A Preliminary Investigation of a Brief Tier 1 Classroom Management Measure

Allison F. Gilmour, PhD[1], Joseph H. Wehby, PhD[2], Jessica Boyle, MEd[2], Howard P. Wills, PhD[3], and Paul Caldarella, PhD[4] iD

Abstract

Multitiered systems for supporting students' behavior in schools rely on teachers using generally effective classroom management strategies, such as Tier 1 practices. However, few tools exist to easily assess classwide Tier 1 practices. In this brief report, we describe one potential tool, the Classroom Atmosphere Rating Scale–Brief (CARS-B) and assess the properties of the tool using item response theory and multilevel models. We found that most items on the CARS-B had acceptable discriminations and difficulty estimates. Teachers' CARS-B latent scores were positively associated with student group on-task. We discuss potential changes to the CARS-B and its use in research and practice.

Schools increasingly use multitiered positive behavioral interventions and supports (PBIS) models to address students' behavioral outcomes. Universal, Tier 1 supports are provided to all students and are implemented at the school-wide and classwide levels. A critical universal component at the classroom level is the extent to which teachers use evidence-based classroom management strategies. Researchers have demonstrated that effective classroom management practices predict student outcomes, such as academic performance and appropriate prosocial behaviors (Garwood & Vernon-Feagans, 2017; Simonsen et al., 2008). Although evidence supports the impact of classroom management practices on student outcomes, general and special educators often struggle to implement effective classroom management practices, particularly when instructing students with disabilities (Hirn & Scott, 2014; Hollo & Hirn, 2015). In addition, teacher preparation programs often exclude classroom management from their curriculum (Freeman et al., 2014; Oliver & Reschly, 2010), and teachers frequently report feeling ill-prepared to implement classroom management (Tillery et al., 2010).

Given the importance of Tier 1 classroom practices as the foundation for multitiered systems of support, there is a need for measures to assess these practices. There are widely used measures to assess *schoolwide* Tier 1 practices (e.g., Benchmarks of Quality [Kincaid et al., 2010]; Tiered Fidelity Inventory [Algozzine et al., 2019]; School-Wide Evaluation Tool [Sugai et al., 2005]), but these measures often do not explicitly assess *classwide* Tier 1 practices or provide specific, actionable items teachers could use to

improve their classroom management skills (Mathews et al., 2014). In addition, there seems to be limited reporting of classroom Tier 1 practices in the overall literature on tiered behavior support (Blair et al., 2021). Existing observational measures that assess general classroom management (e.g., the Classroom Atmosphere Scoring System [Pianta et al., 2008], the Brief Classroom Interaction Observation–Revised [Reinke et al., 2015]) and discrete teacher and student behaviors (Lewis et al., 2014) have limitations due to the extensive training and time necessary to complete the classroom observations in resource-limited school environments. There is a need for brief, resource-efficient measures, correlated with student outcomes, that can be used by researchers and administrators to identify teachers who may require additional support with managing their classrooms, both within and outside the context of implementing PBIS.

The purpose of this study was to examine the use of a brief classroom management tool, Classroom Atmosphere Rating Scale–Brief (CARS-B; Wehby et al., 1993), and its

[1]Temple University, Philadelphia, PA, USA
[2]Vanderbilt University, Nashville, TN, USA
[3]University of Kansas, Lawrence, USA
[4]Brigham Young University, Provo, UT, USA

Corresponding Author:
Allison F. Gilmour, Temple University, 1301 Cecil B. Moore Ave., RH 362, Philadelphia, PA 19122-6008, USA.
Email: allison.gilmour@temple.edu

Associate Editor: Shanna Hirsch

association with student behavior. The CARS-B is intended to assess teachers' use of generally effective (e.g., Tier 1) classroom management skills. We asked the following research questions:

Research Question 1: What are the difficulties and discrimination of each item on the CARS-B?
Research Question 2: Are teachers' scores on the CARS-B associated with student on-task behavior?

Method

Sample

We used observational data from 158 elementary general and special educators in 21 schools across three states in the United States. Teachers were participating in a randomized control trial of a classroom management intervention; all data included in our analyses were collected prior to intervention. We used data from the second baseline data collection session for each teacher in case of reactivity during the first data collection session. The majority of the teachers in our sample were female (95.5%) and White (83%). Some teachers were missing educational information (13%); of those reporting data, 50.7% had a bachelor's degree and 49.3% had a master's degree or higher. On average, the teachers worked in schools where 69% of the students qualified for free or reduced lunch and 55% of students were from minoritized backgrounds. More details about study recruitment and the sample are included in Wills, Wehby, et al. (2018) and Wills, Kamps, et al. (2018).

Measures and Procedures

We trained observers by reviewing definitions of the CARS-B items and the definition of group on-task (supplemental materials available online provide the complete CARS-B and on-task definition). The CARS-B is a classroom management tool adapted from the Classroom Atmosphere Rating Scale (CARS; Wehby et al., 1993). Observers obtained 80% agreement with two master coder–scored videos before conducting data collection in schools. During each 20-min observation, trained observers measured group on-task (described below) before completing CARS-B. The CARS-B included nine items: (a) level of compliance during academic instruction, (b) students follow rules appropriate to setting, (c) transitions are short with only minor disruptions, (d) students are focused and on task, (e) level of structure (organized clear directions, sufficient work to keep students busy), (f) teacher ignores minor inappropriate behaviors, (g) frequent and specific praise given, (h) praise ratio to reprimands approximately 4:1, and (i) three to five clearly and positively stated expectations/rules are visibly posted. At the end of

each observation session, observers rated each item on a 1 to 4 scale, with "4" corresponding to higher levels of the skill. When observers did not observe a transition, Item 3, the observers coded a "0," which was treated as missing in our analyses. We collected inter-observer agreement (IOA) for 21% of sessions. Average IOA was 95% (range = 47%–100%). We used item-level scores for our analyses. In our sample, Cronbach's α was .79.

We assessed student on-task behavior by measuring group on-task using a momentary time-sampling procedure. Students were divided into groups of two to six students based on proximity. Every 30 s, for 20 min, the research assistant would mark whether all members of the group met the definition of on-task or whether at least one member of the group did not meet the definition of on-task at that moment, with each group rated during each interval. We calculated a final score, the percentage of intervals on-task, by dividing the number of intervals on-task by the number of total intervals and multiplying by 100. The definitions and information about the procedures are included in Wills, Wehby, et al. (2018), Wills, Kamps, et al. (2018), and the supplemental materials online. Average on-task behavior was 56.06% (SD = 15.5; range = 12.5%–89.9%), with an approximately normal distribution. The IOA for 23% of sessions averaged 97% for teachers in the treatment group (range = 87%–100%) and 95% for teachers in the control group (range = 77%–100%). Fifteen teachers did not have on-task behavior data corresponding to their second data collection session.

Data Analysis

We used a polytomous item response theory (IRT) rating scale model (RSM) to obtain difficulty estimates for the CARS-B items. Next, we fit a graded response model (GRM) to obtain discrimination estimates for each item and difficulty estimates. We compared model fit using the Akaike information criterion (AIC) and Bayesian information criterion (BIC), with lower numbers indicating better fit. These models use maximum likelihood to handle missing data. We calculated theta scores from the RSM; unfortunately, we were unable to calculate theta scores from the GRM because of model convergence issues likely related to our sample size. Finally, we used a multilevel model to examine whether CARS-B theta scores were associated with on-task behavior, accounting for the nesting of teachers within schools. The final set of analyses included 143 teachers because of missing group on-task data. We completed all analyses using Stata 15.

Results and Discussion

We first compared the fit of the RSM and the GRM with the data. The AIC from the RSM was 3,141.57 and the BIC was 3,178.32; the AIC from the GRM was 2,828.75 and the BIC

Table 1. Discrimination Estimates From Graded Response Model.

Item	Discrimination	SE	z	p value	95% CI
Item 1	9.95	4.92	2.02	.04	[0.31, 19.59]
Item 2	5.59	1.17	4.79	.00	[3.30, 7.87]
Item 3	0.87	0.21	4.17	.00	[0.46, 1.28]
Item 4	4.79	0.83	5.79	.00	[3.17, 6.41]
Item 5	1.53	0.24	6.30	.00	[1.05, 2.00]
Item 6	0.60	0.18	3.41	.00	[0.26, 0.94]
Item 7	0.58	0.20	2.98	.00	[0.20, 0.97]
Item 8	0.62	0.21	2.94	.00	[0.21, 1.03]
Item 9	0.25	0.17	1.45	.15	[−0.09, 0.59]

Note. $N = 158$. CI = confidence interval.

was 2,939.01. The decline in the AIC and BIC supported that the GRM better fit the data. In Table 1, we report the discrimination estimates from the GRM, and in Table 2, we report item difficulties. Items 1, 2, and 4 had the highest discrimination estimates of 9.95, 5.59, and 4.79, respectively. Item 9 had the lowest discrimination estimate and was the only estimate that was not statistically significant. The difficulty results suggested that the rating options adequately differentiated between teachers on most items. The test information function, included in the supplemental materials online, suggested the CARS-B provided the most reliable estimates of ability for teachers with average classroom management and theta scores of 1 and –1.

Next, we examined whether CARS-B theta scores, calculated from the RSM model, were associated with group on-task. The association between CARS-B scores and group on-task was statistically significant and substantively meaningful. A one-logit change in CARS-B score was associated with a 11.7 percentage point change in the percentage of intervals on-task ($SE = 1.31$; $p < .001$; 95% confidence interval [CI] = [9.13, 14.26]). The pseudo R^2 was .32, suggesting that 32% of the variation in group on-task was accounted for by CARS-B score.

Overall, these findings suggest that the CARS-B is a promising tool for assessing teachers' Tier 1 classroom management practices, particularly because scores were associated with a meaningful behavioral outcome providing some evidence of convergent validity. Notably, Items 1, 2, 4, and 5 had the highest discriminations meaning that they best differentiated between teachers with higher and lower classroom management skills. These four items addressed the structure of the classroom. In contrast, Item 9 did not discriminate teachers on classroom management skills. This could be because the item addressed posted rules rather than teachers' behaviors related to the rules, such as providing pre-corrections related to the classroom rules. Most items performed well on difficulty with the different rating selections varying in difficulty. Item 6,

Table 2. Difficulty Estimates From Graded Response Model.

Item	Difficulty	SE	z	p value	95% CI
Item 1					
2 vs. 1	−0.777	0.103	−7.53	.000	[−0.979, −0.575]
3 vs. 2	0.298	0.092	3.22	.001	[0.117, 0.479]
4 vs. 3	1.651	0.165	10.00	.000	[1.327, 1.974]
Item 2					
2 vs. 1	−0.972	0.123	−7.90	.000	[−1.214, −0.731]
3 vs. 2	0.169	0.944	1.79	.074	[−0.016, 0.354]
4 vs. 3	1.503	0.154	9.79	.000	[1.202, 1.804]
Item 3					
2 vs. 1	−2.190	0.560	−3.91	.000	[−3.286, −1.093]
3 vs. 2	−0.365	0.255	−1.43	.153	[−0.865, 0.135]
4 vs. 3	1.963	0.467	4.20	.000	[1.047, 2.878]
Item 4					
2 vs. 1	−0.729	0.111	−6.54	.000	[−0.948, −0.511]
3 vs. 2	0.637	0.106	6.00	.000	[0.429, 0.846]
4 vs. 3	1.797	0.208	8.65	.000	[1.390, 2.205]
Item 5					
2 vs. 1	−1.896	0.291	−6.51	.000	[−2.468, −1.325]
3 vs. 2	−0.311	0.147	−2.11	.035	[−0.600, −0.023]
4 vs. 3	1.147	0.202	5.68	.000	[0.751, 1.542]
Item 6					
2 vs. 1	−4.213	1.253	−3.36	.001	[−6.668, −1.758]
3 vs. 2	−0.906	0.381	−2.38	.017	[−1.652, −0.160]
4 vs. 3	2.433	0.731	3.33	.001	[0.999, 3.866]
Item 7					
2 vs. 1	0.906	0.392	2.31	.021	[0.137, 1.674]
3 vs. 2	2.496	0.834	2.99	.003	[0.862, 4.130]
4 vs. 3	4.516	1.489	3.03	.002	[1.597, 7.434]
Item 8					
2 vs. 1	1.393	0.502	2.78	.005	[0.410, 2.376]
3 vs. 2	2.792	0.927	3.01	.003	[0.974, 4.610]
4 vs. 3	4.120	1.368	3.01	.003	[1.439, 6.802]
Item 9					
2 vs. 1	1.256	1.063	1.18	.238	[−0.828, 3.340]
3 vs. 2	4.185	2.914	1.44	.151	[−1.526, 9.897]
4 vs. 3	5.997	4.130	1.45	.146	[−2.098, 14.091]

Note. $N = 158$. CI = confidence interval.

ignoring minor inappropriate behaviors, captured the widest range of difficulties suggesting wide variability in teachers' abilities to address this item. Again, Item 9 performed most poorly with regard to difficulty; teachers across latent levels of classroom management did not have statistically significant probabilities of receiving scores on the item that corresponded to their ability.

There are some notable limitations to the present investigation. First, the sample size was relatively small, which resulted in our inability to calculate theta scores from the GRM. Second, the same observers who conducted the group on-task completed the CARS-B. The CARS-B scores

could potentially be influenced by observers completing group on-task. Third, the CARS-B was conducted by trained research assistants, not practitioners. Individuals, such as administrators, who have personal relationships with the teachers they rate might use the CARS-B differently than research assistants, and resulting scores could less accurately capture teachers' classroom management skills. Fourth, our sample only included elementary school classrooms. The CARS-B may perform differently in other settings.

Given that CARS-B theta scores were positively associated with students' on-task behavior, and accounted for substantial variation in on-task behavior, the measure could be helpful for screening teachers' Tier 1 classroom management skills when implementing a multitiered system of behavioral support. Given that many existing assessments of PBIS Tier 1 fidelity focus on the school globally, the CARS-B could be used by administrators or instructional coaches to identify teachers in need of more support and begin conversations about improving teachers' use of generally effective classroom management practices. Future research might assess if the number of items could be decreased, perhaps dropping Item 9 and combining the items related to praise that had lower discrimination estimates. Classwide Tier 1 assessments that are relatively simple to conduct are an important and often missing component of multitiered systems of support; the CARS-B could fill the need for such a tool in research and practice.

Declaration of Conflicting Interests

Funding

ORCID iD

Paul Caldarella [iD] https://orcid.org/0000-0002-0883-8890

Supplemental Material

Supplemental material for this article is available on the *Assessment for Effective Intervention* website with the online version of this article.

References

Algozzine, B., Barrett, S., Eber, L., George, H., Horner, R., Lewis, T., Putnam, B., Swain-Bradway, J., McIntosh, K., & Sugai, G. (2019). *School-wide PBIS tiered fidelity inventory*.

OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org

Blair, K.-S. C., Park, E., & Kim, W. (2021). A meta-analysis of Tier 2 interventions implemented within school-wide positive behavioral interventions and supports. *Psychology in the Schools*, *58*(1), 141–161. https://doi.org/10.1002/pits.22443

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, *37*(2), 106–120. https://doi.org/10.1177/0888406413507002

Garwood, J. D., & Vernon-Feagans, L. (2017). Classroom management affects literacy development of students with emotional and behavioral disorders. *Exceptional Children*, *83*(2), 123–142. https://doi.org/10.1177/0014402916651846

Hirn, R. G., & Scott, T. M. (2014). Descriptive analysis of teacher instructional practices and student engagement among adolescents with and without challenging behavior. *Education and Treatment of Children*, *37*(4), 589–610. https://doi.org/h4gm

Hollo, A., & Hirn, R. G. (2015). Teacher and student behaviors in the contexts of grade-level and instructional grouping. *Preventing School Failure*, *59*(1), 30–39. https://doi.org/10.1080/1045988X.2014.919140

Kincaid, D., Childs, K., & George, H. (2010). *Tier 1 benchmarks of quality* (Rev. ed.). https://assets-global.website-files.com/5d3725188825e071f1670246/60303de04f17fab0bc5d854c_Tier_1_Benchmarks_of_Quality.pdf

Lewis, T. J., Scott, T. M., Wehby, J. H., & Wills, H. P. (2014). Direct observation of teacher and student behavior in school settings: Trends, issues and future directions. *Behavioral Disorders*, *39*(4), 177–247. https://doi.org/10.1177/019874291303900404

Mathews, S., McIntosh, K., Frank, J. L., & May, S. L. (2014). Critical features predicting sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, *16*(3), 168–178. https://doi.org/f6kchn

Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders*, *35*(3), 188–199. https://doi.org/10.1177/019874291003500301

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System [CLASS] manual: Pre-K*. Brookes Publishing.

Reinke, W. M., Stormont, M., Herman, K. C., Wachsmuth, S., & Newcomer, L. (2015). The brief classroom interaction observation–revised: An observation system to inform and increase teacher use of universal classroom management practices. *Journal of Positive Behavior Interventions*, *17*(3), 159–169. https://doi.org/10.1177/1098300715570640

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*(3), 351–380. https://doi.org/10.1353/etc.0.0007

Sugai, G., Lewis-Palmer, T., Todd, A. W., & Horner, R. H. (2005). *School-wide evaluation tool* (Version 2.1). https://www.pbis.org/resource/school-wide-evaluation-tool-set

Tillery, A. D., Varjas, K., Meyers, J., & Smith Collins, A. (2010). General education teachers' perceptions of behavior management and intervention strategies. *Journal of Positive Behavior*, *12*(2), 86–102. https://doi.org/10.1177/1098300708330879

Wehby, J. H., Dodge, K. A., & Greenberg, M. (1993). *Classroom Atmosphere Rating Scale* [Unpublished Technical Manual]. Vanderbilt University.

Wills, H., Kamps, D., Caldarella, P., Wehby, J., & Romine, R. S. (2018). Class-wide Function-Related Intervention Teams (CW-FIT): Student and teacher outcomes from a multi-site randomized replication trial. *The Elementary School Journal*, *119*(1), 29–51. https://doi.org/10.1086/698818

Wills, H., Wehby, J., Caldarella, P., Kamps, D., & Romine, R. S. (2018). Classroom management that works: A replication trial of the CW-FIT program. *Exceptional Children*, *84*(4), 437–456. https://doi.org/10.1177/0014402918771321