Hammill Institute on Disabilities

Assessment for Effective Intervention
2023, Vol. 48(2) 100–112

DOI: 10.1177/15345084221111338
aei.sagepub.com
SAGE

A Teacher Self-Assessment of Culturally Relevant Practice to Inform Educator Professional Development Decisions in MTSS Contexts

Lindsay M. Fallon, PhD, BCBA-D[1], Sadie C. Cathcart, MEd[1], Austin H. Johnson, PhD, BCBA[2], Takuya Minami, PhD[1], Breda V. O'Keeffe, PhD[3], Emily R. DeFouw, MEd, BCBA[4], and George Sugai, PhD[5]

Abstract

When students require support to improve outcomes in a variety of domains, educators provide youth with school-based intervention. When educators require support to improve their professional practice, school leaders and support personnel (e.g., school psychologists) provide teachers with professional development (PD), consultation, and coaching. This multi-study article describes how the Assessment of Culturally and Contextually Relevant Supports (ACCReS) was developed with the purpose of assessment driving intervention for teachers in need of support to engage in culturally responsive practice. Items for the ACCReS were created via a multi-step process including review by both expert and practitioner panels. Then, results of an exploratory factor analysis with a national sample of U.S. teachers ($N = 500$) in Study 1 yielded three subscales. A confirmatory factor analysis conducted with a separate sample of teachers ($N = 400$) in Study 2 produced adequate model fit. In Study 3, analyses with another final sample of teachers ($N = 99$) indicated preliminary evidence of convergent validity between the ACCReS and two measures of teacher self-efficacy of culturally responsive practice. Data from the ACCReS can shape the content of educator intervention (e.g., PD) and promote more equitable student outcomes for youth.

Keywords

professional development, culturally responsive practice, instrument development

In the United States, it is projected that over the next decade, racially and ethnically minoritized youth (Proctor & Owens, 2019) will account for 56% of students enrolled in public elementary and secondary schools (National Center for Education Statistics [NCES], 2019a) while the teaching field remains predominately White and female (Hussar et al., 2020). This racial/ethnic "mismatch" can affect how teachers evaluate their students' abilities and behaviors, ultimately affecting students' experiences in school (La Salle et al., 2020). Teachers report graduating from their preservice training programs underprepared (Milner, 2017) and needing support in the classroom (Gregory et al., 2016) to bridge the long-standing opportunity (Bohrnstedt et al., 2015) and discipline gaps (Gopalan & Nelson, 2019) that affect Black, Latinx, and Native American students most acutely (e.g., Gage et al., 2019; Skiba et al., 2016).

Furthermore, teachers often enter the field without a firm understanding of their own biases and the impact students' culture has on their learning (Howard & Navarro, 2016; Peters et al., 2016). For teachers to be responsive to students' culture and foster an effective educational environment, intervention in the form of high-quality teacher professional development (PD), consultation, or coaching is critical (Ellerbrock et al., 2016). For intervention to be effective, assessment data reflecting teachers' perceptions of their use of culturally and culturally relevant classroom

[1]University of Massachusetts Boston, USA
[2]University of California, Riverside, USA
[3]The University of Utah, Salt Lake City, USA
[4]The University of Southern Mississippi, Hattiesburg
[5]University of Connecticut, Storrs, USA

Corresponding Author:
Lindsay M. Fallon, Department of Counseling and School Psychology, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA 02125, USA.
Email: lindsay.fallon@umb.edu

Associate Editor: Jiwon Hwang

supports would ensure teacher training targets the appropriate areas of need.

Culture in the Classroom

Culture refers to dynamic systems of social values, ways of thinking, standards of behavior, and beliefs, with race and ethnicity anchoring identity and expression (Gay, 2018). Culturally responsive teachers acknowledge and understand students' culture, and create a connected, relevant, supportive learning environment. Specifically, teachers using culturally relevant pedagogy (a) build curricula that reflect students' culture, (b) vary their teaching methods dependent on student need, (c) set high expectations for learning, (d) build authentic relationships with students, (e) are reflective in their thinking and practice, and (f) establish relationships with students and their families (Ladson-Billings, 1995). Culturally relevant and responsive practice have been linked to gains across behavioral (Fallon, Cathcart, et al., 2018), academic (Powell et al., 2016), and social-emotional (Castro-Olivo, 2014) domains, leading to positive long-term outcomes (e.g., higher achievement test scores, increased graduation rates; Cammarota & Romero, 2009). Although there is a dearth of research on the prevalence of culturally responsive practice in the classroom, recent reviews (Fallon et al., 2021a, 2022) synthesized the extent to which culturally responsive academic and behavioral practices have been implemented to promote outcomes for racially and ethnically minoritized youth.

Vincent and colleagues (2011) conceptualized culturally responsive multi-tiered systems of support (MTSS) to promote staff members' knowledge and self-awareness, as well as commitment to culturally relevant practice for equitable outcomes. Federal laws in the United States such as the Every Student Succeeds Act (2015–2016) and Individuals with Disabilities Education Improvement Act (2004) encourage educators to adopt an MTSS framework which emphasizes (a) high-quality instruction and behavioral supports for all students (Tier 1 support); (b) universal screening and frequent progress monitoring to determine which students require more intensive supports; and, for students who do, (c) providing more intensive support that matches the level of student need (i.e., Tier 2 and Tier 3 intervention; Sugai & Horner, 2020). Successful implementation requires enablers such as teacher buy-in, adequate resource allocation, and strong administrator support (Pinkelman et al., 2015). With these implementation enablers in place, teachers may be more available to integrate culturally responsive practice into MTSS, and ultimately, be better equipped to reflect on their classroom behavior related to culturally responsive practice.

Vincent and colleagues' (2011) culturally responsive MTSS model includes four domains. First, it calls for integration of universal *behavioral* and *academic practices* that are culturally relevant and empirically validated. The authors reference Cartledge and Kleefeld's (2010) guidance to teach social skills that (a) reflect students' experiences, (b) are aligned with family expectations, (c) are modeled by individuals sharing the students' background, and (d) are delivered in students' language. Second, Vincent and colleagues' (2011) model calls for use of *data* that are culturally and contextually valid for decision-making. Researchers have called in to question the subjective nature of many disciplinary incidents (Skiba et al., 2014); therefore, Vincent et al. (2011) describe the importance of involving educators, families, and other community partners from various backgrounds to cultivate these definitions and/or provide specific examples and non-examples to reduce the prospect of cultural bias.

Third, Vincent and colleagues (2011) call for selection of student *outcomes* that are culturally equitable and promote all students' success in school. These data might include determining the (a) number of individuals disciplined, (b) percentage of suspension and expulsions, and (c) the type of infraction and number of days of missed instruction for each racial/ethnic group. Finally, Vincent and colleagues' (2011) model includes coordinated *systems* of delivery that promote staff members' cultural knowledge and self-awareness. These systems may include time and resources for educators to engage in self-reflection, training, and/or coaching, both individually and collectively as a school staff.

Sugai and colleagues (2012) expanded the model presented in Vincent et al. (2011) to provide specific recommendations for culturally and contextually relevant MTSS that targeted the actual "look, feel and sound" (p. 204) of implementation. Central to these recommendations was reviewing data to guide decision-making, including the targets of intensive and ongoing PD. Intensive and ongoing PD offers teachers more than a "train and hope" approach targeting cultural appreciation activities (Finch, 2012) to first focus on (a) uncovering teachers' biases and building self-awareness; (b) constructing knowledge of cultural, linguistic, and racial diversity; and (c) developing cultural consciousness (Tanguay et al., 2018). Subsequently, PD can focus on changing teacher actions in the classroom, and ultimately aligning action across educators and the school community. To engage in this initial step, assessment is important for understanding teachers' perceptions and informs identification of specific in-service training topics. Practical and efficient assessment tools are needed for this aim.

Teacher Self-Assessment

Self-assessments are efficient to administer and are perceived to be less evaluative by teachers than other means of classroom instruction quality assessment (e.g., classroom observation; Biggs et al., 2008). Although teachers' responses may be influenced by social desirability bias (Fisher, 1993), explicit guidance about the purpose of data collection (e.g., to guide design of PD) can encourage

accurate self-reporting (Fallon, Sanetti, et al., 2018). Selecting validated instruments can also promote confidence in data collected. Minimally, this should include choosing a tool for which internal consistency and factor structure are supported by evidence (Debnam et al., 2015).

Of the few existing assessments that target teachers' cultural responsiveness and possess reliability and validity evidence, many are relatively narrow in scope. Some existing measures focus either exclusively on classroom management self-efficacy (e.g., Culturally Responsive Classroom Management Self-Efficacy Scale (CRCMSES), $\alpha = .97$; Siwatu et al., 2017) or teachers' instruction (e.g., Culturally Responsive Teaching Self-Efficacy Scale (CRTSES), $\alpha = .95$; Siwatu, 2007; Multicultural Efficacy Scale, $\alpha = .80$; Guyton & Wesche, 2005). One instrument, the Double Check Self-Reflection Tool ($\alpha = .65$; Hershfeldt et al., 2009), targets teachers' consideration of students' culture in instruction, as well as efforts to establish supportive relationships with students but does not inquire about teachers' use of data or access to systems of support to guide their efforts (e.g., training, resources). As more schools use MTSS to promote behavioral and academic outcomes, there is a need for a more comprehensive instrument to gauge teachers' cultural responsiveness (Sugai et al., 2012). Alignment with critical features of MTSS will promote efficiency in decision making regarding educator PD. The Assessment of Culturally and Contextually Relevant Supports (ACCReS) was created to serve this purpose.

Development of the ACCReS

The ACCReS was developed using a series of steps (Fallon et al., 2021b). Items were written before being reviewed by a panel of experts and teachers. This is briefly described below prior to the current study's purpose.

Generating items. The initial draft of the ACCReS was based on recommendations made in a comprehensive systematic literature review of culturally and contextually relevant practices and supports (Fallon et al., 2012). Specifically, item stems were added to specific practices recommended in the systematic review (Fallon et al., 2012). For instance, the practice "greet students daily" (which was associated with the recommendation to increase positive interactions), became "Each day, I personally greet all of my students." To ensure items reflected current literature, two follow-up systematic reviews were conducted (targeting instructional and behavioral support, respectively), extending the years of publications reviewed to 2020 (Fallon et al., 2021a, 2022). These subsequent reviews confirmed themes and recommendations found in the original study (e.g., include students' culture in instruction, partner with families).

Expert review of items. It was hypothesized that the 48 items developed based on the above review would align with the core four features of the culturally responsive MTSS model proposed by Vincent and colleagues (2011; see description in Introduction). Items were sent for review to 10 subject-matter experts (i.e., U.S. university professors of education) to evaluate content and face validity as well as item relevance (see Fallon et al., 2018). Experts also offered qualitative feedback including that one item was unclear and should be eliminated ("Critical self-reflection of the decisions I make in the classroom is helpful"), and to add three items ("I frequently ask students questions while I teach," "Students help me define class rules," "I model appropriate behavior for my students"). A teacher panel then reviewed the resulting 50-item instrument.

Teacher review of items. Five elementary school, five middle school, and six high school educators ($n = 16$) participated in the teacher panel. All teachers worked in public schools in the Northeast United States in which there was a large percentage of racially and ethnically minoritized youth. Most panelists were female (87.50%) and White (93.75%), aligning with national trends in teacher demographics (NCES, 2019b), and had a range of teaching experience (43.75% = 0–10 years; 56.25% = 11–15 years). Overall, teachers reported that the directions were clear. Several panelists suggested a revision of specific terms and identified certain items as confusing. These items were removed, and five suggested items were added (e.g., "I review academic data for trends that reflect disproportionality"), resulting in a 48-item instrument.

Purpose of Study

The purpose of this multi-study article is to provide evidence of the psychometric properties of the ACCReS based on an exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and a preliminary convergent validity analysis. The research questions and hypotheses were as follows:

RQ1. What factor structure emerges from conducting an EFA? Based on Vincent and colleagues' (2011) model of culturally responsive MTSS (i.e., pertaining to systems, practices, data, and outcomes), we hypothesized that ACCReS items would map onto a four-factor model.

RQ2. Do data from an independent sample analyzed with a CFA confirm the factor structure extracted in the EFA? We hypothesized that the model specified in the CFA, informed by EFA results, would demonstrate an adequate fit to the data.

RQ3. What reliability coefficients emerge for each factor identified during the CFA process? We hypothesized that reliability coefficients would indicate acceptable internal consistency.

RQ4. What evidence of convergent validity exists between the ACCReS and two similar measures of cultural responsiveness for teachers? We hypothesized that responses on the ACCReS would be positively and significantly correlated with responses on two similar measures of cultural responsiveness for teachers.

General Method

General Overview

We evaluated the psychometric properties of the ACCReS in three separate studies with independent samples of Grades K–12 school teachers in the United States. Study 1 presents results of an EFA. Study 2 presents results of a CFA and an evaluation of the ACCReS' internal consistency. Study 3 presents a preliminary exploration of convergent validity. Below, we describe the measures and methodologies applied across all studies. Methods and results unique to each study then follow.

Measures

Assessment of Culturally and Contextually Relevant Supports. Participants completed the ACCReS items on a 6-point Likert-type scale: *strongly disagree, disagree, somewhat disagree, somewhat agree, agree*, and *strongly agree*.

Demographic questionnaire. Participants were asked to respond to items about personal characteristics as well as items about their work credentials, experience, and setting.

Procedures

Recruitment. Qualtrics Panel Management Services was enlisted to recruit a national sample of teacher respondents in all studies. To participate, respondents were required to be employed as an elementary, middle, or high school teacher and were offered a US$10 gift card for taking part in the study. Qualtrics staff solicited participation from eligible teacher participants who had previously registered as panelists with Qualtrics. Use of a paneling service for recruitment ensured data efficiency as well as quality in recruitment and data collection. Incomplete responses or complete responses that took less than 3 min to produce were excluded from the data set.

Statistical analysis. R (version 1.1.423; R Core Team, 2016) was used for all factor analytic procedures, as well as to calculate descriptive statistics, reliability coefficients, and correlation matrices. The packages used to conduct analyses were ez (Lawrence, 2016), lavaan (Rosseel, 2012), MVN (Korkmaz et al., 2014), and rstatix (Kassambara, 2020; all packages are available by request from second author). The calculation of descriptive statistics provided insight into participant response patterns. Reliability coefficients were generated to examine internal consistency. McDonald's omega is reported due to its superiority to Cronbach's alpha when factor loadings are unequal (Trizano-Hermosilla & Alvarado, 2016). Also, coefficients $> .75$ were interpreted to indicate acceptable internal consistency (Reise et al., 2013). Finally, correlation matrices reflected Pearson's product–moment coefficients for the purpose of conducting a preliminary convergent validity analysis.

Study 1

Study 1 contains an EFA to identify factors underlying the ACCReS.

Method

Sample. The 500 respondents were predominately White (85.20%) and female (78.47%), consistent with national teacher trends (NCES, 2019b). Although most teachers indicated >25% of their students were racially and ethnically minoritized (see Table 1), national student trends indicate racially and ethnically minoritized youth make up 52% of students nationwide (NCES, 2019a).

Instrumentation. In Study 1, the ACCReS included 48 items: 11 hypothesized to align with the academic practices factor, 16 hypothesized to align with the behavior practices factor, nine hypothesized to align with the use of data and monitoring outcomes factor, and 12 hypothesized to align with the systems to support staff factor.

Statistical procedures. Items on the ACCReS produce ordinal data; however, with six response categories, estimation methods for continuous indicators were deemed acceptable (Rhemtulla et al., 2012). We used principal axis factoring (PAF) and oblimin rotation as we hypothesized factors were intercorrelated. Relationships between items were examined through review of correlation coefficients. High inter-item correlations can indicate that multiple items may be measuring the similar constructs and are thus redundant. Items found to be weakly related to all other components of the instrument may also be problematic (McCoach et al., 2013). To identify the number of factors to retain in the model, we first conducted a scree test and parallel analysis. Visual analysis of the scree plot of eigenvalues provided an estimate of the maximum number

Table 1. Demographic Data for Studies 1, 2, and 3.

	Study 1 ($N = 500$)		Study 2 ($N = 400$)		Study 3 ($N = 99$)	
Participant Demographic Variables	%	*n*	%	*n*	%	*n*
Teacher gender						
Female	78.47	390	71.21	282	83.84	83
Male	21.33	106	28.79	114	16.17	16
Nonbinary or other	0.20	1	0.00	0	1.01	1
Teacher race and ethnicity[a]						
White	85.20	426	79.25	317	77.78	77
Black or African American	5.00	25	10.00	40	11.11	11
Hispanic or Latinx	4.00	20	5.75	23	9.09	9
Other	8.6	43	10	40	8.08	8
Teacher years of teaching experience						
0–5 years	24.70	123	26.70	106	33.33	33
6–10 years	19.48	97	23.68	94	25.25	25
≥11 years	55.82	278	49.62	197	42.42	42
School community						
City	44.78	223	42.25	169	58.58	58
Suburban	35.54	177	37.00	148	32.32	32
Rural	19.68	98	20.75	83	10.10	10
Grades taught[a]						
Elementary (K–5th grade)	53.00	265	41.50	166	52.53	52
Secondary (6th–8th grade)	33.60	168	29.75	119	26.26	26
High school (9th–12th grade)	37.40	187	40.00	160	27.27	27
Percentage of racially and ethnically minoritized students in school						
0%–25%	40.68	203	38.84	155	27.27	27
26%–50%	17.84	89	20.80	83	22.22	22
51%–75%	17.64	88	17.79	71	26.26	26
76%–100%	15.63	78	15.29	61	20.20	20
Not sure	8.22	41	7.27	29	5.05	5

[a]Questions were "Check all that apply," so percentages may > 100%.

of factors to extract (Cattell, 1966). Parallel analysis estimated the number of factors to extract by identifying eigenvalues greater than those generated with random data. Consistent with the procedures used in the development of similar measures, we retained items that loaded ≥ .40 on one factor only, and if crossloadings were < .32 across factors (Spanierman et al., 2011; see Table 2).

Results

Results of the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was .96 and Bartlett's test of Sphericity was statistically significant ($p < .001$), providing a preliminary indication that the sample was adequate to conduct the EFA. Descriptive statistics indicated that responses to items were negatively skewed, implying that respondents tended to indicate favorable practices reflected in median response categories as follows: *agree* (28 items), *somewhat agree* (16 items), and *strongly agree* (4 items) (see Table 3 for mean, standard deviation, skew, and kurtosis for each item). However, standard deviations across items indicated reasonable variability in response choices. Based on review of factor loadings, 37 items were retained.

Factor selection. We hypothesized a four-factor solution based on the model of Vincent and colleagues (2011). However, initial assessments of factor structure through scree test and parallel analysis suggested a three- and five-factor solution, respectively. Therefore, we considered three-, four-, and five-factor solutions (Table S1 in online supplemental materials). The four-factor solution showed just two items loading on to the fourth factor without strong theoretical justification. This was also the case for the five-factor solution (i.e., two items loading on both the fourth and fifth factor without strong theoretical justification). The three-factor model, however, was supported by the scree test solution and (a) included factors with at least three items each, (b) demonstrated sufficient internal consistency (as

Table 2. Factor Loadings From Exploratory Factor Analysis.

Item		Factor		
		ECP	AIS	CCC
Items retained				
1	I use explicit instruction when I teach (e.g., clearly describe, model, and practice content with students).	**0.63**	–0.17	0.17
2	I differentiate instruction to support the different learners I teach.	**0.54**	0.18	0.13
3	I provide additional (or more intensive) academic support when a student needs it.	**0.59**	0.12	–0.02
4	I plan lessons that are designed to actively engage all learners when I teach.	**0.61**	–0.05	0.21
5	I listen actively to students when they express concerns.	**0.65**	–0.09	0.01
6	I engage in more positive interactions with students than negative interactions.	**0.73**	0.00	–0.02
7	I am consistent and fair when it comes to discipline.	**0.69**	0.01	–0.05
8	I explicitly teach social skills (e.g., ways to ask for help appropriately).	**0.41**	0.18	0.10
9	I explicitly teach students about my expectations for classroom behavior.	**0.67**	–0.02	0.03
10	Each day, I personally greet all of my students.	**0.50**	0.19	–0.09
11	I work to build a positive relationship with each student I teach.	**0.75**	0.04	–0.10
12	I deliver praise equitably in my classroom.	**0.55**	–0.01	0.05
13	I actively monitor all parts of the classroom.	**0.65**	0.11	–0.07
14	I ask families to help define my classroom expectations.	–0.06	**0.56**	0.05
15	I collect classroom data to inform the equity of my interactions across students (e.g., frequency and distribution of positive interactions).	0.04	**0.82**	–0.05
16	I collect classroom data to inform the equity of my disciplinary actions across students (e.g., evidence of consistent consequences administered).	–0.03	**0.66**	0.12
17	I review academic data for trends that reflect disproportionality (e.g., students of a certain race not achieving in mathematics vs. students from other groups).	–0.03	**0.66**	0.16
18	I seek professional development opportunities (e.g., attend conferences, workshops, trainings) to learn about how to engage in culturally and contextually relevant practice.	0.08	**0.58**	0.13
19	I request the resources (e.g., time, staff, training) I need to implement culturally and contextually relevant instruction.	0.00	**0.72**	0.16
20	I request the resources (e.g., time, staff, training) I need to implement culturally and contextually relevant behavior support.	–0.02	**0.68**	0.16
21[a]	I request to meet with support personnel (e.g., instructional coaches, lead teachers, consultants) to help me consider cultural and contextual factors that might affect how I teach.	0.02	**0.83**	0.02
22	I request to meet with support personnel (e.g., instructional coaches, lead teachers, consultants) to help me consider cultural and contextual factors that might affect how I support students' behavior.	0.04	**0.90**	–0.09
23	I meet with support personnel (e.g., instructional coaches, lead teachers, consultants) to help me to find evidence of disproportionality (e.g., racial, gender) in my classroom data.	0.03	**0.82**	–0.09
24	I talk to administrators in my building about accessing the resources I need to provide culturally and contextually relevant academic supports.	0.03	**0.65**	0.17
25	I seek the resources (e.g., time, access, translators) I need to partner with families to support students.	0.26	**0.43**	0.17
26	Culturally and contextually relevant instruction is important to how I teach.	0.04	–0.04	**0.73**
27	I know how to provide culturally and contextually relevant instruction.	0.10	–0.01	**0.69**
28	I modify the curriculum to be culturally and contextually relevant, when appropriate.	0.11	0.13	**0.59**
29	I consider students' culture when I decide on the type of instructional support I will provide.	–0.05	0.24	**0.61**
30	I understand that behavior may be context-specific (e.g., different behaviors may be more appropriate at home or school).	0.30	–0.20	**0.55**
31	I consider a student's culture when selecting a research-based intervention strategy.	–0.05	0.29	**0.57**
32	I self-assess my cultural biases regularly.	–0.01	0.07	**0.51**
33	I understand that some students are at risk for being disproportionally excluded from the learning environment (e.g., sent to the office, suspended, expelled).	0.16	–0.03	**0.44**
34	I gather information about my students' families (e.g., customs, languages spoken, cultural traditions).	0.16	0.23	**0.42**
35	I consider students' culture and language when I select assessment tools.	–0.07	0.12	**0.64**
36[a]	I know where to find information about culturally and contextually relevant academic practices.	0.09	0.22	**0.53**
37	I know where to find information about culturally and contextually relevant behavior management practices.	0.06	0.15	**0.51**

Note. Response options across all ACCReS items were presented on a 6-point Likert-type scale, and dummy coded for analysis (*strongly disagree* = 0, *disagree* = 1, *somewhat disagree* = 2, *somewhat agree* = 3, *agree* = 4, and *strongly agree* = 5).

Upon selection of the three-factor model, items were removed from the instrument if [a]high inter-item correlations ($r > .70$). See Table S1 in online supplemental materials. Factor 1 was named Accessing Information and Support (AIS), inter-item correlations $M = 0.56$, $SD = 0.08$; $\omega_h = 0.86$. Factor 2 was named Equitable Classroom Practices (ECP), inter-item correlations $M = 0.47$, $SD = 0.08$; $\omega_h = 0.87$. Factor 3 was named Consideration of Culture and Context (CCC), inter-item correlations $M = 0.47$, $SD = 0.09$; $\omega_h = 0.77$.

Table 3. Item-Level Descriptive Summaries From ACCReS Responses in EFA and CFA Teacher Samples.

	EFA sample				CFA sample			
Item	*M*	*SD*	Skew	Kurtosis	*M*	*SD*	Skew	Kurtosis
Q1	3.72	1.11	–0.88	0.79	3.73	1.15	–1.04	1.23
Q2	3.66	0.99	–0.95	1.70	3.70	1.05	–0.94	1.22
Q3	4.30	0.86	–1.41	2.68	4.19	0.93	–1.51	3.32
Q4	4.16	0.87	–1.03	1.35	4.19	0.89	–1.20	1.95
Q5	4.34	0.76	–1.03	0.95	4.32	0.85	–1.56	3.65
Q6	4.23	0.82	–1.17	2.21	4.25	0.87	–1.43	3.16
Q7	3.68	1.06	–0.92	1.17	3.69	1.10	–1.06	1.59
Q8	3.36	1.22	–0.82	0.43	3.43	1.23	–0.94	0.72
Q9	4.44	0.71	–1.45	3.68	4.36	0.83	–1.71	4.55
Q10	4.06	0.87	–0.90	1.21	4.11	0.91	–1.38	3.29
Q11	4.31	0.83	–1.21	1.85	4.21	0.89	–1.42	3.10
Q12	4.29	0.74	–1.09	2.40	4.23	0.86	–1.42	3.26
Q13	3.92	1.08	–0.98	0.84	3.92	1.14	–1.08	0.98
Q14	4.42	0.73	–1.05	0.44	4.41	0.84	–1.75	4.09
Q15	4.22	1.00	–1.43	2.08	4.00	1.14	–1.25	1.29
Q16	4.53	0.66	–1.24	0.97	4.41	0.81	–1.93	6.07
Q17	2.48	1.42	–0.06	–0.84	2.57	1.45	–0.04	–0.97
Q18	4.33	0.74	–0.87	0.24	4.25	0.88	–1.63	4.30
Q19	4.25	0.76	–0.97	1.29	4.21	0.89	–1.46	2.98
Q20	3.33	1.21	–0.68	0.23	3.39	1.21	–0.80	0.57
Q21	3.33	1.19	–0.68	0.33	3.36	1.21	–0.76	0.44
Q22	3.83	1.08	–1.00	1.22	3.96	1.10	–1.35	2.17
Q23	3.07	1.31	–0.44	–0.49	3.35	1.29	–0.74	–0.04
Q24	3.11	1.32	–0.48	–0.53	3.36	1.32	–0.73	–0.09
Q25	3.03	1.30	–0.48	–0.35	3.29	1.32	–0.64	–0.18
Q26	3.54	1.17	–0.80	0.53	3.48	1.24	–0.81	0.39
Q27	3.43	1.25	–0.78	0.19	3.43	1.32	–0.88	0.27
Q28[a]	3.45	1.16	–0.88	0.65	NA	NA	NA	NA
Q29	3.37	1.17	–0.78	0.39	3.44	1.15	–0.65	0.14
Q30	3.41	1.30	–0.77	0.13	3.47	1.35	–0.80	–0.04
Q31	3.22	1.25	–0.66	0.05	3.28	1.24	–0.75	0.13
Q32	3.14	1.20	–0.53	–0.02	3.52	1.18	–0.76	0.40
Q33[a]	2.96	1.30	–0.42	–0.51	NA	NA	NA	NA
Q34	2.96	1.33	–0.32	–0.58	3.18	1.34	–0.63	–0.22
Q35	2.80	1.44	–0.19	–0.84	3.01	1.46	–0.41	–0.74
Q36	3.21	1.28	–0.60	–0.08	3.26	1.38	–0.72	–0.17
Q37	3.55	1.13	–0.92	0.91	3.49	1.17	–0.89	0.67

Note. Response options across all ACCReS items were presented on a 6-point Likert-type scale, and dummy coded for analysis (*strongly disagree* = 0, *disagree* = 1, *somewhat disagree* = 0, *somewhat agree* = 3, *agree* = 4, and *strongly agree* = 5). ACCReS = Assessment of Culturally and Contextually Relevant Supports; EFA = exploratory factor analysis; CFA = confirmatory factor analysis.

[a]Denotes items excluded due to high inter-item correlations ($r > .70$) across both EFA and CFA datasets.

indicated below), and (c) was interpretable and consistent with our conceptualization of culturally and contextually relevant supports (Tabachnick & Fidell, 2019).

As depicted in Table 2, the three-factor solution presented a distribution of items across themes representing teachers' (a) instructional style and behavior management practices (named Equitable Classroom Practices [ECP]), (b) data collection practices and access to PD (named Accessing Information and Support [AIS]), and (c) explicit consideration of student culture and the educational context (named Consideration of Culture and Context [CCC]). We found internal consistency to be acceptable for the AIS ($\omega_h = .87$), CCC ($\omega_h = .83$), and ECP ($\omega_h = .77$) factors.

Study 2

Study 2 contains an CFA to test the three-factor solution.

Method

Sample. In this sample, the 400 respondents were again predominately White (79.25%), female (71.21%), licensed or certified (88.41%), and worked in a public school (82.00%). The majority indicated that > 25% of their students were racially or ethnically minoritized youth (see Table 1).

Instrumentation. To conduct the CFA, participants completed the revised 37-item ACCReS.

Statistical analysis. For CFA procedures, we utilized maximum likelihood (ML) estimation with robust (i.e., Huber–White) standard errors to address potential issues relating to non-normality (Li, 2015). Prior to calculating model fit, we removed two items that were highly correlated (>.70) across datasets (see note in Table 2). This was to reduce redundancy and shorten the instrument (McCoach et al., 2013). To establish model fit, we calculated the Tucker–Lewis index (TLI), the comparative fit index (CFI), root mean squared error of approximation (RMSEA), standardized root mean squared residual (SRMR), chi-square, Akaike information criterion (AIC), and Bayesian information criterion (BIC). To evaluate fit indices, we used the following cutoffs: ≥.95 for TLI and CFI, ≤.06 for the RMSEA, and <.08 for the SRMR (Hu & Bentler, 1999; Sivo et al., 2006). For chi-square (χ^2), we determined if the ratio of χ^2 to degrees of freedom (*df*) was ≤ 3 and considered a lower value for AIC and BIC to indicate a better fit (Schreiber et al., 2006).

Results

Screening revealed that data violated multivariate normality. Although descriptive statistics indicated that participants provided the full range of response options, respondents again demonstrated a preference for *agree* and *strongly agree* (see Table 3 for means, standard deviations, skew, and kurtosis). The most popular response was *agree* (the median response category for 26 of the 35 items). Mean standard deviations across items were similar in both datasets (EFA = 1.09; CFA = 1.11). Two items were both highly correlated with other items and thus excluded from the final instrument (Items 28 and 33; see Table 2). These items were worded similarly to other items (Items 29 and 34), which were retained. Raw data were used for the CFA. The path diagram (see Figure S1 in online supplemental materials) shows all items and latent factors.

Model evaluation and internal consistency. The three-factor model demonstrated mixed results with regard to fit. Values for RMSEA (0.06, 90% confidence interval [CI] = [0.06, 0.07]), SRMR (0.07), and χ^2/df (2.50) were in the acceptable range, but TLI and CFI were < .95 (CFI = 0.88; TLI = 0.87). In addition, AIC and BIC were determined to be the lowest of comparison models (AIC = 34,830.72; BIC =35,122.09). All factor loadings were found to be statistically significant. As we noted AIS and CCC factors were correlated (r = .84), we examined a two-factor model for comparison. The AIS and CCC factors were collapsed into one factor, and the ECP domain stood alone. Results did not demonstrate a superior fit (e.g., higher AIC [35,177.32] and BIC [35,460.72]), and the two-factor model lacked theoretical justification (see Table S2 in online supplemental materials). Therefore, the three-factor model was retained. Estimates indicated acceptable internal consistency across all latent constructs in the final instrument: AIS (ω_h = .86), ECP (ω_h = .87), CCC (ω_h = .77) (see Table 2).

Study 3

Study 3 presents a preliminary convergent validity analysis. Convergent validity is fundamental to construct validity. Evidence of convergent validity supports a relationship between two measures of the same or similar construct and can be helpful when interpreting data produced by an instrument (e.g., ACCReS; American Educational Research Association et al., 2014).

Method

Participants. In this sample, 99 respondents were again predominately White (77.78%), female (83.84%), licensed or certified (95.96%), and taught in a public school (79.80%). Most indicated that >25% of their students were racially or ethnically minoritized youth (see Table 1).

Instrumentation

Assessment of Culturally and Contextually Relevant Supports. In this study, the 35-item ACCReS was administered.

Culturally Responsive Teaching Self-Efficacy Scale (CRTSES). Participants also completed the CRTSES (Siwatu, 2007), a 40-item unidimensional scale that evaluates teachers' perceived self-efficacy to engage in culturally responsive teaching practices in the classroom with strong internal consistency (α = .96; Siwatu, 2007). Teachers are instructed to rate the confidence with which they feel they can engage in items on a 0 to 100 scale, with zero indicating *no confidence at all* and 100 indicating *completely confident.* Sample items include,

> Rate how confident you are in your ability to engage in specific culturally responsive practices: (a) Adapt instruction to meet the needs of my students, (b) Teach students about their cultures' contributions to science, (c) Build a sense of trust in my students.

Culturally Responsive Classroom Management Self-Efficacy Scale (CRCMSES). Participants also completed the CRCMSES (Siwatu et al., 2017), a 35-item unidimensional scale that evaluates teachers' self-efficacy to implement culturally responsive behavior support strategies with strong internal consistency ($\alpha = .97$; Siwatu et al., 2017). The response format for the CRCMSES is similar to the CRTSES (i.e., 0–100; *no confidence at all* to *completely confident*). Sample items include,

> Rate how confident you are in your ability to successfully accomplish each of the tasks listed below: (a) Assess students' behaviors with the knowledge that acceptable school behaviors may not match those that are acceptable within a student's home culture, (b) Clearly communicate classroom policies, (c) Address inappropriate behavior without relying on traditional methods of discipline such as office referrals.

Analysis. To examine relationships between instrument scores, bivariate correlation analyses were conducted using Pearson's *r* (calculated using both subscale and overall raw scores). Correlational significance was established after application of the Holm–Bonferroni method to account for the effects of multiple comparisons (Holm, 1979). A sensitivity analysis ($\alpha = .05$, power = .80) indicated a sufficient sample for identification of a significant correlation coefficient.

Results

In comparison with the ACCReS, respondents engaged with a more limited range of response options within the 0 to 100 scale on the CRCMSES. Respondents neglected to interact with a full range of options across all CRCMSES items, and 13 of the 35 items had minimum response ratings of 20 or above (reflecting interaction limited to 80% or fewer of potential response options). The mean of minimum responses across all CRCMSES items was 72.11, and the mean of maximum responses was 90.06. A negative skew was notable. Results of respondent interactions with the CRTSES represent more variance in response selection than that observed in the CRCMSES. Respondents neglected to interact with a full range of options in only 12 of the 40 CRTSES items, and only seven of the total items had minimum response ratings of 20 or above.

As hypothesized, higher scores on the ACCReS subscale and total scale scores were significantly, positively correlated with total scores on the CRCMSES and CRTSES (see Table S3 in online supplemental materials). This provides preliminary evidence of convergent validity. Correlational analyses indicated a strong relationship between responses to both the CRCMSES and CRTSES measures ($r = .85$, $p < .001$). Correlations between the ACCReS and the CRCMSES and CRTSES were also positive and significant, but in the moderate range. This may be because the ACCReS was designed to align with MTSS, a framework which includes the consideration of not only teaching and classroom management practices but also the information and systems needed to support implementation (e.g., data, training, administrative support).

General Discussion

As the United States continues to become increasingly racially and ethnically diverse, school systems must be prepared to support all learners. This requires school staff members to be culturally responsive (Gay, 2018). When staff understand and value students' cultures, they are better able to design environments for students that are relevant and rigorous (Muñiz, 2019). These systems must include time and resources for educators to engage in self-reflection and high-quality in-service PD, both individually and collectively. The ACCReS was developed as a practical tool to assist educators in reflecting to improve their practice, and to provide assessment data to inform staff intervention needs. Results of this study produced a 35-item instrument measuring teachers' (a) use of ECP, (b) effort toward AIS, and (c) explicit CCC in the classroom.

The ACCReS items were developed based on results of a comprehensive literature review. Originally, items were hypothesized to align with a four-factor structure based on Vincent and colleagues' (2011) conceptualization of cultural responsiveness MTSS. We expected that each item would encourage teachers to consider students' culture in relation to the educational context. However, some items encouraged this consideration more explicitly (e.g., "I know how to provide culturally and contextually relevant instruction") than others (e.g., "I work to build a positive relationship with each student I teach"). Analyses indicated a three-factor configuration as the best model fit for the ACCReS, in which classroom instructional and behavior management practices were assessed within the same domain (ECP), PD and data were assessed on the second domain (AIS), and items encouraging explicit consideration of culture loaded onto a unique factor (CCC).

Upon testing the three-factor solution, findings from the CFA indicated mixed results with regard to model fit. Although some absolute fit indices indicating adequate fit (RMSEA, SRMR) and others fell below recommended cutoffs (TLI, CFI), it has been suggested that attention to SRMR and RMSEA may help retain the true model when discrepancies among indices are present (Sivo et al., 2006). Furthermore, Lai and Green (2016) caution against overinterpreting fit indices, indicating that there is still a need for an agreed upon standard for model fit interpretation, particularly when fit indices indicate mixed findings. In the future, researchers might target investigating the reason for mixed findings with regard to model fit. However, as the

ACCReS is meant to guide decisions about appropriate PD for educators (and not high-stakes clinical decisions, for instance), these findings present adequate evidence for the instrument's intended use.

In Study 3, we found significant correlations between total scores on the ACCReS and total scores on the CRCMSES and CRTSES. Conceptually, this positive and significant association stands to reason; Bandura's (1997) theory of self-efficacy supports the notion that teachers who perceive themselves as able to engage in culturally responsive practices (as evidenced based on responses to the CRCMSES and CRTSES) will also likely report their implementation of those practices on the ACCReS. Although relationships between scales were positive and significant, correlations were moderate, potentially indicating that whereas the CRCMSES and CRTSES scales target classroom management and teaching practices, respectively, the ACCReS items target behavioral supports, instructional practice, as well as access to data and systems of support. The CRCMSES, CRTSES, and ACCReS may function similarly, but not identically, and each may offer unique insights into teachers' perceptions and practice.

Limitations

Limitations should be considered when interpreting results. First, the majority of teachers indicated that at least one-quarter of their students were racially and ethnically minoritized youth, yet national student trends indicate 52% that racially and ethnically minoritized youth make up 52% of students nationwide. This may have affected how favorably teachers endorsed ACCReS items, and future studies might ensure these student trends are more represented in the participant sample. Also, although the teacher participants across the three studies were homogeneous, this is indicative of teacher demographics in the United States (i.e., White, female). Furthermore, social desirability bias is always a limitation when using self-report measures. Yet, recruitment occurred via a paneling service. Although the use of a paneling service limits the opportunity to determine a response rate and could introduce sampling bias (as certain teachers may choose to opt-in to serve as panelists), participants were aware that their responses were completely anonymous. Therefore, it is unlikely that participants felt it necessary to misrepresent themselves as researchers did not know their identity. In the future, researchers might also administer a brief social desirability scale with the ACCReS. Relatedly, as described in Debnam and colleagues (2015), teachers tended to provide high ratings related to their cultural responsiveness, seen in this study on items within the ECP subscale. Although teachers may produce data that bias more favorable responses, relative intraindividual weakness in any area may provide topic areas for which PD is useful.

A high number of variables per factor may have both misleadingly improved model fit and compromised stability (Hogarty et al., 2005). However, overdetermination can be a strength to a degree as five or more items per factor is recommended (Comrey & Lee, 1992). Also, although some researchers indicate there are limitations to the use of Pearson's product–moment coefficients (Holgado-Tello et al., 2008), others contend it is acceptable to use in factor analysis (Murray, 2013). Finally, in Study 3, the sample was deemed adequate and representative, yet the relatively small number of participants may limit the extent to which these findings are generalizable. However, results provide a necessary piece of the larger puzzle of validation procedures conducted to examine the psychometric properties of scores derived from the ACCReS.

Implications

Additional research is needed to understand the reason for model fit findings (Lai & Green, 2016). It is possible that the factor structure might be improved by reducing or adding items, or altering the content of current items and repeating analyses. However, this instrument was created for teacher reflection and to inform selection of PD topics. As such, the current version is suitable for this applied purpose. Future research might also target concurrent and predictive validity, and differential item functioning according to teacher characteristics. Specifically, tests of invariance by teacher race/ethnicity may provide valuable insight. It is also important to determine if there is evidence of generalizability of scores over time, across individuals in various contexts, and between ACCReS and other data sources (e.g., observation). Future research might include student outcome data as well as both observer and teacher self-report data to run comprehensive and comparative analyses. Research might also target if completing the ACCReS changes teachers' practice, and measure more distal outcomes (e.g., improved student achievement) over time.

Conclusion

Results of the current study indicate preliminary reliability and validity evidence for the 35-item, three-factor ACCReS, but additional validation endeavors are needed. In practice, the ACCReS may prove to be a valuable tool to assess teachers' perceptions and actions related to cultural responsiveness, particularly within an MTSS context. Data from the ACCReS could guide decisions regarding educator intervention (e.g., PD), promote change in classroom practice, and ultimately benefit racially and ethnically minoritized youth who have historically been disadvantaged in the U.S. education system. As teachers often enter the field with a lack of understanding of their own biases and the impact of students' culture

on learning, efforts toward assessing teachers' perceptions and practices may be a critical first step in designing effective PD that will dismantle systemic barriers to equitable learning environments.

Declaration of Conflicting Interests

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

Funding

The U.S. Department of Education's Institute of Education Sciences supported this research through Grant R324B170010 to the University of Massachusetts Boston. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Supplementary Material

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

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Henry Holt & Co.

Biggs, B. K., Vernberg, E. M., Twemlow, S. W., Fonagy, P., & Dill, E. J. (2008). Teacher adherence and its relation to teacher attitudes and student outcomes in an elementary school-based violence prevention program. *School Psychology Review*, *37*(4), 533–549. https://doi.org/10.1080/02796015.2008.12087866

Bohrnstedt, G., Kitmitto, S., Ogut, B., Sherman, D., & Chan, D. (2015). *School composition and the black–white achievement gap* (NCES 2015-018). U.S. Department of Education, National Center for Education Statistics.

Cammarota, J., & Romero, A. (2009). The Social Justice Education Project: A critically compassionate intellectualism for Chicana/o students. In W. Ayers, T. Quinn, & D. Stovall (Eds.), *Handbook for social justice education* (pp. 465–476). Lawrence Erlbaum.

Cartledge, G., & Kleefeld, J. (2010). *Working together: Building children's social skills through folktales, Grades 3–6* (2nd ed.). Research Press.

Castro-Olivo, S. M. (2014). Promoting social-emotional learning in adolescent Latino ELLs: A study of the culturally adapted Strong Teens Program. *School Psychology Quarterly*, *29*(4), 567–577. https://doi.org/10.1037/spq0000055

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Comrey, A. L., & Lee, H. B. (1992). Interpretation and application of factor analytic results. In A. L. Comrey & H. B. Lee (Eds.), *A first course in factor analysis* (2nd ed., pp. 240–262). Psychology Press.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Fourrecommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 10(7), 1–9. https://doi.org/10.7275/jyj1-4868

Debnam, K. J., Pas, E. T., Bottiani, J., Cash, A. H., & Bradshaw, C. P. (2015). An examination of the association between observed and self-reported culturally proficient teaching practices. *Psychology in the Schools*, *52*(6), 533–548. https://doi.org/10.1002/pits.21845

Ellerbrock, C. R., Cruz, B. C., Vásquez, A., & Howes, E. V. (2016). Preparing culturally responsive teachers: Effective practices in teacher education. *Action in Teacher Education*, *38*(3), 226–239. https://doi.org/10.1080/01626620.2016.1194780

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015–2016).

Fallon, L. M., Cathcart, S. C., DeFouw, E. R., O'Keeffe, B. V., & Sugai, G. (2018). Promoting teachers' implementation of culturally and contextually relevant classwide behavior plans. *Psychology in the Schools*, *55*, 278–294. http://doi.org/10.1002/pits.22107

Fallon, L. M., Cathcart, S. C., & Johnson, A. H. (2021b). Assessing differential item functioning in a teacher self-assessment of cultural responsiveness. *Journal of Psychoeducational Assessment*, *39*(7), 816–831. https://doi.org/10.1177/07342829211026464

Fallon, L. M., DeFouw, E. R., Cathcart, S. C., Berkman, T. S., O'Keeffe, B. V., & Sugai, G. (2021a). Supports to improve academic outcomes with racially and ethnically minoritized youth: A review of research. *Remedial and Special Education*. Advance online publication. https://doi.org/10.1177/07419325211046760

Fallon, L. M., DeFouw, E. R., Cathcart, S. C., Berkman, T. S., Robinson-Link, P., O'Keeffe, B. V., & Sugai, G. (2022). School-based supports and interventions to improve social and behavioral outcomes with racially and ethnically minoritized youth: A review of recent quantitative research. *Journal of Behavioral Education*, *31*, 123–156. https://doi.org/10.1007/s10864-021-09436-3

Fallon, L. M., O'Keeffe, B. V., & Sugai, G. (2012). Consideration of culture and context in school-wide positive behavior support: A review of current literature. *Journal of Positive Behavior Interventions*, *14*(4), 209–219. https://doi.org/10.1177/1098300712442242

Fallon, L. M., Sanetti, L. M. H., Chafouleas, S. M., Faggella-Luby, M. N., & Briesch, A. M. (2018). Direct training to increase agreement between teachers' and observers' treatment integrity ratings. *Assessment for Effective Intervention*, *43*, 196–211. https://doi.org/10.1177/15345084177387

Finch, M. E. (2012). Special considerations with response to intervention and instruction for students with diverse backgrounds. *Psychology in the Schools*, *49*(3), 285–296. https://doi.org/10.1002/pits.21597

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315. https://doi.org/10.1086/209351

Gage, N. A., Whitford, D. K., Katsiyannis, A., Adams, S., & Jasper, A. (2019). National analysis of the disciplinary

exclusion of Black students with and without disabilities. *Journal of Child and Family Studies*, *28*(7), 1754–1764. https://doi.org/10.1007/s10826-019-01407-7

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.

Gopalan, M., & Nelson, A. A. (2019). Understanding the racial discipline gap in schools. *AERA Open*, *5*(2), 1–26. https://doi.org/10.1177/2332858419844613

Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the racial discipline gap in classrooms by changing teacher practice. *School Psychology Review*, *45*(2), 171–191. https://doi.org/10.17105/spr45-2.171-191

Guyton, E. M., & Wesche, M. V. (2005). The multicultural efficacy scale: Development, item selection, and reliability. *Multicultural Perspectives*, *7*(4), 21–29. https://doi.org/10.1207/s15327892mcp0704_4

Hershfeldt, P. A., Sechrest, R., Pell, K. L., Rosenberg, M. S., Bradshaw, C. P., & Leaf, P. J. (2009). Double-check: A framework of cultural responsiveness applied to classroom behavior. *TEACHING Exceptional Children Plus*, *6*(2), 2–18.

Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, *65*(2), 202–226. https://doi.org/10.1177/0013164404267287

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70. https://www.jstor.org/stable/4615733

Howard, T. C., & Navarro, O. (2016). Critical race theory 20 years later: Where do we go from here? *Urban Education*, *51*(3), 253–273. https://doi.org/10.1177/0042085915622541

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., & Cui Dilig, R. (2020). *The condition of education 2020* (NCES 2020-144). National Center for Education Statistics.

Individuals With Disabilities Education Improvement Act (IDEA), 20 U.S.C. § 1400 *et seq.* (2004).

Kassambara, A. (2020). Rstatix: Pipe-friendly framework for basic statistical tests. *R package version 0.6. 0*.

Korkmaz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*(2), 151–162. https://doi.org/10.32614/RJ-2014-031

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, *32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, *51*(2–3), 220–239. https://doi.org/10.1080/00273171.2015.1134306

La Salle, T. P., Wang, C., Wu, C., & Rocha Neves, J. (2020). Racial mismatch among minoritized students and white teachers: Implications and recommendations for moving forward. *Journal of Educational and Psychological Consultation*, *30*(3), 314–343. https://doi.org/10.1080/10474412.2019.1673759

Lawrence, M. (2016). *ez: Easy analysis and visualization of factorial experiments* (R Package Version 4.4-0). https://CRAN.R-project.org/package=ez

Li, C. (2015). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. Springer.

Milner, H. R. (2017). Race, talk, opportunity gaps, and curriculum shifts in (teacher) education. *Literacy Research: Theory, Method, and Practice*, *66*(1), 73–94. https://doi.org/10.1177/2381336917718804

Muñiz, J. (2019). *Culturally responsive teaching: A 50-state survey of teaching standards*. https://files.eric.ed.gov/fulltext/ED594599.pdf

Murray, J. (2013). Likert data: What to use, parametric or non-parametric? *International Journal of Business and Social Science*, *21*(11), 258–264.

National Center for Education Statistics. (2019a). *Common core of data (CCD), "State nonfiscal survey of public elementary and secondary education," 2000–01 and 2017–18; and National elementary and secondary enrollment by race/ethnicity projection model*.

National Center for Education Statistics. (2019b). National Teacher and Principal Survey (NTPS), "Public School Teacher Data File," 2017–18. National Teacher and Principal Survey (NTPS).

Peters, T., Margolin, M., Fragnoli, K., & Bloom, D. (2016). What's race got to do with it? Preservice teachers and White racial identity. *Current Issues in Education*, *19*(1), 1–22. https://cie.asu.edu/ojs/index.php/cieatasu/article/view/1661

Pinkelman, S. E., McIntosh, K., Rasplica, C. K., Berg, T., & Strickland-Cohen, M. K. (2015). Perceived enablers and barriers related to sustainability of school-wide positive behavioral interventions and supports. *Behavioral Disorders*, *40*(3), 171–183. https://doi.org/10.17988/0198-7429-40.3.171

Powell, R., Cantrell, S. C., Malo-Juvera, V., & Correll, P. (2016). Operationalizing culturally responsive instruction: Preliminary findings of CRIOP research. *Teacher's College Record*, *118*(1), 1–46. https://doi.org/10.1177/016146811611800107

Proctor, S. L., & Owens, C. (2019). School psychology graduate education retention research characteristics: Implications for diversity initiatives in the profession. *Psychology in the Schools*, *56*(6), 1037–1052. https://doi.org/10.1002/pits.22228

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/joer.99.6.323-338

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *The Journal of Experimental Education*, *74*, 267–288. https://doi.org/10.3200/JEXE.74.3.267-288

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education*, *23*(7), 1086–1101. https://doi.org/10.1016/j.tate.2006.07.011

Siwatu, K. O., Putman, S. M., Starker-Glass, T. V., & Lewis, C. W. (2017). The culturally responsive classroom management self-efficacy scale: Development and initial validation. *Urban Education*, *52*(7), 862–888. https://doi.org/10.1177/0042085915602534

Skiba, R. J., Arredondo, M. I., Gray, C., & Rausch, M. K. (2016). What do we know about discipline disparities? New and emerging research. In R. J. Skiba, K. Mediratta, & M. Karega Rausch (Eds.) *Inequality in school discipline* (pp. 21–38). Palgrave Macmillan.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, *51*(4), 640–670. https://doi.org/10.3102/0002831214541670

Spanierman, L. B., Oh, E., Heppner, P. P., Neville, H. A., Mobley, M., Wright, C. V., Dillon, F. R., & Navarro, R. (2011). The multicultural teaching competency scale: Development and initial validation. *Urban Education*, *46*(3), 440–464. https://doi.org/10.1177/0042085910377442

Sugai, G., & Horner, R. H. (2020). Sustaining and scaling positive behavioral interventions and supports: Implementation drivers, outcomes, and considerations. *Exceptional Children*, *86*(2), 120–136. https://doi.org/10.1177/0014402919855331

Sugai, G., O'Keeffe, B. V., & Fallon, L. M. (2012). A contextual consideration of culture and school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *14*(4), 197–208. https://doi.org/10.1177/1098300711426334

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics*. Pearson.

Tanguay, C. L., Bhatnagar, R., Barker, K. S., & Many, J. E. (2018). AAA+ professional development for teacher educators who prepare culturally and linguistically responsive teachers. *Curriculum and Teaching Dialogue*, *20*(1/2), 87–181.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions. *Frontiers in Psychology*, *7*, 769–778. https://doi.org/10.3389/fpsyg.2016.00769

Vincent, C. G., Randall, C., Cartledge, G., Tobin, T. J., & Swain-Bradway, J. (2011). Toward a conceptual integration of cultural responsiveness and schoolwide positive behavior support. *Journal of Positive Behavior Interventions*, *13*(4), 219–229. https://doi.org/10.1177/1098300711399765

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. https://doi.org/10.1177/0011000006288127