733714
jebd.sagepub.com
SAGE

Profiles of Problem Behavior in Children With Varying Language Ability

Jason C. Chow, PhD[1] and Joseph H. Wehby, PhD[2]

Abstract

The purpose of this study was to explore the relationships between language behavior in students with or at risk of emotional and behavioral disorders. School-age students in K-4 grades (N = 300, 76% male) were sampled across three school districts. Students were grouped based on oral language ability and measured on teacher-rated subscales of problem behavior and direct observation of classroom behavior. Profile analyses revealed that language ability was associated with direct observation measures of classroom behaviors. Lower language was associated with higher rates of aggression, and higher language was associated with higher rates of academic engagement. Incongruent results relative to teacher-rated and directly observed behavior, limitations, and implications for future research are discussed.

Keywords

behavior(s), language, disorders/disabilities, measurement

Language forms the necessary foundation for academic and behavioral success in school contexts. Understanding language is required for students to benefit from academic instruction that is presented orally, and comprehension of verbal instruction is necessary to be able to respond appropriately. Although a general significant association between language and behavior has been estimated (Chow & Wehby, 2016), there has been increasing recognition of the linguistic deficits of students with emotional and behavioral disorders (EBD). A deeper understanding of the relation between language and behavior in students with EBD is important, because the co-occurrence of language deficits and behavioral disorders is detrimental to students' academic and social outcomes (Im-Bolter & Cohen, 2007). Descriptive literature has demonstrated trends of these often comorbid deficits in both children identified with language impairment and EBD. For example, children with language deficits are 10 times more likely to present antisocial behaviors than their peers (Donahue, Cole, & Hartas, 1994; Warr-Leeper, Wright, & Mack, 1994) and twice as likely to develop future internalizing, externalizing, and attention deficit hyperactivity disorder (ADHD) symptoms (Yew & O'Kearney, 2013). However, Camarata, Hughes, and Ruhl (1988) estimated that 97% of students with emotional disturbance (ED) fell at least one standard deviation below the mean on the *Test of Language Development* (TOLD-I), and Nelson, Benner, and Cheney (2005) estimated that 68% of students with ED also had clinical language deficits. Furthermore, in a narrative review of 26 studies, 71% of students with EBD also experienced clinical language deficits, and 57% of students identified with language deficits also identified with EBD (Benner, Nelson, & Epstein, 2002). These trends suggest that students who exhibit problem behavior often show meaningful language deficits, and students with language deficits are often identified to have comorbid behavior problems or are at significantly higher risk of manifesting later behavior problems as compared with their peers.

More recently, and of particular concern, researchers have estimated that 81% of students with EBD had unidentified comprehensive language deficits, and 47% of those were moderate to severe (Hollo, Wehby, & Oliver, 2014). This alarming finding suggests that many students with EBD may have meaningful language deficits that are going undetected, and thus, untreated. Research also has shown that language deficits are often misinterpreted as a myriad of other indicators, including deficient cognitive ability, inattention, noncompliance, and intentional dishonesty and disrespect (Cohen, Davine, Horodezky, Lipsett, & Isaacson, 1993; Donahue et al., 1994). Therefore, research beyond prevalence rates of comorbidity of disorders is needed to help inform the design of interventions that specifically aim to improve academic and social outcomes for children who present both language and behavior problems. One area of promise involves deepening our understanding of what

[1]Virginia Commonwealth University, Richmond, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Jason C. Chow, Department of Counseling and Special Education, Virginia Commonwealth University, 1015 W. Main Street, Richmond, VA 23284, USA.
Email: jcchow@vcu.edu

types and topographies of behavior are more strongly associated with language ability.

Behavior Measurement

To date, the literature that examines the overlapping symptomology of language and behavior problems has used rating scales or special education status to index problem behavior. Although these methods are established and widely used, they may not be conducive to capturing specific types of behaviors that teachers may consider problematic in classrooms. That is, although rating scales allow for a more general assessment of behaviors across contexts, they may be less sensitive to behaviors where precision is important (e.g., frequency counts, intervals) such as frequency of aggressive behavior and duration engaged in academic instruction. These behaviors require more intensive, observational methods for measurement. Comparing the relation between these two types of frequently used behavior measures and language may provide insight into (a) what types of behaviors is language most related to and (b) the direction for future research and subsequent intervention.

Current Study

In the current study, we aim to identify and evaluate patterns of behavior in children ($N = 300$) who exhibit problem behavior in classrooms based on their language ability. Identifying differences in classroom behavior based on language ability is an important step to identifying targets for intervention or remediation that will best support the social and academic adjustment of this population of students. To do this, we identify subgroups of children based on cutoff scores of language abilities and examine patterns in teacher-rated problem behavior and directly observed classroom behavior. Through profile analysis, we can (a) elucidate patterns of both rated and observed behavior and (b) compare relative patterns between these language skill–based subgroups of children. Most studies have only reported rating scales of problem behavior, and are typically small-N samples. We aim to extend the current literature by providing an analysis of a large, multisite sample of students with or at risk of EBD, and by exploring and comparing how different levels of language relate to teacher-rated and observed behavior in the same sample of students.

Although this study is exploratory in nature, we do expect to see differential patterns in behavior based on language skill–based subgroups of children. Because low language skills are highly prevalent in children with EBD (Hollo et al., 2014), and there is a significant association between low language and problem behavior in the general school-age population (Chow & Wehby, 2016), we expect the behavioral profiles of children with or at risk of EBD to differ based on language ability.

Method

Participants

The data analyzed in the current study were collected as a part of a multiyear efficacy trial assessing the effects of a comprehensive teacher-level behavioral management intervention with an academic tutoring component. Data in this article were collected across two cohorts ($N = 300$) during their preliminary assessment wave. Students were recruited from both general and special education classrooms (K-4) from participating schools across three school districts. The current sample is a subset of a larger intervention study that screened in 422 original participants. Here, we utilize cases who had complete pretest data on all relevant variables. See Table 1 for full participant demographics.

Students were selected for inclusion into the sample via two methods. First, to identify at-risk students, the first two stages of the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1992) were used. Following consent procedures, general education teachers rank ordered consented students in their respective classrooms based on SSBD-defined characteristics for externalizing disorders. The top three ranked students in each classroom moved to Stage 2 of the SSBD. At this stage, teachers completed three brief rating scales—the Adaptive Behavior (AB) scale, the Maladaptive Behavior (MB) scale, and the critical events index (CEI). The AB scale is a 12-item scale that describes a student's current functioning with items such as "Follows classroom rules" and "Exhibits positive social peer interactions." The MB scale is an 11-item scale that describes a student's maladaptive behavior with items such as "Creates a disturbance during classroom activities" and "Refuses to participate with other children at recess." The CEI rates low-frequency, high-intensity events such as theft, setting fires, and property damage. Student who exceeded SSBD norms on the AB scale and MB scale or exceeded more than five critical events on the CEI were included in the study. Of the 300 participants in this study, 112 students in general education classrooms were identified as at risk based on the SSBD.

For the second approach, students receiving special education services in self-contained special education classrooms for students with behavioral difficulties were selected for participation, and, thus, 188 were included in the study because of their status as special education students. Given there is some evidence of differential eligibility patterns across schools and districts (e.g., Siperstein, Wiley, & Forness, 2011; Wiley, Siperstein, Bountress, Forness, & Brigham, 2008), we analyzed this as a single aggregated sample of at-risk and identified students that allowed us to leverage the variability in language and behavior scores and measures that were important characteristics for the current study's purpose.

Table 1. Demographic Characteristics of Included Sample (*N* = 300).

Sample	Grade	Sex	Ethnicity	Label	Lunch	ELL
SPED (63%)	9% K 11% first 14% second 19% third 10% fourth	55% M 14% F	53% Black 10% White 4% Hispanic 1% Other	32% emotional disturbance 20% developmental delay 12% learning disability 11% speech/language impairment 11% ADD/ADHD 5% intellectual disability 4% other health impairment 5% other	54% free 6% reduced 9% N/A	<1%
At risk (37%)	8% K 10% first 10% second 4% third 5% fourth	21% M 10% F	23% Black 4% White 2% Hispanic 3% Other		26% free 2% reduced 3% N/A	<1%
Total (100%)	17% K 21% first 24% second 23% third 15% fourth	76% M 24% F	76% Black 14% White 6% Hispanic 4% Other		80% free 8% reduced 12% N/A	<2%

Note. ELL = students receiving English language learning services; SPED = students receiving special education services; ADD = attention deficit disorder; ADHD = attention deficit hyperactivity disorder.

Procedure

We used multivariate profile analysis to examine a subset of measures from the larger intervention study. Trained research assistants (RAs) administered the *Woodcock Johnson Test of Achievement–Third Edition* (WJ-III; Woodcock, McGrew, & Mather, 2001) to obtain our measure of oral language. The project coordinator trained RAs on administration procedures. Training consisted of a 2-hr meeting that covered an assessment overview and description, administration, modeling, scoring, data entry, and reliability. The RAs then practiced administration, giving the test a minimum of three times to other RAs. To proceed to assessing participants, RAs were required to achieve 90% accuracy on administration and scoring. Teacher-rated behavior variables were collected via a standardized teacher rating scale, and RAs conducted all direct observations in classrooms.

Measures

Language. To estimate oral language ability, we use the WJ-III Oral Language Cluster—a composite of linguistic competency, listening ability, and comprehension. The cluster is an aggregate of the Story Recall and Understanding Directions subtests. In the present study, the oral language cluster was used as the between-subjects factor (group formation), and the individual subtests were used as within-subject dependent variables.

Behavior ratings and direct observation. Teacher ratings of behavior dimensions were collected via the Teacher Report Form (TRF; Achenbach & Rescorla, 2001). Direct observation of classroom behavior was collected via the Multiple Option Observation System for Experimental Studies (MOOSES; Tapp, Wehby, & Ellis, 1995), a flexible electronic system that is widely used in applied research to collect real-time observational data.

TRF. The TRF is a 120-item checklist used to rate children (ages 4–8) on syndrome scales for anxious and withdrawn, somatic complaints, thought and attention problems, rule-breaking and aggressive behavior, and conduct problems. To complete the scale, respondents rate the degree on a scale of 0 (*not true*), 1 (*sometimes true*), or 2 (*true*). In the present study, we use the TRF for both group formation and as a within-subjects dependent measure. Alpha for the TRF is .82 for the U.S.-normed sample.

Direct observation. We observed students for four 15-min sessions during language arts instructional activities. Each observation occurred on a separate day. The following measures represent proportions or frequencies of pooled observations. That is, for each student, the proportion or frequency of behavior represents 60 min of total classroom observation conducted across four observation periods. We accounted for any time a target student was out of the classroom during an observation in the denominator when calculating proportions.

Prior to live classroom data collection, all observers were trained to code student behaviors to a minimum of 80% interobserver reliability on three consecutive observation sessions when compared with a master data file for each of the five classroom demonstration videos. Observers

were then required to complete three consecutive practice sessions in classrooms with a master coder with 80% interrater agreement. Interobserver agreement (IOA) during training was calculated automatically by the MOOSES software based on the number of agreements divided by the number of agreements plus disagreements within a 3-s window multiplied by 100 to obtain a proportion of overall agreement for each event-based behavioral code. Agreement for duration measures was calculated on a second-by-second basis.

Engagement. Engagement represents the proportion of pooled sessions engaged. Engagement is defined as the target student is appropriately engaged in working on assigned/approved activities. Characteristics of this behavior include (a) attending to the material and the task, (b) making appropriate motor responses (e.g., writing, following directions, orienting to the teacher or speaker), (c) asking for assistance in an appropriate manner (e.g., raising hand), and (d) waiting appropriately for the teacher to begin or resume instruction.

Active responding. Active responding is indexed by the proportion of instructional stimuli appropriately responded to. Examples of active student responses include works read orally, questions answered, and words written. The proportion is calculated by dividing the number of active responses to instructional stimuli by the total number of instructional stimuli directed at the target child (individual and group).

Negative talk. Negative talk represents the number of statements or vocalizations made with the intent to provoke, annoy, pester, mock, whine, complain, tattle, or make fun of another. This includes threats of physical aggression against another individual or property, arguing or disagreeing with another (e.g., protest), as well as any verbal refusal to comply with a demand.

Aggression. Frequency of aggression is defined as the number of deliberate physical contacts that are potentially harmful to self, others, or property during observation sessions. Aggression was also coded for posturing or gesturing with the intention of provoking another individual.

Agreement estimates were obtained for 20% of all observation sessions. IOA was conducted via live reliability sessions. IOA for engagement, active responding, negative talk, and aggression were 89.9%, 91.8%, 100%, and 98.5%, respectively.

Data Analysis

Preliminary analyses and variable transformations. Table 2 includes the means, standard deviations, and correlations for the sample of 300 students of language and behavior measures; all analyses in the present study were conducted in Stata 12 (StataCorp, 2011). Direct observation–dependent variables were transformed into within-sample *z* scores ($M = 0.00$, $SD = 1.00$). We plotted the bivariate relations between language and each behavior construct–dependent measure. No outliers were identified; thus, all observations were included in further analyses. Then, we examined the distributions of each dependent variable to determine whether transformations were necessary. For frequency of aggression and negative talk, we used log-based transformations to help account for the expected moderate skew.

Language group formation. We used the Oral Language Cluster from the WJ-III to group cases into three groups. The low language group is defined as having a language cluster composite score greater than 1 *SD* below the mean (<85). Moderate language is defined as a composite score between 1 *SD* below the mean and the mean (85–100). High language, in the context of this sample, is defined as a composite score above the mean (>100). We selected these cutoffs because, due to the nature of the current sample and past literature, students with or at risk of problem behavior likely have lower average language scores than a normative sample. For example, in a comprehensive meta-analysis, the mean comprehensive language score for students with EBD across 22 studies was 76.33, which was significantly below average (Hollo et al., 2014).

Sociodemographic comparisons. To determine whether groups based on language ability were different, we used chi-square difference tests to examine whether groups were different by gender, special education status, free or reduced lunch, grade, and ethnicity. Ethnicity was the only sociodemographic variable that was significantly different between language groups ($p = .04$). Due to this significant difference, ethnicity was included as a covariate in all analyses.

Site- and school-level comparisons. First, we examined intraclass correlations (ICC) to determine whether multilevel modeling was necessary to account for clustering effects. With clustering effects, we would violate the independence assumption of ANOVA, potentially resulting in spurious significance (Raudenbush, Bryk, & Congdon, 2002). Site (i.e., district) accounted for less than 1%, and school contributed less than 5% unique variance across dependent measures. Using Raudenbush and Liu's (2000) standards that identify small, medium, and large as 5%, 10%, and 15%, we did not account for clustering and proceeded to the overall analysis.

Overall analyses. Because the effect of site and school contributed little unique variance in the sample, we conducted profile analyses using repeated measures ANOVA. Language was the between-subjects factor, and teacher ratings

Table 2. Sample Means, Standard Deviations, and Correlations (N = 300).

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12
1. EXT[total]	68.65	8.57	—											
2. INT[total]	61.79	10.28	.32	—										
3. Language cluster[a]	88.73	13.15	−.07	.06	—									
4. Affect[b]	62.64	8.34	.42	.83	.04	—								
5. Anxiety[b]	60.39	7.56	.26	.76	−.01	.56	—							
6. Attention[b]	65.88	9.76	.67	.13	−.16	.28	.23	—						
7. Oppositional[b]	65.78	7.60	.89	.26	−.01	.38	.23	.58	—					
8. Conduct[b]	67.99	11.16	.89	.20	−.08	.30	.12	.62	.71	—				
9. Engagement[c]	0.00	1.00	−.29	−.02	.23	−.10	−.13	−.29	−.28	−.26	—			
10. Active responding[c]	0.00	1.00	−.06	−.005	−.07	−.08	−.02	−.07	−.10	−.07	.23	—		
11. Negative talk[d]	0.00	1.00	.25	.06	−.10	.07	.11	.16	.29	.15	−.49	.001	—	
12. Aggression[d]	0.00	1.00	.11	.06	−.16	.03	.09	.02	.08	.08	−.49	−.02	.67	—

Note. EXT = total externalizing *t* score; INT = total internalizing *t* score (Teacher Report Form; Achenbach & Rescorla, 2001).
[a]Group formation. [b]Subscales of teacher ratings of behavior. [c]Direct observation proportion. [d]Direct observation frequency.

of behavior and directly observed behavior were the within-subject factors. After examining level and shape effects, we conducted univariate follow-up contrasts to explore how profiles differed by group.

Results

We aimed to examine profiles of teacher-rated and observed classroom behavior based on student language ability. With teacher-rated behavior subscales of problem behavior as the dependent variables, we found no significant differences between language groups, $F(2, 327) = 2.05$, $p = .130$; however, when averaged across groups, mean scores of teacher-rated behavior was significantly different, $F(4, 324) = 28.44$, $p < .001$. This difference suggests that language does not differentiate between ratings of separate subscales, but that teachers rate students' behavior differently on average based on language group (see Figure 1).

Next, we examined profiles of students based on language on direct observation outcomes. Here, we aimed to determine whether patterns of direct observation data differed from patterns of teacher-rated behavior (first analysis; see Figure 1) in groups formed based on language ability. The first dependent variables we examined were positive classroom behaviors (engagement and active responding). When aggregated across language groups, there was no significant main effect, $F(1, 327) = 0.036$, $p = .849$. However, this analysis yielded a significant shape effect, $F(2, 327) = 5.19$, $p = .006$, suggesting a Language × Positive Behavior interaction (see Figure 2). That is, group mean scores depend on language, and there is a significant difference between groups on positive classroom behavior. Students in the high language group are associated with the highest mean within-sample z score, whereas the low language group is associated with the lowest z score. Students with the lowest language were engaged in class for the least amount of time during observation sessions. Alpha-adjusted contrasts on individual-dependent variables reveal that the low language group is significantly different from both medium ($p = .04$) and high groups ($p = .009$) on engagement.

For negative classroom behavior, groups based on language ability differ significantly when negative behaviors are averaged, $F(2, 327) = 3.79$, $p = .024$. The effect for shape was not statistically significant, $F(4, 654) = 2.087$, $p = .081$; however, visual interpretation of Figure 2 suggests potential practical differences. Students with low language ability are associated with more negative behavior; specific to the high language group, their mean level of negative talk is highest across groups.

Discussion

The aim of this article was to examine the behavioral profiles of students with or at risk of EBD with varying language ability and to explore differences between profiles of teacher-rated and directly observed behavior. Because students with behavior problems have been associated with significant linguistic deficits, it is important to understand how language is associated with different types and measures of classroom behavior. The current findings suggest that language ability may underpin behavioral phenomena that students exhibit in classrooms, as well as how students experience their classroom environments. Noteworthy differences between teacher ratings and observed behavior imply a mismatch between how behaviors and language impact teacher perceptions and how behaviors manifest themselves in relation to researcher observations.

Profiles of observed classroom behaviors based on language groups provide some interesting points for discussion. First, with positive behavior–dependent variables,

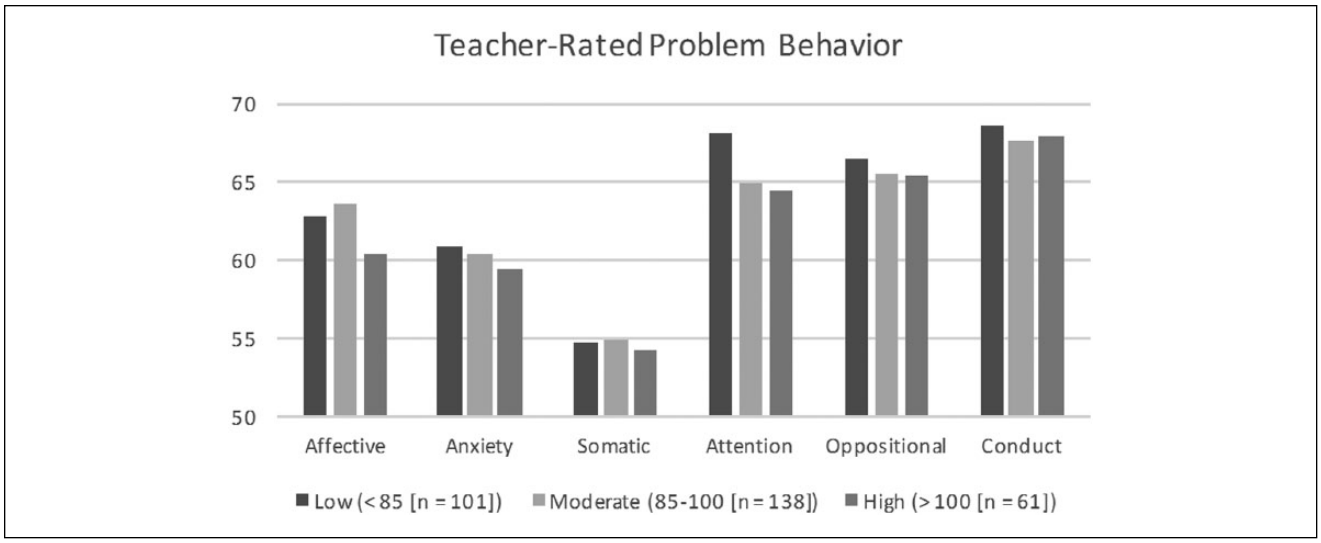

Figure 1. Subscale *t* scores of the TRF based on language scores.
Note. TRF = Teacher Report Form.

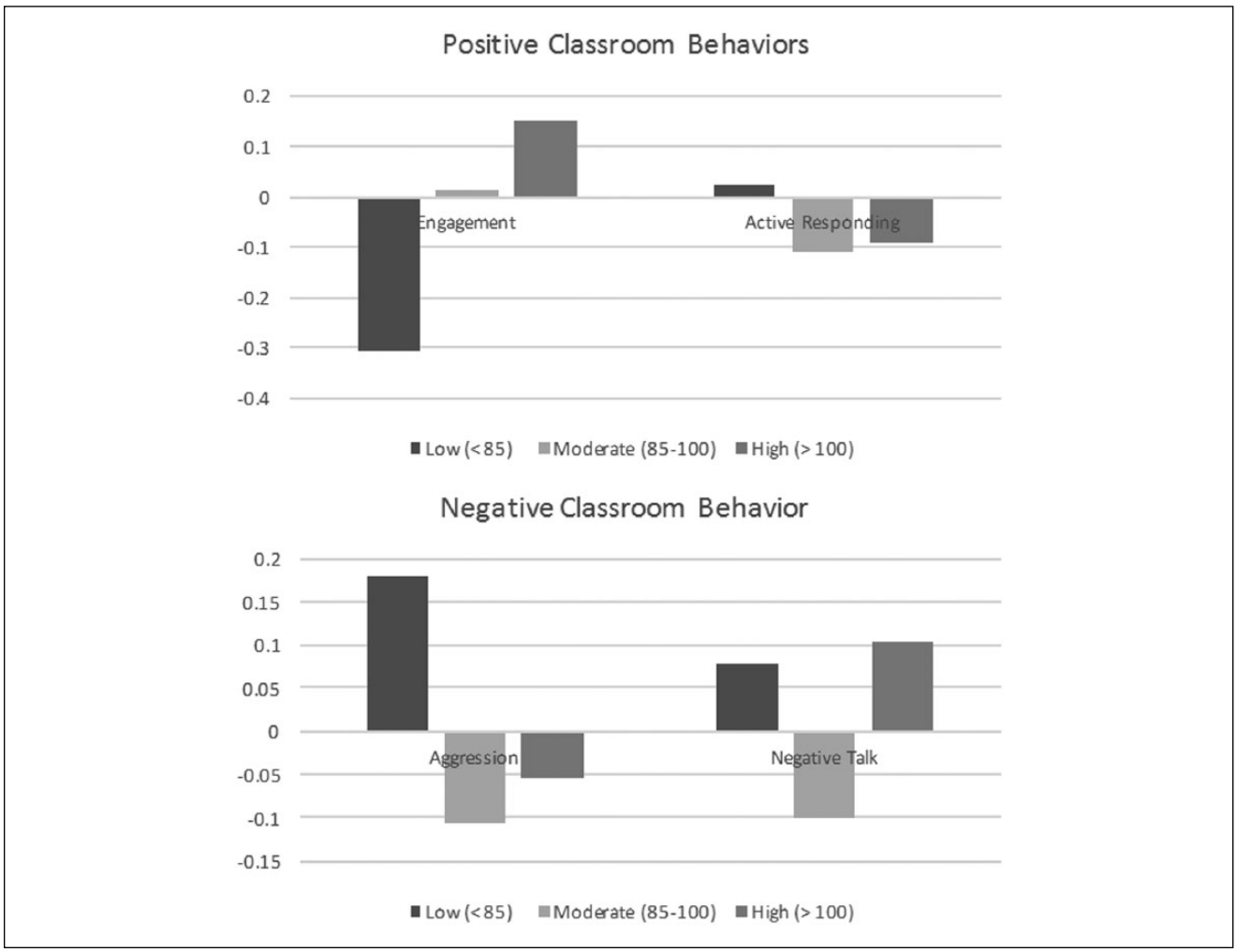

Figure 2. Profiles of positive and negative classroom behaviors; within-sample *z* scores are on the *y* axis.

students with high oral language were engaged for the highest proportion of pooled observations, and students with low language were engaged for the least amount of time. In this study, language group membership predicted rates of general engagement during academic instruction, but did not predict response to opportunities to respond. It is likely that language skills are needed to self-regulate and actively participate in the classroom, which manifests itself in higher rates of general engagement. Conversely, because opportunities to respond are isolated events that are teacher initiated, it may be less likely that children need higher language skills to attend to a teacher-delivered prompt.

Relative to negative classroom behaviors, the low language group is associated with higher frequency of aggressive behavior. Both moderate and high language groups are below the mean of the full sample. However, for frequency of negative talk, high language group membership is associated with the highest level of this behavior. Although minimal language ability is needed to engage in negative talk, these data suggest the possibility that the topography of problem behavior for students with higher language may more frequently be verbal than for students with lower language skills. Students who have more developed language skills may resort to verbal aggression when engaging in problem behaviors to access attention or escape from academic tasks. In addition, students with higher language skills may appear to be less problematic to teachers as they appear more engaged and be viewed as less aggressive.

In addition, comparing the relative association between oral language and measures of teacher-rated behavior and directly observed behavior is an important step in moving forward with decisions and subsequent interpretation of the measurement of behavior. Although behavior observation has been long considered a superior method for measuring behavior (Johnston & Pennypacker, 2010; Walker, 2015), observational coding systems are subject to methodological and measurement problems (Yoder & Symons, 2010). Relative to teacher ratings, there are inconsistencies in the literature. In a longitudinal analysis of six large-scale data sets, adult-rated problem behavior did not emerge a significant predictor of later achievement (Duncan et al., 2007). Conversely, some studies have reported that teacher ratings predict achievement outcomes relatively well. This may be, in part, due to the definitions of "outcomes." For example, Lane, Menzies, Kalberg, and Oakes (2012) used grade point average and course failure as their achievement outcome. The current study uses subtests from the WJ-III. Teacher ratings may be a better predictor of outcomes that are tied more closely to teachers (e.g., grades) than standardized assessments. In a study comparing language assessments of children with ED, Chow and Hollo (2017) reported meaningful discrepancies between teacher-rated language ability and standardized assessments of language. Teachers were much more likely to underestimate than overestimate the presence of language risk in students with ED. Along these same lines and when thinking about measuring behavior, teacher ratings may not be tapping the same underlying construct as direct observation of behavior.

The current study provides descriptive evidence that oral language may be more strongly associated with observed classroom behavior than teacher ratings. It is possible that teacher ratings of behavior and direct observation may not measure the same construct, because the analysis with teacher-rated behavior as the within-subjects factor did not produce significant level or shape effects, whereas the direct observation within-subjects factor analyses did. Also, these findings provide additional evidence that teachers may be unaware or untrained to notice the language deficits that are associated with many children who have identified behavioral difficulties and do not account for these differences in their instruction. Although there appeared to be no significant difference in the percentages of active engagement by students (i.e., opportunities to respond responded to by students), it was not coded whether these instructional prompts were linguistically different based on student language ability. If not, such a finding would be additional evidence of an instructional mismatch between these groups of students.

When comparing teacher ratings to direct observation of behavior, it is important to consider the substantive difference. Teacher ratings are typically checklists that are completed outside of an active instructional context, whereas direct observations are usually conducted by a trained observer in the classroom, often when the teacher is engaged in instruction. Thus, these measures provide very little basis on which to make qualitative comparisons, and, findings from the present study suggest that teacher ratings and direct observation of behavior may potentially be measuring different constructs of behavior. However, one potential interpretation is that rating scales are more reflective of a teacher's overall impression of student behavior while observations are capturing levels of behavior at a more molecular level. In addition, it is unclear how intensity of behavior affects both ratings and observations. For example, low-frequency/high-intensity behavior events may skew teacher ratings toward a more negative rating. Future measurement studies should use multiple measures of both observed and teacher-rated student behavior to explore issues relative to agreement and construct validity.

Taken together, however, our data reveal that what teachers perceive as problematic behavior or report on as important negative behaviors may not align with how researchers define behavior problems in classrooms. Observational coding systems are complex, precise, and often are employed for brief, systematic periods of time within the context of a research study or an intervention program. Teacher ratings, however, likely draw from a broader perspective a student's behavior, and may potentially be biased depending on the

behavioral characteristics of peers and their classroom composition. That is, how a teacher rates a student's behavior may depend on other student's level of problematic behavior, as well as a teacher's personality, perspective, and past teaching experiences. Conversely, as designed, direct observation of behavior provides a discrete characterization of the behaviors present during observation.

Limitations and Future Directions

Although this exploratory study suggests important implications for the field, our findings should be interpreted with a certain level of caution. First, our data are concurrent and correlational. A limitation of the direct observation measures is that the distribution of frequency counts for aggression and negative talk were moderately skewed and required transformation. During observations, several students included in the analysis did not exhibit any aggression or negative talk. We included these exploratory analyses because, although with limitations, the data suggest a potential association between language ability and to topography of problem behavior. Because the association between verbal proficiency and verbal aggression has been speculated in toddlers (Dionne, 2005), the present study offers preliminary evidence of this association in school-age children. If this research is replicated and students with lower language are more likely to engage in nonverbal types of problem behavior, a closer look at the link between language and the behavioral function may be merited.

Another consideration that is important when conducting research with diverse samples is that our sample was predominantly African American, and it is important to consider this in implications of our findings. The WJ-III Oral Language Cluster indexed language ability in the current study. This measure may not be best suited for capturing language proficiency and communicative ability. However, the urban composition of the school districts may mitigate any general concern that the overall sample demographics limit our contextualized findings. Future studies should consider using additional measures, including rating scales of communicative ability that may be less prone to underestimation as well as measures that are designed to account for dialectical variation. Relative to the diversity of our sample, we included students with or at risk of EBD. The at-risk group was screened in using the SSBD, whereas students who were receiving special education services for problem behavior were automatically enrolled in the study. We included a combined sample of students with or at risk of EBD to leverage the likely variability in language scores that allowed for us to examine our research questions of interest. In addition, as noted in the introduction, there is some evidence (Wiley et al., 2008) that there may be variability by districts and individual schools on who is identified as needing special education services for EBD. Although the nonsignificant chi-square difference test for special education status provides support for combining the at-risk and special education students, future research should examine differences between at-risk groups of students and those already receiving special education services for behavior problems.

Relative to measurement, we used a broad cluster score to create language groups. Cluster subtests require the student to listen to a passage or set of directions. Story recall includes expressive language, where students verbally recall details of passages. Our measure of language may also be tapping working memory, as opposed to a measure of pure receptive vocabulary (see MacDonald & Christiansen, 2002), and future studies should consider control for verbal working memory. Finally, we conducted tests for demographic group differences based on behavior and language. Language groups were equivalent on all demographic variables except for ethnicity, which was significantly different ($p = .04$). Although we expected that our groups would be significantly different in ethnic composition because the groups were based on oral language skill, this is important to consider when interpreting these results.

The present exploratory study extends the literature by examining profiles of teacher-rated and directly observed behavior of students with or at risk of EBD. The goal was to determine whether groups based on language ability produced significantly different profiles on teacher-rated problem behavior and direct observation of classroom behavior. This study corroborates previous research that there is meaningful association between language ability and behavior ratings. However, our study adds to the existing literature by demonstrating significant differences in observed behavior based on language ability. Overall, both general education and special education teachers of students with EBD should be aware of these co-occurring and significant deficits, and work with qualified clinicians (e.g., speech-language pathologists, school psychologists) to actively integrate linguistic and behavioral support. Future researchers should actively consider the type of measures they use to index problem behavior when further probing the relation between language and behavior, and during subsequent development of much-needed intervention.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Research and preparation of this manuscript was supported in part by the Institute of Educational Sciences grant (H324P040013) and the Office of Special Education Programs Leadership Training grant (H325D100027). The opinions expressed in this article are

those of the authors and do not necessarily reflect those of the funding agencies.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: Research Center for Children, Youth, and Families, University of Vermont.

Benner, G. J., Nelson, J. R., & Epstein, M. H. (2002). Language skills of children with EBD: A literature review. *Journal of Emotional and Behavioral Disorders*, *10*, 43–56. doi:10.1177/106342660201000105

Camarata, S. M., Hughes, C. A., & Ruhl, K. L. (1988). Mild/moderate behaviorally disordered students: A population at risk for language disorders. *Language, Speech, and Hearing Services in Schools*, *19*, 191–200. doi:10.1044/0161-1461.1902.191

Chow, J. C., & Hollo, A. (2017). Language ability of students with emotional disturbance: Discrepancies between teacher ratings and direct assessment. *Assessment for Effective Intervention*. Advance online publication. doi:10.1177/15345084177020

Chow, J. C., & Wehby, J. H. (2016). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review*. Advance online publication. doi:10.1007/s10648-016-9385-z

Cohen, N. J., Davine, M., Horodezky, N., Lipsett, L., & Isaacson, L. (1993). Unsuspected language impairment in psychiatrically disturbed children: Prevalence and language and behavioral characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*, 595–603. doi:10.1097/00004583-199305000-00016

Dionne, G. (2005). Language development and aggressive behaviour. In R. E. Tremblay, W. W. Hartup, & J. Archer (Eds.), *Developmental origins of aggression* (pp. 330–352). New York, NY: The Guilford Press.

Donahue, M., Cole, D., & Hartas, D. (1994). Links between language and emotional/behavioral disorders. *Education and Treatment of Children*, *17*, 244–254.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., & Klebanov, P. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

Hollo, A., Wehby, J. H., & Oliver, R. M. (2014). Unidentified language deficits in children with emotional and behavioral disorders: A meta-analysis. *Exceptional Children*, *80*, 169–186.

Im-Bolter, N., & Cohen, N. J. (2007). Language impairment and psychiatric comorbidities. *Pediatric Clinics of North America*, *54*, 525–542. doi:10.1016/j.pcl.2007.02.008

Johnston, J. M., Pennypacker, H. S., & Green, G. (2010). *Strategies and tactics of behavioral research*. New York, NY: Routledge.

Lane, K. L., Menzies, H. M., Kalberg, J. R., & Oakes, W. P. (2012). A comprehensive, integrated three-tier model to meet students' academic, behavioral, and social needs. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA Educational Psychology Handbook: Vol. 3. Application to learning and teaching* (pp. 551–581). Washington, DC: American Psychological Association.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.

Nelson, J. R., Benner, G. J., & Cheney, D. (2005). An investigation of the language skills of students with emotional disturbance served in public school settings. *The Journal of Special Education*, *39*, 97–105.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2002). *Hierarchical linear modeling*. Thousand Oaks, CA: SAGE.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*, 199–213. doi:10.1037//1082-989x.5.2.199

Siperstein, G. N., Wiley, A. L., & Forness, S. R. (2011). School context and the academic and behavioral progress of students with emotional disturbance. *Behavioral Disorders*, *36*, 172–184.

StataCorp. (2011). *Stata statistical software: Release 12*. College Station, TX: Author.

Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*, *27*, 25–31. doi:10.3758/bf03203616

Walker, H. M. (2015). Perspectives on seminal achievements and challenges in the field of emotional and behavioral disorders. *Remedial and Special Education*, *36*, 39–44. doi:10.1177/0741932514554106

Walker, H. M., & Severson, H. H. (1992). *Systematic screening for behavior disorders (SSBD)*. Longmont, CO: Sopris West.

Warr-Leeper, G., Wright, N. A., & Mack, A. (1994). Language disabilities of antisocial boys in residential treatment. *Behavioral Disorders*, 159–169.

Wiley, A. L., Siperstein, G. N., Bountress, K. E., Forness, S. R., & Brigham, F. J. (2008). School context and the academic achievement of students with emotional disturbance. *Behavioral Disorders*, *33*, 198–210.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*, *54*, 516–524. doi:10.1111/jcpp.12009

Yoder, P., & Symons, F. (2010). *Observational measurement of behavior*. New York, NY: Springer.