Promoting Educators' Use of Culturally Responsive Practices:

A Systematic Review of In-Service Interventions

Jessika H. Bottiani - University of Virginia Curry School of Education
Kristine E. Larson - Johns Hopkins Bloomberg School of Public Health
Katrina J. Debnam - Johns Hopkins Bloomberg School of Public Health
Christina M. Bischoff - Johns Hopkins Bloomberg School of Public Health
Catherine P. Bradshaw - University of Virginia Curry School of Education

This is a post-peer-review, pre-copyedit version of an article published in Journal of Teacher Education. The final authenticated version is available online at: https://doi.org/10.1177/0022487117722553

Bottiani, J.H., Larson, K. Debnam, K., Bischoff, C., & Bradshaw, C.P. (2017). Promoting educators' use of culturally responsive practices: A systematic review of in-service interventions. Journal of Teacher Education, 1–19. doi:10.1177/0022487117722553.

Agency: Institute of Education Sciences Grant Number: R305A150221

Abstract

Few educators are equipped to bridge cultural differences to ensure that all students have opportunities to learn and succeed. Existing frameworks for culturally responsive practices (CRP) suggest its potential for promoting equitable learning environments, yet the state of the science has not been assessed. This systematic review aimed to 1) describe the features of empirically examined in-service CRP interventions; 2) analyze the quality of the empirical studies; and 3) characterize study measures, outcomes, and conclusions regarding intervention impact. We found a total of just 10 empirical studies of the impact of CRP in-service training models (2 quantitative and 8 qualitative). Study methods universally failed to meet standards of evidence for efficacy, effectiveness, and dissemination; none employed rigorous design features to allow causal inference. Findings suggest that the research base is inadequate to draw conclusions regarding effectiveness and that more rigorous CRP in-service intervention research is needed.

Keywords: culturally responsive, professional development, school equality, discipline gap, race, school/teacher effectiveness

Promoting Educators' Use of Culturally Responsive Practices: A Systematic Review of In-Service Interventions

To effectively and equitably support student learning, educators must learn to leverage culture and cross-cultural differences in the classroom (Delpit, 2006; Gay, 1995, 2010; Moll, Amanti, Neff, & Gonzalez, 2005; Paris, 2012). Shifts in the racial, ethnic, and cultural diversity of students in U.S. schools are calling teachers to this increasingly complex task. By 2024, students of color (including Black, Latino, Asian, Native Hawaiian, Pacific Islander, American Indian, and multi-racial youth) are projected to represent 56% of the student enrollment in U.S. public schools (Kena et al., 2015). Despite changes in student diversity, the racial and ethnic composition of our teaching workforce has remained monolithic. Nearly 82% of U.S. public school teachers are non-Hispanic White, and the proportion of teachers entering the workforce who are of color (21%) does not suggest significant change on the horizon (NCES, 2013). Even if there were better representation of people of color in the teaching workforce, a student-teacher racial or ethnic match does not necessarily translate to effective, culturally responsive, and unbiased classroom practices (Bradshaw, Mitchell, & Leaf, 2010; Ehrenberg, Goldhaber, & Brewer, 1995), nor does it guarantee that teachers have skills needed to support positive peer cross-cultural interactions. Given accumulating and persistent evidence of racial and ethnic disparities in school disciplinary and special education practices (Artiles, Kozleski, Trent, Osher, & Ortiz, 2010; Losen & Orfield, 2002; Porowski, O'Conner, & Passa, 2014; Sullivan, 2011; Skiba et al., 2011; cf. Morgan et al., 2015), interventions to promote educator competencies and school-wide procedures to bridge cultural divides in the classroom have become increasingly critical (Delpit, 2006; Gay, 1995, 2010).

Several researchers have proposed frameworks to help teachers develop skills to promote inclusive, respectful, and equitable classroom environments through the use of culturally responsive and sustaining practices. These frameworks have largely focused on enhancing the fit between students' home and school cultural ways of knowing and learning (Au & Jordan, 1981; Cazden & Legett, 1981; Erickson & Mohatt, 1982); addressing structural inequities which create a mismatch between students and schools (Ogbu, 1994; Villegas, 1988); exploring learning as cultural practice (Gutierrez & Rogoff, 2003); increasing teachers' understanding of students' cultures as 'funds of knowledge' (Moll et al., 2005); and the use of other culturally responsive practices (CRP) in diverse, urban classrooms (Au, 2009; Auerbach, 2009; Bonner, 2012; Epstein, Mayorga, & Nelson, 2011; Gay, 2000; Garza, 2009; Gregory & Ripski, 2008; Lalas, 2007; Gloria Ladson-Billings, 1994; Gloria Ladson-Billings 1995; Nieto, 2013; Phillipo, 2012; Sampson & Garrison-Wade, 2011; Risko & Walker-Dalhouse, 2008; Toldson & Lemmons, 2013; Ware, 2006). There has also been recent interest in school-wide approaches to improving schools' cultural responsiveness through cultural adaptation of school-wide positive behavior support (SWPBS; Sugai & Horner, 2002; Vincent et al. 2011). Together, the extant work on CRP highlights the importance of tapping students' cultural backgrounds and differences as a resource to promote engagement and learning, and as a resource that should be sustained by schools, rather than as a problem to address (Paris, 2012). This study adopts this view and employs a definition of CRP developed by Geneva Gay (2000): "It teaches to and through [students'] strengths. . . it builds bridges of meaningfulness between home and school experiences as well as between academic abstractions and lived sociocultural realities; it uses a wide variety of instructional strategies that are connected to different learning styles; it incorporates

multicultural information, resources, and materials in all the subjects and skills routinely taught in schools" (p. 29).

Despite extensive literature on CRP and movement in the field to increase the use of such approaches in the classroom, progress towards establishing an evidence base on effective strategies to promote educators' use of CRP has been slow (Bottiani, Bradshaw, Rosenberg, Hershfeldt, Pell, & Debnam, 2012; Fiedler et al., 2008; Griner & Stewart, 2013). There are several inter-related issues with the CRP literature hindering progress in this area, including questions regarding which in-service interventions to promote use of CRP are effective with educators, what study designs researchers should use to assess effectiveness, and which outcome measures are reliable and sensitive to change. A review of the literature to examine these issues is necessary to advance the research base for strategies to promote CRP, and, in turn, to support investments in effective models for scale-up and dissemination. Systematic reviews of the evidence supporting intervention can provide a bridge to help efforts to close the gap between research and practice (Bero, Grilli, Grimshaw, Harvey, Oxman, & Thomson, 1998).

The Current Study

To strengthen our understanding of the state of the science of strategies to promote CRP, we systematically reviewed the literature with a focus on interventions targeted to in-service administrators and teachers, in the context of general education settings. In this review, our purpose was to generate knowledge to inform ongoing, real world school-based interventions to promote CRP. Therefore, we did not focus our review on pre-service (i.e., university-level teacher education coursework on culturally responsive practices). We employ the term 'interventions' in this paper to refer to a broad range of professional development and school-wide reforms (e.g., training, coaching and consultation, policy and systems changes, and peer-to-peer learning

models) affecting the implementation of CRP within the classroom and school-wide. Our review centered on three specific aims: 1) to describe the features of in-service CRP interventions that have been empirically examined; 2) to analyze the quality of empirical studies of in-service interventions to promote CRP; and 3) to characterize study measures, outcomes, and conclusions regarding intervention impact on schools' and teachers' use of CRP and on disparities in student outcomes. Our overarching goal was to inform educational researchers and practitioners regarding in-service interventions that have demonstrated improvements in CRP or reductions in disparate student outcomes. We summarized findings from the systematic review and discussed implications for future empirical research to determine effective in-service CRP interventions.

Method

This study employed a systematic review methodology based on the Campbell Collaboration guidelines for conducting systematic reviews (Hammerstrøm et al., 2010). The Campbell Collaboration highlights the importance of formulating clearly defined research questions, using sensitive search methods, and methodically screening studies (rigorous application of explicit inclusion and exclusion criteria). Applying these methods, we followed a six-step review process, described in detail below.

Step 1: Formulating the Review Aims

The purpose of this study was to determine the state of the evidence in support of in-service interventions to promote teacher and school administrator CRP in public K-12 schools in the U.S. Specifically, our first aim (Features of In-Service Interventions) was to characterize the interventions that had been studied in terms of the *dosage* of the in-service intervention (periodicity, intensity, duration), the *target for change* (knowledge, attitudes, procedures, skills); and, the *mode of delivery* used for the in-service intervention (e.g., individualized coaching, all-

staff training). We also documented who was responsible for *implementation* of the in-service intervention; whether staff participation was *voluntary* or mandated; and whether in-service intervention was designed to be *specialized* to a particular student population or to be generalizable across diverse student groups. These features of dosage, target for change, mode of delivery, implementer, voluntary participation, and specialized population were selected as key intervention characteristics that influence intervention exposure and uptake based on implementation fidelity research in the field of education (Bickman et al., 2009). For example, if a researcher comes into the school to implement the intervention, teachers may value their expertise enough to participate, but may have less reinforcement to implement relative to an intervention delivered by a peer or their administrator. Our second aim (Study Quality) was to characterize the empirical studies that have been employed to examine CRP in-service intervention impact. Specifically, we sought to analyze the *quality* of the empirical studies identified using relevant standards of evidence for quantitative and qualitative research (Creswell & Miller, 2000; Gottfredson et al., 2015). This assessment of quality applied standards of evidence to each study to determine how confident we can be that the intervention caused the outcomes reported. The standards of evidence we utilized are specified in Step 6 (Analyze Study Quality). For our third aim (Impact of In-Service Interventions), we identified promising or evidence-based practices in light of study quality, and document the *definitions*, *measures*, and other criteria that were used to assess CRP and student outcomes. This aim is the most actionable component of the study and was intended to provide a guide for change based on the findings of the review.

Step 2: Defining Inclusion and Exclusion Criteria

To optimize the specificity of our search, inclusion and exclusion criteria were set applying the PICO framework (Population, Interventions, Comparators, and Outcomes; Schardt, Adams, Owens, Keitz, & Fontelo, 2007). We defined our *Population* of interest as in-service teachers and school administrators due to our intervention implementation focus in classrooms and school-wide; support or clinical school staff were excluded as their responsibilities were relevant to contexts outside the scope of the study (i.e., clinical therapeutic contexts). We included U.S. schools at any grade level (K-12); thus, we excluded studies that examined CRP in other settings (e.g., juvenile justice, child welfare). In addition, we excluded any studies of interventions that were exclusively focused on impacting practice within self-contained special education or English as a Second Language (ESL) classrooms, as they were not relevant to general education practice. We did not otherwise limit or exclude studies of interventions focused on impacting teacher or administrators' practices with specific student groups (i.e., by race, ethnicity, or gender). We limited our *Intervention* focus to interventions delivered to in-service school administrators and teachers to promote CRP and related student outcomes, including but not limited to training, coaching and consultation, policy and systems changes, peer-to-peer learning models, and teacher induction. Thus, we excluded interventions providing pre-service training and education; a review of research on pre-service training and education to promote CRP is also a gap in the literature and merits a full, separate review. Note that we did not exclude studies based on their CRP definition for their intervention. The *Comparators* component of the PICO framework refers to whether the study was designed to support causal inference regarding the impact of the intervention (i.e., allowing conclusions that the intervention, and not some other factor, caused the measured changes); typically, these study designs require a comparison group. In applying this component of the PICO model, we

recognized that some flexibility would be necessary to accommodate differences in methodological approach and in the context in which the interventions or programs were delivered (Pettigrew & Roberts, 2006). Therefore, we did not require that a comparison group be part of the study design for inclusion in our review; however, we did exclude articles that did not empirically report, either qualitatively or quantitatively, on impacts associated with the intervention. Although we kept the target of the intervention focus on school staff (i.e., teachers and administrators), *Outcomes* of interest included both CRP and disparities in outcomes between student groups. See Table 1 for a summary of the inclusion and exclusion criteria applied to the present review.

Step 3: Developing and Documenting the Search Strategy

We developed a written protocol to ensure a comprehensive and replicable search strategy, adhering to detailed information retrieval guidelines given by the Campbell Collaboration (Hammerstrøm et al., 2010). Search limiters were 1) Peer-reviewed journals only and 2) publication dates between January 1, 1998 and December 31, 2014. The Boolean search phrase was as follows: “school” or “teacher” or “classroom” AND “culturally responsive” or “disproportionality” or “discipline gap” or “cultural proficiency” AND “study” or “empirical” or “research” NOT “special education”. We a used centralized search function through EBSCOHost to conduct the searches in *Academic Search Complete*, *ERIC*, *PsycINFO*, and *Social Sciences Full Text* simultaneously order to minimize duplication of results. We avoided controlled vocabulary, filters, delimiters, and/or approaches for handling truncation and related terms unique to any specific database.

Steps 4-5: Select Studies and Extract Data

Studies were systematically selected through the application of inclusion and exclusion criteria defined above through a phased review of abstracts, including initial identification, screening, determination of eligibility, and final inclusion. We then characterized each study by the study design (e.g., randomized controlled trial, quasi-experimental study, case study with quantitative outcomes, qualitative case study); sample and setting; definition of CRP; intervention characteristics (including dosage, target for change, implementer, specific or generalized student focus, and voluntary or mandated participation); staff and student outcome measures; and, findings.

Step 6: Analyze Study Quality

We then followed the Department of Education's What Works Clearinghouse (WWC) framework for categorizing studies by degrees of evidence (i.e., ineligible for review, does not meet standards of evidence, meets standards with reservations, meets without reservations; U.S. Department of Education, 2015). The WWC (https://ies.ed.gov/ncee/WWC/) reviews research on different education programs and policies with the goal of providing educators information needed to make evidence-based decisions. For quantitative studies, we applied rigorous standards for characterizing an intervention as "evidence-based" (Flay et al., 2005; Gottfredson et al., 2015). These standards of evidence are organized around the types of validity critical for intervention research, as described in Cook and Campbell (1979) and Shadish et al. (2002), which include statistical conclusion validity (i.e., appropriate use of statistical methods, including adjusting for baseline measures in the analysis), internal validity (i.e., study designs include equivalent comparison group and random assignment, or quasi-experimental designs that allow causal inference), construct validity (i.e., psychometric properties of outcome measures are sound), and external validity (i.e., the target population, setting, intervention, and method of

sampling are described well enough to evaluate the generalizability of the results). To be eligible for the quantitative study quality analysis, the study minimally had to have a group design that would allow for a quantitative, comparative analysis between an intervention group and an equivalent control group. To analyze the quality of qualitative or mixed methods studies, we used quality standards for qualitative research outlined by Brantlinger, Jimenez, Klingner, Pugach, and Richardson (2005) and Creswell and Miller (2000). These indicators included whether the reporting demonstrated researchers' use of triangulation, disconfirming evidence, researcher reflexivity, member checks, external auditors, peer debriefing, audit trail, prolonged field engagements, thick detailed description, and/or particularizability. After individually coding each article on these quality indicators (present or not), we randomly selected half (four of eight) to subject to a secondary review to ensure consistency between coders.

Results

Our initial search yielded a total of 622 articles. Of these, 179 unduplicated articles met basic inclusion criteria (i.e., U.S., educator sector, general education, K-12 settings). We then excluded 117 articles because they did not feature an intervention to improve CRP (i.e., we excluded papers that included only descriptions of culturally responsive practices or descriptions of teacher characteristics associated with cultural competency). Of the remaining 62 articles, we excluded another 34 because the studies' interventions did not directly intervene with teachers or administrators (e.g., the intervention was for students only). Thereafter, we examined the full articles reporting on the remaining 28 studies to ensure that each paper reported on an empirical examination of the impact of an in-service intervention to promote CRP. This process led to the exclusion of another 18 articles, leaving only 10 peer-reviewed articles eligible for inclusion from the 16-year period of 1998 – 2014 (see decision tree in Figure 1). To address the three aims

of this study, we describe key findings of our analysis of these ten articles below, which are summarized in Table 2.

Aim 1: Features of In-Service Interventions

Target for change. Each study in our review covered a different in-service intervention, with the exception of the studies by Vincent, Swain-Bradway, Tobin, and May (2011) and Jones Caravaca, Cizek, Horner, and Vincent (2006). Both of these studies focused on Positive Behavioral Interventions and Supports (PBIS; Sugai & Horner, 2002), with attention to its utility as a framework for enhancing CRP or reducing discipline disproportionality. However, in Jones et al. (2006), PBIS was culturally adapted to promote the schools' responsiveness to a specific Dine (Navajo) student population in New Mexico; this culturally-adapted version of PBIS was the only intervention in our review that targeted *school-wide procedures*, *skills*, and *knowledge/beliefs* for change (all three in combination).

Five studies had a singular target for change. Two of these focused on *procedural* changes only. In Vincent, Swain-Bradway, et al. (2011), a secondary analysis of national data from PBIS schools, there were no specific cultural adaptations and the intervention targeted changes in school-wide procedures only. Similarly, the intervention in the study by Shriberg et al. (2012) focused on a culturally responsive, participant-driven data collection and decision-making approach utilizing a Participatory Action Research (PAR) framework, and thus targeted school-wide procedural changes only. Two study interventions focused on improving *knowledge/beliefs* only (Fickel, 2005; Hammerness & Matsko, 2013). The intervention examined by Fickel (2005) was a professional development immersion model (building on a PDS initiative) which took place in a rural Alaska Native village and in which participations engaged with Native Elders to gain insights into Native ethnotheories of learning. The

intervention examined in Hammerness and Matsko (2013) employed a coaching intervention during teacher induction to shift teacher beliefs and attitudes (e.g., local socio-cultural context-specific supports to understand and maintain commitment to urban teaching). The study by Ryan, Chandler, and Samuels (2007) explored an intervention to promote culturally responsive school-based evaluation (called the Culture in Evaluation Project, a pseudonym to protect confidentiality), which was *skills*-focused only and designed to equip teachers and principals with skills needed to plan and conduct a culturally-sensitive school-based evaluation (aimed at promoting school improvement, external accountability, and equity).

Otherwise, the remaining four studies' interventions were designed to enhance both teacher *knowledge/beliefs* (e.g., CRP theory, cultural self-awareness, other cultural awareness) and *skills* (e.g., instructional strategies, self-evaluation, behavior management) in combination (Eberly et al., 2010; McAllister & Irvine, 2002; McCormick et al., 2013; Thompson & Byrnes, 2011). The case study by Eberly and colleagues described an interactive, three-session cultural responsiveness professional development series which focused on informing teachers about how culture and context can shape parent involvement, child-rearing practices, attitudes towards education, and ways teachers can tap students' cultural backgrounds when planning lessons through interactive simulation, role playing, and self-reflection exercises. The study by McAllister and Irvine (2002) described a similar intervention approach, called the CULTURES professional development seminar, which supported teacher learning through interactive, role-playing (e.g., Bafa Bafa cross-cultural simulation game), and reflective thinking exercises. The intervention studied by McCormick and colleagues (2013) describes a professional development school (PDS) initiative and partnership with a university that implemented a yearlong intervention called Project REACH, which employed book study groups and discussion of

awareness and practices related to CRP. The study by Thompson and Byrnes (2011) examined outcomes of students whose teachers participated in the district's REACH (Respecting Ethnic and Cultural Heritage) training, which included content related to culturally responsive teaching, theory, instructional and assessment strategies for English Language Learner (ELL) students, family engagement and empowerment, and community involvement.

Mode of delivery. Professional development training in a group setting was most frequently the modality of choice, with four of 10 interventions employing this approach to improve CRP (Eberly et al., 2010; McAllister & Irvine, 2002; Ryan et al., 2007; Thompson & Byrnes, 2011). Often, the professional development approaches featured interactive sessions that employed role playing, simulations, group and individual analysis of case scenarios, rather than taking a didactic approach. Other modes included immersion experience (Fickel, 2005), coaching for new teacher induction (Hammerness & Matsko, 2013), and an interactive in-person and online book study group (McCormick et al., 2013). For those interventions that were procedural in nature, the mode of delivery was less clearly defined, though some included needs assessment approaches (e.g., Jones et al., 2006; Shriberg et al., 2012).

Dosage. Details regarding the dosage of intervention, including frequency (e.g., number of sessions per month), intensity (e.g., how long each session met), and duration (e.g., over two years) was inconsistently provided. Sufficient studies identified their duration to merit some categorization. Specifically, one intervention was implemented over 10 days (Fickel, 2005), as an immersion experience, this information documents the frequency, intensity, and duration of the intervention. In addition, two interventions were implemented over 3 months (Eberly et al., 2010; McCormick et al., 2013), two interventions were implemented over one to two years (Ryan et al., 2007; Shriberg et al., 2012), and one intervention was implemented over a five-year

span (Hammerness & Matsko, 2013). The remaining four studies' intervention durations were not specified (Jones et al., 2006; McAllister & Irvine, 2002; Thompson & Byrnes, 2011; Vincent, Swain-Bradway, et al., 2011); however, two of these studies reported on intervention intensity, which was 40 hours (McAllister & Irvine, 2002) and 18 university credits (Thompson & Byrnes, 2011). In addition, Eberly et al. (2010) reported intensity of 2.5 hours for each of its three sessions over 3 months. Overall, the 8 studies that reported information on intervention dosage suggest wide variability on this factor.

Implementation. The entity responsible for implementing the intervention varied across the 10 studies, if specified at all. In three studies, outside researchers implemented the intervention (Eberly et al., 2010; Fickel, 2005; Hammerness & Matsko, 2013); in one study, the district implemented the intervention (Thompson & Byrnes, 2011); in another, the school staff as a whole implemented the intervention (a needs assessment; Shriberg et al., 2012); in another, the intervention (a book study) was implemented by grade-level chairs (McCormick et al., 2013).

Voluntary or mandated participation. Four of 10 studies reported that participation in the intervention was voluntary, either by staff participants or at the school-level, by the school administration (Fickel, 2005; Hammerness & Matsko, 2013; McAllister & Irvine, 2002; McCormick et al., 2013). In only one intervention were school staff mandated to participate (Shriberg et al., 2012). The remaining four studies did not specify whether participation in the intervention was voluntary or mandated (Eberly et al., 2010; Jones et al., 2006; Thompson & Byrnes, 2011; Vincent, Swain-Bradway, et al., 2011).

Specific to a cultural group, or generalizable across diverse groups. Eight of 10 studies featured interventions that were designed to improve teacher or administrator CRP in a way that was generalizable across diverse cultural groups. Only two studies featured

interventions that were tailored to a specific cultural group. The first was specific to Dine (Navajo) culture (Jones et al., 2006); the second was specific to the culture of an Alaskan native village (Fickel, 2005).

Aim 2: Study Quality

Quantitative analysis. Of the 10 empirical studies identified, only two studies utilized group designs that would allow quantitative, comparative analyses between intervention and control groups (Thompson & Byrnes, 2011; Vincent, Swain-Bradway, et al., 2011). Neither study featured random assignment. Thompson and Byrnes used a retrospective, nonrandomized matched control design, which is considered a causal design that yields credible results only rarely, under certain conditions (Gottfredson et al., 2015); however, these conditions were not met. Specifically, nonrandomized matched control designs that 1) statistically adjust for baseline measures of the outcome and important covariates and 2) ensure that assignment is not by self-selection, but by some other factor may be adequate to support causal statements regarding the intervention. Due to the retrospective design of the study by Thompson and Byrnes, it was not possible to collect baseline measures. This weakness in the design seems to be particularly important in this study, as matching on demographics was limited (i.e., there was not sufficient variability to match fully on gender or race/ethnicity). In addition, the teacher assignment mechanism to the district-sponsored cultural competency trainings was not reported, but was presumably by self-selection and not universally mandated, as there was a pool of teachers within the same schools to choose from who did not participate in the trainings. Moreover, the subsequent comparative analysis was not conducted at the level of assignment; teachers (level 2) were assigned intervention or control status, but the analysis was conducted at level 1 (students) – that is, comparative analyses (ANOVAs) were conducted on participating teachers' *students'*

survey responses on a measure of mutual friendship across gender, race, and socioeconomic status. Yet to mitigate selection bias, it is necessary that the analysis assess the treatment effect at the level at which assignment took place (Gottfredson et al., 2015). Thus, the study design met neither of these conditions for causal inference and internal validity of the study could not be established. With regard to the psychometric properties of the outcome measure (i.e., to establish construct validity), none were reported. In terms of effectiveness and readiness for implementation, the intervention was not well-described in the study, and only nominal participant demographics were provided, precluding assessment of external validity.

In the only other quantitative, comparative group design study, Vincent, Swain-Bradway, and colleagues (2011) conducted a quasi-experimental secondary analysis of 153 intervention and control group elementary schools following schools' voluntary implementation of School-Wide Positive Behavior Support (SW-PBS). The outcome of interest was *disparities* in disciplinary actions, with the hypothesis that SW-PBS would be an effective, school-wide in-service intervention approach to reduce disciplinary inequity. The mechanism for assignment to intervention (i.e., implementers of SW-PBS) and control (i.e., non-implementers) for purposes of the analysis was either of two measures of implementation fidelity, the School-wide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001) and the Team Implementation Checklist (TIC; Sugai, Horner, & Lewis-Palmer, 2001), with thresholds set at 80% fidelity. As there was no randomization or other methodological approach to adjust for systematic bias, the design did not meet standards of evidence for statements regarding efficacy. Therefore, both studies (Thompson & Byrnes, 2011; Vincent et al., 2011) were categorized as 'does not meet standards of evidence' within the WWC framework.

Qualitative analysis. The remaining eight studies qualitatively examined the impact of in-service interventions through ethnographic fieldwork, case study analyses, grounded theory methods of analyzing open-ended survey responses, and document review (Eberly, Joshi, Jonzal, & Galen, 2010; Fickel, 2005; Hammerness & Matsko, 2013; Jones, Caravaca, Cizek, Horner, & Vincent, 2006; McAllister & Irvine, 2002; McCormick, Eick, &Womack, 2013; Shriberg, Schumacher, McMahon, Flores, Moy, Swidzinski, & Tompkins, 2012) or employed mixed methods (using quantitative descriptive data to support case study findings; Ryan, Chandler, & Samuels, 2007). These studies yielded meaningful and useful knowledge regarding in-service approaches to promote culturally responsive practices in schools, which we discuss in greater depth in the Aim 3 results. Yet it is important to underscore that none of these studies were eligible for review within the WWC framework. Specifically, none attempted to draw conclusions relative to a comparison group of any kind (equivalent or not), none collected baseline information, and none attempted to avoid problems due to selection bias. Nonetheless, in subjecting the eight studies to quality analysis using a credibility framework for qualitative research (Brantlinger et al., 2005; Creswell & Miller, 2000), we were able to identify a number of strengths within this body of research that are noteworthy (see Table 3 for an overview). We summarize the quality analysis below.

***Triangulation*.** In examining studies for this feature, we searched for researchers' use of convergence and consistency across multiple and varied data sources to support conclusions. A strong example of triangulation was found in the study by Ryan et al. (2007), in which multiple focus groups with different target groups, multiple interviews with varied stakeholders, a document review to assess evaluation plans and reports, and video analysis of a school forum all

were employed to draw conclusions. This was a fairly common research practice, with six of eight studies clearly demonstrating use of triangulation.

Disconfirming evidence. We identified no studies that presented evidence inconsistent with researchers' identified themes (i.e., outliers). The study by Eberly et al. (2010) highlighted findings which showed the difficulties that participants experienced in translating this new knowledge into practice in their classrooms, which was in contrast to conclusions that teacher participants in the professional development experienced improvements in knowledge of CRP; however, an instance of disconfirming evidence would have been if the researchers also presented an anomalous case in which one teacher had been able to translate new knowledge.

Researcher reflexivity. Similarly, researchers' use of reflexivity, or attempts to self-disclose their perspective, were limited in the articles reviewed. Although several studies included direct mention of their efforts to mitigate researcher bias (e.g., McCormick et al., 2013), and others openly disclosed their theoretical orientation, values, and/or positionality within the research (e.g., Hammerness & Matsko, 2013; Shriberg et al., 2012), we identified none that provided a thoughtful discussion of how the researchers' values, perspectives, and assumptions may have biased or otherwise influenced the conclusions drawn.

Member checks. Reported use of member checks (in which participants review and validate transcripts, analyses, and interpretations of data) was found in two studies (Eberly et al., 2010; Shriberg et al., 2012). This was a particular strength of the Shriberg et al. (2012) article, which formally collected critical feedback on findings through a participatory framework, including with parents, at several points in the two-year study, as well as informally collected extensive feedback at regularly scheduled intervals.

Collaborative work. The majority of studies (six of eight) employed multiple researchers or other partners in designing a study or drawing conclusions, including a number of university partnerships, with one highlighting the role of the schools as part of the research team within a participatory action research frame (Ryan et al, 2007).

External auditors/peer debriefing. Four articles reported use of external auditors or a peer debriefing process, such as having a colleague or 'outsider' provide critical feedback on interpretations and analyses (Fickel, 2005; McAllister & Irvine, 2002; McCormick et al., 2013; Shriberg et al., 2012).

Audit trail. Two articles reported evidence of an audit trail (Fickel, 2005; McAllister & Irvine, 2002). For example, in McAllister and Irvine (2002), the researchers reported keeping a history of all coding changes and a journal of methodological decisions and processes.

Prolonged field engagement. The majority (seven of eight studies) featured prolonged field engagement (reported repeated contact in the field over a specific or phased time frame), including several in which the researchers were engaged in the field for over one year (Fickel, 2005; McCormick et al., 2013; Ryan et al., 2007; Shriberg et al., 2013).

Thick, detailed description. Six of eight studies featured rich, descriptive information to support themes and conclusions. For example, the case study conducted by Eberly et al. (2010) provided extensive quotes in its final evaluation phase to document the specific strategies employed in the intervention (e.g., simulation exercises to uncover bias, journaling, parent panel, role plays, group and individual analyses of case examples) and to differentiate their development along a continuum and highlight the challenges of putting into practice new learning.

Particularizability. Across the eight studies, only half provided sufficient information regarding the particular context of the study such that readers could determine the extent to which conclusions could be generalized to their own contexts. That is, four studies provided information about at least four of these five dimensions: population density/geographic region, sample size, school grade-level, school sector (e.g., public, parochial), and teacher and/or principal sample demographics (Fickel, 2005; McAllister & Irvine, 2002; McCormick et al., 2013; Shriberg et al., 2012). Specifically, five studies reported being set in major metropolitan or other urban school districts (Hammerness & Matsko, 2013; McAllister & Irvine, 2002; Ryan et al., 2007; Shriberg et al., 2012); one study was set in a small, but increasingly diverse school district (McCormick et al., 2013); and, three studies were set in a Native Alaskan village (Fickel, 2005) or other rural, American Indian settings (Jones et al, 2006; Ryan et al., 2007[1]). One study did not provide information regarding the geographic location or population density of the study setting (Eberly et al., 2010). The grade-level school settings included primarily elementary schools (six studies: Eberly et al., 2010; Fickel, 2005; Jones et al., 2006; McCormick et al., 2013; Thompson & Byrnes, 2011; Vincent, Swain-Bradway, et al., 2011), with one study set in a combined elementary and middle (pre-K to 8) school (Shriberg et al., 2012) and another in both elementary and middle schools (McAllister & Irvine, 2002), none set in high school, and two studies with unspecified grade-levels (Hammerness & Matsko, 2013; Ryan et al., 2007). Most studies were conducted with faculty from public schools; only one was conducted with faculty from a private, Catholic school (Shriberg et al., 2012), and two did not specify (Eberly et al., 2010; Fickel, 2005). Study sample sizes also widely varied, from small (i.e., < 15 individuals; Hammerness & Matsko, 2013; Shriberg et al., 2012) to moderate (< 300 individuals; Eberly et al,

[1] The study by Ryan and colleagues (2007) was set in three urban schools and one Navajo reservation school.

2010; Fickel, 2005; McAllister & Irvine, 2002; McCormick et al., 2013). Notably, two of the eight studies did not specify the total sample size (Jones et al., 2006; Ryan et al., 2007). Despite the diversity of settings and sample sizes, sample demographics were often reflective of the demographics of the teaching workforce in the U.S. (i.e., predominantly White, female teachers). Four of the eight studies included primarily White, female teachers (Eberly et al., 2010; Fickel, 2005; McCormick et al., 2013; Shriberg et al., 2012), whereas only one study focused on African American female teachers (McAllister & Irvine, 2002). The remaining three studies did not specify the demographic characteristics of their teacher or principal sample (Hammerness & Matsko, 2013; Jones et al., 2006; Ryan et al., 2007).

Aim 3: Impact of In-Service Interventions

In this section, we briefly summarize the definitions, measures and other criteria used to assess impact, then highlight intervention strategies that show potential for effectiveness.

Definitions of culturally responsive practices. Across the 10 studies, a wide range of foci were identified to operationalize CRP. Three studies operationalized CRP as related to school improvement procedures (school-wide systemic focus; Jones et al., 2006; Ryan et al., 2007; Vincent et al., 2011); three studies operationalized CRP as tapping students cultural 'funds of knowledge' (Moll et al., 1992) to transform the traditional curriculum or otherwise modifying pedagogical approaches in the classroom (academic instruction focus; Fickel, 2005; McAllister & Irvine, 2002; McCormick et al., 2013); two studies operationalized CRP as responsive cross-cultural parent-teacher interactions (school-family partnership focus; Eberly et al., 2010; Shriberg et al., 2012); one study operationalized CRP as a "context-specific…means of looking at general issues through a lens of the particular—such as how race, ethnicity and language can influence one's teaching decisions," (teacher induction and behavior management focus;

Hammerness & Matsko, 2013, p. 561); and one study operationalized CRP in a way that incorporated elements from several of these categories (Thompson & Byrnes, 2011).

Outcome measures. Three studies examined school effects, whereas the remaining studies reported effects based on individual student or teacher report. School effects were assessed through racial equity in office disciplinary referrals (Vincent et al., 2011), staff reported school improvement and school accountability measures (Ryan et al., 2007), and staff reported integration of culture into school behavior management system (Jones et al., 2006). The majority of studies (eight out of 10) did not examine student outcomes, but those that did assessed student knowledge of behavioral expectations (Jones et al., 2006) and student-perceived social inclusiveness among students (Thompson & Byrnes, 2011). Of the 7 studies that reported staff outcomes, six focused on CRP skills and knowledge/beliefs, and one focused on CRP knowledge/beliefs only. Among those studies focused on staff outcomes, the measures and methods of assessing CRP skills and knowledge were inconsistent and ranged in mode of administration (e.g., open-ended survey questions, semi-structured interview protocols, closed-ended self-report on a Likert-type scale, document review of staff journals, and observation of staff behaviors). Of the studies that used a quantitative instrument to assess CRP, none reported on the psychometric properties of the measure. There was overlap between only two studies in their use of a specific outcome indicator (i.e., disciplinary outcomes of PBIS intervention; Jones et al., 2006; Vincent, Swain-Bradway et al., 2011), but no studies used the same instrument.

Study conclusions. It is important to consider the reported study conclusions in light of our findings on study quality in Aim 2. Because eight of the 10 studies in this review failed to meet eligibility criteria for the quantitative study analysis (our application of the Institute of Sciences' WWC methodology), and the other two were analyzed but nonetheless were found not

to meet standards of evidence, we cannot identify any of the interventions as evidence-based per se, nor can we corroborate any conclusions the authors may have drawn regarding intervention effects. Nevertheless, we summarized the 10 studies' conclusions here with this caveat, and present implications of the study conclusions in the discussion.

All ten studies concluded that the studied interventions were associated with at least some gains in schools' and teachers' knowledge, skills, or use of CRP or improvement in equity with regard to student outcomes. Most studies provided some acknowledgement of limitations regarding causal inference or generalizability, including Hammerness and Matsko (2013), Jones et al. (2006), McAllister and Irvine (2002), Thompson and Byrnes (2011), Vincent, Swain-Bradway, et al. (2011). The two quantitative papers applied a test of significance ($p < .05$) in examining improvements in for the intervention groups relative to a comparison group (Thompson & Byrnes, 2011; Vincent, Swain-Bradway, et al., 2011). A small number of studies also highlighted findings that the intervention was not sufficient alone to achieve success in working cross-culturally with diverse students or their parents. For example, one study noted that teachers struggled to translate their newly acquired cultural knowledge into their classroom practice (Eberly et al., 2010), and another noted that the focus of the intervention (building teacher empathy) was necessary but not sufficient to promote CRP (McAllister & Irvine, 2002). Another study that focused on school-wide outcomes (i.e., Vincent, Swain-Bradway, et al., 2011) concluded that the intervention was associated with reduced the racial/ethnic gap in office disciplinary referrals, but these gaps were not eliminated. In light of limitations in the study designs, we highlight several strategies that emerged from this review as potentially promising directions for future testing in the discussion.

Discussion

Our findings suggest that empirical research examining the impact of interventions to improve culturally responsive practices is in a relatively nascent stage, with the majority of studies (6 of 10) being published in the past five years. We identified a very limited pool of only 10 out of 179 unduplicated peer-reviewed articles reporting results of an empirical examination of teacher, principal, or school-targeted intervention to promote CRP and equitable student outcomes. The 10 studies identified employed designs which either did not meet WWC criteria for review (i.e., 8 qualitative of mixed method studies) or met criteria for review but failed to meet standards of evidence for efficacy (i.e., 2 studies). With regard to external validity (i.e., generalizability in quantitative research or particularizability in qualitative research), several studies neglected to report basic information regarding the sample, context, and procedures. With regard to construct validity and psychometrics, differing indicators were used across 8 of 10 studies, no studies used the same measures, and report on the psychometrics of the measures was extremely limited. Taken together, the findings suggest that the state of the science of in-service educator interventions to improve CRP is inadequate to draw conclusions regarding efficacy, effectiveness, and readiness for dissemination (Flay et al., 2005; Gottfredson et al., 2015).

Credit should be given to researchers attempting to implement and assess CRP interventions despite several conceptual, methodological, and practical complexities (e.g., operationalizing and measuring CRP, measurement of outcomes at multiple levels for teachers and students, working within parameters of schools' often bureaucratic structures). As classrooms in U.S. public schools become increasingly culturally, racially, and ethnically diverse landscapes for teaching and learning, the potential of CRP to support equity and excellence in student outcomes becomes increasingly important. Further rigorous empirical research is needed

to identify and refine the extant promising models of intervention and support to promote school staff use of CRP. The findings of the systematic review identified several future research directions to strengthen the evidence-base and realize the potential of CRP.

Future Research Directions

Measurement. The wide variation in operational definitions of CRP suggests more work is needed to achieve clarity and promote consensus on the proximal (teacher-level) outcome of in-service CRP interventions. Specifically, we are advocating for consistent operationalization and agreement upon a well-defined set of indicators that can serve as a comprehensive and integrative reflection of CRP, rather than the generation of new or additional definitions. There is certainly no shortage of definitions in the literature (see Brown-Jeffy & Cooper, 2011 for a review). We acknowledge the complexity of CRP and the need to implement CRP flexibly based on the local context of school and diverse backgrounds of students; as such, we suggest this measure may be better presented as a multi-dimensional index, rather than a unidimensional scale. We suggest the underlying principles of CRP, consistent with the definition we highlighted for the purpose of this study (Gay, 2000), can be measured and that greater consensus in the area of CRP outcome measurement would help to advance the field and would hold researchers and interventionists accountable to the same high and comprehensive standard of CRP implementation. A focus on establishing measures that demonstrate content validity is a critical next step in our efforts to establish an evidence-base in support of CRP and in-service CRP interventions.

Similarly, defining and operationalizing the more distal, student-level outcomes theorized to be impacted by in-service CRP interventions requires attention. What specific student outcomes are theorized to be impacted by teachers' and administrators' effective use of CRP

(e.g., social, emotional, academic, disciplinary outcomes), and do these vary by the specific type of CRP (e.g., culturally responsive pedagogy, culturally responsive classroom behavior management, culturally responsive parent-teacher interactions)? If more distal, student-level outcomes of CRP interventions are to be studied, there needs to be consideration of how quickly, and to what extent, we can expect to see an impact on students who are not directly exposed to the intervention, and what mechanisms could be put in place to track teacher fidelity of implementation of CRP in their classrooms.

Finally, attention is needed not just to improvement in student outcomes, but also to track progress towards equity as an outcome itself. Measuring equity as an outcome is complex and raises questions about appropriate comparisons (e.g., equitable relative White students, to all students, or to some specified standard for educational excellence). Researchers assessing whether intervention to promote CRP yielded more equitable outcomes may benefit from interdisciplinary exchange with the field of public health, which has extensively examined methodological issues related to the measurement of equity (i.e., health disparities; Keppel et al., 2005).

It is worthwhile to consider what may be the most appropriate mode of administering assessments of CRP practices, or whether a multi-informant approach is necessary. Self-report of CRP is known to be subject to social desirability bias (Constantine, 2001; Constantine & Ladany, 2000; Granello & Wheaton, 1998; Katz, & Hoyt, 2014; Liu, Sheu & Williams, 2004; Ohm & Rosen, 2011; Sodowsky, Kuo-Jackson, Richardson, & Corey, 1998). Social desirability bias suggests that teachers may report implementing practices they believe they should be doing, regardless of whether they are in fact implementing them. Perhaps more nuanced assessments are needed to improve the validity of self-report; alternately, measures of social desirability bias

(e.g., Reynolds, 1982) can be included in teacher self-report surveys as a statistical control. Likewise, establishing valid observational measures of CRP poses challenges, including how measures can accommodate the fluidity with which CRP must be applied, contingent on the local context of the school and the varying backgrounds of students within the classroom (Debnam, Pas, Bottiani, Cash, & Bradshaw, 2015). Employing qualitative methods to assess CRP can enrich the development of content-valid, locally responsive observational measures. The solution to shortcomings of various measures may be to triangulate data collected through multiple informant and mixed method approaches. For example, perhaps a complementary, integrated set of measures for teacher report, student report, and independent observation could be developed, and a preliminary criterion of concordance between the measures established, in order to ascertain significant change in the CRP outcome. To our knowledge, the field currently lacks this type of measurement approach for CRP.

Intervention features and foci. Central, intervention-specific issues that require attention, as revealed by the systematic review, are primarily related to differences in the interventions' theories of change. A number of interventions in the present review featured a sustained approach (i.e., not just a one-time training) and all included hands-on simulation activities, role plays, case analysis, reflection exercises, direct contact with culture, and/or other interactive learning opportunities to support participants' discovery of culture, cultural difference, and bias. It is possible that these interactive approaches, as opposed to more didactic, consciousness-raising presentations or 'food and festivals' type events covering superficial aspects of a specific culture, may gain more and more lasting traction with regard to knowledge and skill in CPR. And yet, a few studies pointed out that, without supports to translate improvements in these knowledge and skill domains to more effective use of CRP in the

classroom, these efforts may be fruitless. It is interesting that the majority of interventions drew voluntary participants, ostensibly circumnavigating the important issue of motivation to change. Particularly with regard to CRP, it may be realistic to expect that the school administrators or teachers most in need of intervention also may be highly resistant to participation and lack motivation necessary to make changes. Thus, effective professional development on CRP should aim to improve motivation and self-efficacy to implement CRP in the classroom, and not only focus on knowledge alone (Gusky & Yoon, 2009). To comprehensively target all of these important mechanisms, in-service CRP interventions should include motivational coaching, job-embedded opportunities to practice and refine new skills, and methods for monitoring and reinforcing implementation in the classroom. Another area for research exploration is the determination of a sufficient dosage of the various modalities necessary to achieve significant improvements in CRP and targeted student outcomes.

Finally, an important intervention consideration is the racial/ethnic background of the teachers and administrators who will participate. Only six out of 10 papers in this review reported on the races/ethnicities of study participants. Of note, only one paper reported on a primarily non-White sample. Nationally, over 82% of teachers K-12 are White (NCES, 2016). White teachers who lack experience with diversity in their own lives may need different supports to help them consider the experiences of students from marginalized or historically marginalized backgrounds, relative to teachers who come from Black or Latino backgrounds. Beyond race/ethnicity, future research reports on CRP interventions providing more contextual and demographic information on participants, such as years of service in education, gender and gender identity, preservice exposure to CRP practices, will be critical to help draw conclusions on generalizability of the studies.

Study design. While our review found that the extant research is insufficient to identify evidence-based interventions to promote CRP and equitable student outcomes, it is worth noting that most of the individual study authors in this review concluded that gains in CRP and/or reductions in disparate student outcomes were linked to the in-service intervention. Although these studies presented evidence of statistically significant improvements or other data establishing the credibility of these findings, due to flaws in the study designs, the possibility remains that other factors, which were unaccounted for and unmeasured in the study, influenced the changes observed (Gottfredson et al., 2015). The fact that most studies concluded improvements were made underscores the importance of pursuing rigorous, empirical research methods that meet the standards of evidence for establishing intervention efficacy (see Gottfredson et al., 2015).

Limitations and Strengths of the Systematic Review

Although we took care to structure our review of the literature to address several possible threats to validity of the findings, there are some inherent limitations in conducting a systematic review that limit the generalizability of our findings. Namely, through the PICO process we followed, we eventually excluded studies based on our criteria which otherwise may have yielded findings relevant to our central research questions. For example, we excluded studies focused on improving cultural competence of school clinicians (e.g., Moore-Thomas & Day-Vines, 2008). Additionally, in our effort to locate highly credible research, we searched only the peer-reviewed literature; yet, there may be a great deal of work examining CRP in-service interventions that had not yet been published in peer reviewed journals, and which could have informed this review (e.g., My Teaching Partner; Gregory, Allen, Mikami, Hafen & Pianta, 2013). Although there is a potential loss of studies resulting from our consistent adherence to a

priori review criteria through the PICO process, this trade-off allowed us to achieve precision in our research questions, supported the internal validity of our study, and allowed us to draw meaningful conclusions from the literature on the discrete topic of teacher, principal, and school targeted in-service interventions to promote CRP and equity in schools.

The issue of how systematic reviews can lead to change is an important one embedded in the broader question of effectively disseminating evidence-based practices. Knowledge-translation scholars have identified barriers to the use of systematic reviews by policy-makers and pratitioners, including difficulty locating reviews, their lack of user-friendliness, and the "real or perceived failure to…make actionable policy recommendations," (Chambers et al., 2011, p. 132). In this case, our review highlights that more standardization and rigor in research practices is necessary before such recommendations can be made to support specific CRP and CRP interventions in schools. Thus, the practical importance of the review's findings are inward-looking, to fellow CRP researchers, to spur change in our research practices. Interventions that have proven to be effective at improving educator CRP and students' equitable and inclusive school and classroom experiences are vitally needed to close gaps in achievement and discipline. We crafted this report as a guiding framework to researchers to stimulate more rigorous methods in the development and testing of interventions to promote CRP in schools.

Conclusion

This systematic literature review assessed the state of the evidence in support of in-service interventions to promote teacher and school administrator CRP and reduce disparate student outcomes in public K-12 schools in the U.S. This review provided evidence of a substantial lack of outcome-focused research assessing the effectiveness of such in-service interventions. Of the 179 unduplicated articles yielded from the basic search, only 10 articles

reported on an empirical examination of the impact of a CRP in-service intervention. Of these, none met standards of evidence to establish efficacy of the intervention, with the majority of studies being qualitative impact assessments. This central finding highlights the need to strengthen empirical research examining whether in-service intervention to support teachers' and administrators' CRP translate to 1) improvements in the use of CRP in schools (e.g., responsive interactions with students, classroom structure, instructional materials) and 2) equity and improvement in student academic, social, emotional, and disciplinary outcomes. Aside from inadequate study designs, an important factor hindering research progress to support evidence-based CRP is the lack of consistent, integrated, psychometrically sound measurement approaches. The review findings suggest more rigorous research approaches are needed to determine what CRP in-service interventions are effective and therefore could be disseminated and brought to scale to ensure equitable, high quality learning environments for all students.

References

(* indicates studies included in the systematic review)

Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., & Ortiz, A. (2010). Justifying and explaining disproportionality, 1968–2008: A critique of underlying views of culture. *Exceptional Children, 76,* 279-299.

Au, K. (2009). Isn't culturally responsive instruction just good teaching? *Social Education, 73,* 179–183.

Au, K., & Jordan, C. (1981). Teaching reading to Hawaiian 164 children: Finding a culturally appropriate solution. In H. Trueba, G. Guthrie, & K. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp, 69-86). Rowley, MA: Newbury House.

Auerbach, S. (2009). Walking the walk: Portraits in leadership for family engagement in urban schools. *School Community Journal, 19,* 9–32.

Bero, L. A., Grilli, R., Grimshaw, J. M., Harvey, E., Oxman, A. D., & Thomson, M. A. (1998). Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal, 317,* 465-469.

Bickman, L., Riemer, M., Brown, J. L., Jones, S. M., Flay, B. R., Li, K. K., ... & Massetti, G. M. (2009). Approaches to measuring implementation fidelity in school-based program evaluations. *Journal of Research in Character Education, 7,* 75.

Bonner, E. P., & Adams, T. L. (2012). Culturally Responsive Teaching in the Context of Mathematics: A Grounded Theory Case Study. *Journal of Mathematics Teacher Education, 15,* 25-38. doi:10.1007/s10857-011-9198-4

Bottiani, J.H., Bradshaw, C.P., Rosenberg, M.S., Hershfeldt, P.A., Pell, K.L., Debnam, K.D. (2012). Applying Double Check to Response to Intervention: Culturally responsive practices for students with learning disabilities. *Insights on Learning Disabilities, 9,* 93-107.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes results from a randomized controlled effectiveness trial in elementary schools. Journal of Positive Behavior Interventions, 12(3), 133-148.

Brantlinger, E., Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional children*, *71*, 195-207.

Brown-Jeffy, S. & Cooper, J. E. (2011). Toward a conceptual framework of culturally relevant pedagogy: An overview of the conceptual and theoretical literature. *Teacher Education Quarterly*, *38*, 65-84. Retrieved from http://www.eric.ed.gov/PDFS/EJ914924.pdf.

Bruner, J. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.

Cazden, C., & Leggett, E. (1981). Culturally responsive education: Recommendations for achieving Lau remedies II. In H. Trueba, G, Guthrie, & K. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp. 69-86). Rowley, MA: Newbury House.

Chambers, D., Wilson, P. M., Thompson, C. A., Hanbury, A., Farley, K., & Light, K. (2011). Maximizing the impact of systematic reviews in health care decision making: a systematic scoping review of knowledge-translation resources. *Milbank Quarterly*, *89*(1), 131-156.

Constantine, M. G. & Ladany, N. (2000). Self-report multicultural counseling competence scales: Their relation to social desirability attitudes and multicultural case conceptualization ability. *Journal of Counseling Psychology, 47*, 155-164. doi:10.1037//0022-0167.47.2.155

Cornell, D. G., Allen, K., & Fan, X. (2012). A randomized controlled study of the Virginia Student Threat Assessment Guidelines in kindergarten through grade 12. *School Psychology Review, 41*, 100.

Debnam, K., Pas, E., Bottiani, J.H., Cash, A.H., & Bradshaw, C.P. (2015). An examination of the association between observed and self-reported culturally responsive teaching practices. *Psychology in the Schools, 52*, 533-548.

Delpit, L. (2006). *Other people's children: Cultural conflict in the classroom*. New York, NY: The New Press.

Dray, B. & Wisneski, D. (2009). Examining diversity through mindful reflection and communication. *Focus on Teacher Education, 9*, 2–8.

*Eberly, J. L., Joshi, A., Konzal, J., & Galen, H. (2010). Crossing cultures: Considering ethnotheory in teacher thinking and practices. *Multicultural Education, 18*, 25-32.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial & Labor Relations Review, 48*, 547-561.

Epstein, T., Mayorga, E., & Nelson, J. (2011). Teaching about race in an urban history class: The effects of culturally responsive teaching. *Journal of Social Studies Research, 35*, 2-21.

Erickson, F., & Mohan, C. (1982). Cultural organization and participation structures in two classrooms of Indian students. In G. Spindler, (Ed.), *Doing the ethnography of schooling* (pp. 131-174). New York: Holt, Rinehart & Winston.

*Fickel, L. H. (2005). Teachers, tundra, and talking circles: Learning history and culture in an Alaska Native village. *Theory and Research in Social Education, 33*, 476-507. doi:10.1080/00933104.2005.10473292

Fiedler, C., Chiang, B., Van Haren, B, Jorgensen, J., Halberg, S, & Boreson, L. (2008). Culturally responsive practices in schools: A checklist to address disproportionality in special education. *Teaching Exceptional Children, 40*, 52-59.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., ... & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, *6*, 151-175. doi:10.1007/s11121-005-5553-y

Flay, B. R., Graumlich, S, Segawa, E., Burns, J. L., & Holliday, M. Y. (2004). Effects of 2 prevention programs on high-risk behaviors among African American youth. *Archives of Pediatrics and Adolescent Medicine, 158*, 377-384.

Garza, R. (2009). Latino and White high school students' perceptions of caring behaviors: Are we culturally responsive to our students? *Urban Education, 44*, 297-321. doi:10.1177/0042085908318714

Gay, G. (1995). Building cultural bridges: A bold proposal for teacher education. *Multicultural education: Strategies for implementation in colleges and universities, 4*, 95-106.

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education, 53*, 106-116. doi:10.1177/0022487102053002003

Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, *16*, 893-926. doi:10.1007/s11121-015-0555-x

Granello, D. H. & Wheaton, J. E. (1998). Self-perceived multicultural competencies of African American and European American vocational rehabilitation counselors. *Rehabilitation Counseling Bulletin*, *42*, 2-15.

Gregory, A., Allen, J., Mikami, A., Hafen, C., & Pianta. R. (2013, January). The promise of a teacher professional development program in reducing the racial disparity in classroom exclusionary discipline. Center for Civil Rights Remedies National Conference. Closing the School to Research Gap: Research to Remedies Conference. Washington, DC.

Gregory, A., & Ripski, M. B. (2008). Adolescent trust in teachers: Implications for behavior in the high school classroom. *School Psychology Review*, *37*, 337-353.

Griner, A. C., & Stewart, M. L. (2013). Addressing the achievement gap and disproportionality through the use of culturally responsive teaching practices. *Urban Education*, *48*, 585-621. doi:10.1177/0042085912456847

Guskey, T.R. & Yoon, K.S. (2009). What Works in Professional Development? American Institutes for Research, Washington, D.C. Retrieved from http://k12.wa.us/Compensation/pubdocs/Guskey2009whatworks.pdf

Gutierrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*, 19-25. doi:10.3102/0013189x032005019

*Hammerness, K., & Matsko, K. K. (2013). When context has content: A case study of new teacher induction in the University of Chicago's Urban Teacher Education Program. *Urban Education, 48*, 557-584. doi:10.1177/0042085912456848

Hammerstrøm, K., Wade, A., & Jørgensen, A. (2010). Searching for studies: A guide to information retrieval for Campbell Systematic Reviews. Retrieved from http://www.campbellcollaboration.org/resources/research/new_information_retrieval_guide.php

*Jones, C., Caravaca, L., Cizek, S., Horner, R. H., & Vincent, C. G. (2006). Culturally Responsive Schoolwide Positive Behavior Support: A case study in one school with a high proportion of Native American students. *Multiple Voices for Ethnically Diverse Exceptional Learners, 9,* 108-119.

Katz, A. D. & Hoyt, W. T. (2014). The influence of multicultural counseling competence and anti-Black prejudice on therapists' outcome expectancies. *Journal of Counseling Psychology, 61,* 299-305. doi:10.1037/a0036134

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., Wilkinson-Flicker, S., Barmer, A., and Dunlop Velez, E. (2015). *The Condition of Education 2015* (NCES 2015-144). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch

Keppel, K., Pamuk, E., Lynch, J., et al. (2005). Methodological issues in measuring health disparities. National Center for Health Statistics. *Vital Health Statistics, 141*, 495-499. Retrieved from http://www.cdc.gov/nchs/data/series/sr_02/sr02_141.pdf

King, K. A., Harris-Murri, N. J., & Artiles, A. J. (2006). *Proactive culturally responsive discipline*. Tempe, AZ: National Center for Culturally Responsive Educational Systems.

Retrieved from http://www.nccrest.org/Exemplars/exemplar_culturally_responsive_discipline.pdf

Ladson-Billings, G. (1994). *The dreamkeepers: Successful teaching for African-American students*. San Francisco: Jossey-Bass.

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

Lalas, J. (2007). Teaching for social justice in multicultural urban schools: Conceptualization and classroom implication. *Multicultural Education, 14*, 17-21.

Liu, W. M., Sheu, H. B., & Williams, K. (2004). Multicultural competency in research: Examining the relationships among multicultural competencies, research training and self-efficacy, and the multicultural environment. *Cultural Diversity and Ethnic Minority Psychology, 10*, 324-339. doi:10.1037/1099-9809.10.4.324

Losen, D. J., & Gillespie, J. (2012). Opportunities suspended: The disparate impact of disciplinary exclusion from school. Retrieved from University of California, Civil Rights Retrieved from http://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder

Losen, D.J. (2015). *Closing the school discipline gap: Equitable remedies for excessive exclusion*. New York: Teachers College Press.

Losen, D. J., & Orfield, G. (2002). *Racial inequity in special education*. Cambridge, MA: Harvard Education Publishing Group.

*McAllister, G., & Irvine, J. J. (2002). The role of empathy in teaching culturally diverse students: A qualitative study of teachers' beliefs. *Journal of Teacher Education, 53,* 433-443. doi:10.1177/002248702237397

*McCormick, T. M., Eick, C. J., & Womack, J. S. (2013). Culturally responsive teaching: Awareness and professional growth through a school-university collaboration. *School-University Partnerships, 6,* 6-14.

Moll, L., Amanti, C., Neff, D., & Gonzalez (2005). Funds of knowledge for teaching: Using a Qualitative approach to connect homes and classrooms. In N. Gonzalez, L. Moll, & C. Amanti (Eds.), *Funds of knowledge: Theorizing practices in households, communities, and classrooms* (pp. 71-88). Mahwah, NJ: Lawrence Erlbaum.

Moore-Thomas, C., & Day-Vines, N.L. (2008). Culturally competent counseling for religious and spiritual African American adolescents. *Professional School Counseling, 11,* 159-165. doi:10.5330/psc.n.2010-11.159

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities Are Disproportionately Underrepresented in Special Education Longitudinal Evidence Across Five Disability Conditions. *Educational Researcher, 44,* 278-292.

National Center for Education Statistics (NCES). (2013). *Schools and staffing survey (SASS): Public school teacher data file 2011–12.* U.S. Department of Education. Retrieved from https://nces.ed.gov/surveys/sass/tables/sass1112_2013314_t1s_001.asp

National Center for Education Statistics (NCES). (2016). *The State of Racial Diversity in the Educator Workforce.* U.S. Department of Education. Retrieved from

http://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversityworkforce.pdf

Nieto, S. (2013). *Finding joy in teaching students of diverse backgrounds: Culturally responsive and socially just practices in U.S. classrooms*. Portsmouth, NH: Heinemann.

Ogbu, J. U. (1994). Racial stratification and education in the United States: Why inequality persists. *Teachers College Record, 96*, 264-298.

Ohm, R., & Rosen, L. (2011). Psychometric evaluation of the perceived prejudice in health care scale–modified (PPHC-M) among baccalaureate student nurses. *Journal of Transcultural Nursing*, *22*, 282-289. doi:10.1177/1043659611404430

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*, 93-97. doi:10.3102/0013189x12441244

Pettigrew, M. & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden, MA: Blackwell Publishing.

Phillippo, K. (2012). "You're trying to know me": Students from nondominant groups respond to teacher personalism. *The Urban Review*, *44*, 441-467. doi:10.1007/s11256-011-0195-9

Porowski, A., O'Conner, R., & Passa, A. (2014). Disproportionality in school discipline: An assessment of trends in Maryland, 2009-12. REL 2014-017. *Regional Educational Laboratory Mid-Atlantic*. Retrieved from http://ies.ed.gov/ncee/edlabs/projects/project.asp?projectID=365

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119-125.

Richards, H., Brown, A., & Forde, T. (2007). Addressing diversity in schools: Culturally responsive pedagogy. *Teaching Exceptional Children, 39*, 64-68.

Risko, V., & Walker-Dalhouse, D. (2008). Tapping students' cultural funds of knowledge to address the achievement gap. *Reading Teacher, 61*, 98-100. doi:10.1598/rt.61.1.12

*Ryan, K. E., Chandler, M., & Samuels, M. (2007). What should school-based evaluation look like? *Studies in Educational Evaluation, 33*, 197-212. doi:10.1016/j.stueduc.2007.07.001

Sampson, D., & Garrison-Wade, D. (2011). Cultural vibrancy: Exploring the preferences of African American children toward culturally relevant and non-culturally relevant lessons. *Urban Review, 43*, 279-309. doi:10.1007/s11256-010-0170-x

Savage, C., Hindle, R., Meyer, L. H., Hynds, A., Penetito, W., & Sleeter, C. E. (2011). Culturally responsive pedagogies in the classroom: indigenous student experiences across the curriculum. *Asia-Pacific Journal of Teacher Education, 39*, 183-198. doi:10.1080/1359866x.2011.588311

Schardt, C., Adams, M., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilizing the PICO framework to improve searching PubMed for clinical questions. *Biomedical Informatics and Decision Making, 7*, 16. doi:10.1186/1472-6947-7-16

*Shriberg, D., Schumacher, R., McMahon, K. C., Flores, S., Moy, G. E., Swidzinski, J., & Tompkins, N. A. (2012). Utilizing participatory action research to foster effective family/school collaboration at an urban pre-K-8 Catholic school. *Catholic Education: A Journal of Inquiry and Practice, 15*, 226-252.

Skiba, R., Horner, R., Chung, C., Rausch, M., May, S., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*, 85-107.

Sodowsky, G. R., Kuo-Jackson, P. Y., Richardson, M. F., & Corey, A. T. (1998). Correlates of self-reported multicultural competencies: Counselor multicultural social desirability,

race, social inadequacy, locus of control racial ideology, and multicultural training. *Journal of Counseling Psychology, 45*, 256-264. doi:10.1037//0022-0167.45.3.256

Sugai, G., & Horner, R. H. (2002). The evolution of discipline practices: School-wide positive behavior supports. In J. K. Luiselli & C. Diament (Eds.), *Behavior psychology in the schools: Innovations in evaluation, support, and consultation* (pp. 23–50). New York, NY: Haworth.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, *77*, 317-334. doi:10.1177/001440291107700304

*Thompson, J., & Byrnes, D. (2011). A more diverse circle of friends. *Multicultural Perspectives, 13*, 93-99. doi:10.1080/15210960.2011.571552

Toldson, I. A., & Lemmons, B. P. (2013). Social demographics, the school environment, and parenting practices associated with parents' participation in schools and academic success among Black, Hispanic, and White students. *Journal of Human Behavior in the Social Environment, 23*, 237-255. doi:10.1080/10911359.2013.747407

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, *What Works Clearinghouse*. Retrieved on July 25, 2016 from http://ies.ed.gov/ncee/wwc

Villegas, A. (1988). School failure and cultural mismatch: Another view. *The Urban Review, 20*, 153-265. doi:10.1007/bf01120137

Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers rethinking the curriculum. *Journal of Teacher Education, 53*, 20-32.

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*, 219-229. doi:10.1177/1098300711399765

*Vincent, C. G., Swain-Bradway, J., Tobin, T. J., & May, S. (2011). Disciplinary Referrals for Culturally and Linguistically Diverse Students with and without Disabilities: Patterns Resulting from School-Wide Positive Behavior Support. *Exceptionality, 19*, 175-190. doi:10.1080/09362835.2011.579936

Wald, J. & Losen, D. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth, 2003*, 9-15. doi:10.1002/yd.51

Ware, F. (2006). Warm demander pedagogy: Culturally responsive teaching that supports a culture of achievement for African American students. *Urban Education, 41*, 427-456. doi:10.1177/0042085906289710

Weinstein, C.S., Tomlinson-Clarke, S., & Curran, M. (2004). Toward a conception of culturally responsive classroom management. *Journal of Teacher Education, 55*, 25-38. doi:10.1177/0022487103259812