

Georgia State University

ScholarWorks @ Georgia State University

Counseling and Psychological Services
Dissertations

Department of Counseling and Psychological
Services

Summer 8-11-2020

Cultural Humility in the Context of Teacher-Student Relationships

Kathryn M. McPhee
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/cps_diss

Recommended Citation

McPhee, Kathryn M., "Cultural Humility in the Context of Teacher-Student Relationships." Dissertation, Georgia State University, 2020.

doi: <https://doi.org/10.57709/18738009>

This Dissertation is brought to you for free and open access by the Department of Counseling and Psychological Services at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Counseling and Psychological Services Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ACCEPTANCE

This dissertation, CULTURAL HUMILITY IN THE CONTEXT OF TEACHER-STUDENT RELATIONSHIPS, by KATHRYN MCPHEE, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education & Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

Joel Meyers, Ph.D.
Committee Chair

Kenneth Rice, Ph.D.
Committee Member

Emily Graybill, Ph.D.
Committee Member

Don Davis, Ph.D.
Committee Member

Date

Brian J. Dew, Ph.D.
Chairperson, Department of Counseling and
Psychological Services

Paul A. Alberto, Ph.D.
Dean, College of Education &
Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education & Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

Kathryn McPhee

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Kathryn Mary McPhee
Counseling and Psychological Services
College of Education & Human Development
Georgia State University

The director of this dissertation is:

Joel Meyers
Counseling and Psychological Services
College of Education & Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Kate McPhee

ADDRESS: 121 Crumley Street SE
Atlanta, GA 30312

EDUCATION:

Ph.D.	2020	Georgia State University Counseling and Psychological Services
Master's Degree	2016	Georgia State University School Psychology
Bachelor's Degree	2011	University of Florida English

PROFESSIONAL EXPERIENCE:

2019-Present	Psychology Intern Psychological and Social Services, Dallas Independent School District
2018-2019	Psychology Intern Georgia Network for Educational and Therapeutic Schools, Dekalb-Rockdale County
2017-2018	Intern Dream Big Center for Learning and Development
2016-2017	School Psychology Advanced Practicum Atlanta Public Schools
2015-2016	School Psychology Practicum Clayton County Public Schools

PRESENTATIONS AND PUBLICATIONS:

Srisarajivakul, E. N., Choe, E., **McPhee, K.**, Varjas, K., Meyers, J., Rice, K., & Davis, D. (2019, April). *Validity of the Cultural Humility Scale for Students (CHS-S)*. Poster presented at the American Educational Research Association Convention, Toronto, Canada.

McPhee, K., Srisarajivakul, E. N., Rice, K. G., & Meyers, J. (2018, August). *Assessing levels of mental illness stigma in diverse adolescent populations*. Poster session presented at the 126th Annual American Psychological Association Convention, San Francisco, CA.

McPhee, K., Srisarajivakul, E. N., Ramirez, V., Guillen, N., Case, M., Varjas, K., Perkins, C., & Meyers, J. (2018, February). *Leveraging the power of relationships to improve school outcomes*. Mini-skills session presented at the National Association of School Psychologists Conference, Chicago, IL.

Srisarajivakul, E. N., **McPhee, K.,** Case, M., Hanna, M., Rice, K., Varjas, K., & Meyers, J. (2018, February). *Assessing levels of mental health stigma among diverse adolescent populations*. Poster presented at the National Association of School Psychologists Conference, Chicago, IL.

Givhan, K., **McPhee, K.,** Ramirez, V., Pimpalkhute, S., Wilkins, S., Hanna, M., Varjas, K., Perkins, C., & Meyers, J. (2018, February). *Cultural mismatch: Bridging the gap between school personnel and students*. Mini-skills session presented at the National Association of School Psychologists Conference, Chicago, IL.

McPhee, K., Srisarajivakul, E. N., Ramirez, V., Varjas, K., Perkins, C., & Meyers, J. (2017, November). *Teacher-student relationships: Strengthening the impact of PBIS on school climate*. Paper presented at the 10th Annual Conference of the Georgia Association for Positive Behavior Support, Duluth, GA.

McPhee, K., Givhan, K., Srisarajivakul, E. N., Taylor, E., Ramirez, V., Varjas, K., Perkins, C. & Meyers, J. (2017, February). *Improving school climate through teacher-student relationships: PBIS stakeholder perspectives*. Paper presented at the National Association of School Psychologists Conference, San Antonio, TX.

Zabek, F., Harper, E. A., **McPhee, K.,** Henry, S., Givhan, K., Perkins, C., Meyers, J., & Kruger, A. C. (2017, February). *School-based prevention of CSEC and other gendered aggression*. Mini-skills session presented at the National Association of School Psychologists Annual Convention, San Antonio, TX.

McPhee, K., Givhan, K., Srisarajivakul, E. N., Taylor, E., Ramirez, V., Varjas, K., Perkins, C., Rice, K., & Meyers, J. (2016, December). *PBIS teams' perceptions of connections between culture and PBIS implementation*. Paper presented at the 9th Annual Conference of the Georgia Association for Positive Behavior Support, Duluth, GA.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2018-Present	American Psychological Association
2014-Present	National Association of School Psychologists
2012-Present	Leadership for Educational Equality Network

CULTURAL HUMILITY IN THE CONTEXT OF TEACHER-STUDENT RELATIONSHIPS

by

KATE MCPHEE

Under the Direction of Joel Meyers, Ph.D.

ABSTRACT

Though cultural humility is emerging as an important quality for helping professionals across many fields of research, it has been minimally explored in the context of primary and secondary education settings. The purpose of this dissertation is to explore the implications of cultural humility for the professional development of teachers and to investigate the impact of cultural humility for teachers who instruct children with cultural backgrounds that differ from their own. A theoretical argument for incorporating cultural humility into teacher professional development programs is presented. This theoretical perspective is examined by extending a recent systematic review of research on inservice training to address teacher cultural responsiveness by incorporating the construct of cultural humility into this review. Review findings suggest that while the literature on teacher professional development for cultural responsiveness is growing,

there are still many methodological and theoretical concerns that limit the evidence-based nature of this type of professional development for teachers. A key finding is that cultural humility may be a construct that can strengthen the research and practice of professional development addressing culturally responsive teaching.

To further investigate the potential importance of cultural humility in education, an empirical study was designed to develop a data base about the importance of cultural humility for teachers who instruct students with cultural backgrounds different from their own. An adapted measure of cultural humility was used along with measures of teacher-students relationships and externalizing behavior. Data were collected in a racially and socioeconomically diverse school district. Students were asked rate their teacher's cultural humility as well as the quality of their relationship with the teacher. Results suggested that teacher cultural humility is a statistically significant, positive moderator of the relationship between students' perceptions of teacher-student relationship quality and externalizing behavior. The magnitude of the moderating effect was significantly stronger with boys. Implications are considered for research and the development of strategies to use cultural humility to promote culturally responsive teacher student relationships with students from culturally diverse backgrounds.

INDEX WORDS: Cultural Humility, Teacher Professional Development, Teacher-Student Relationships

CULTURAL HUMILITY IN THE CONTEXT OF TEACHER-STUDENT RELATIONSHIPS

by

Kate McPhee

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

School Psychology

in

Counseling and Psychological Services

in

the College of Education & Human Development

Georgia State University

Atlanta, GA

2019

Copyright by
K M McPhee
2019

DEDICATION

This dissertation is dedicated first to my husband, Justin, who has supported my work on this project with incredible generosity and encouragement. Justin – doing a PhD wasn't in the plan when I went to graduate school five years ago, but as my passions for school psychology grew and changed, you supported me in pursuit of this degree and all that went into it. I would not have made it without your patience, kind words, cooking skills, and flexibility. You have truly labored through the dissertation process with me, more than earning the husband/honey/helper “h” in PhD. I love you; thank you.

It is also dedicated to my parents – especially my mom, who went through this process recently and could share in the frustrations and joys that come during data analysis, writing, and revising (and revising again). Mom – I am so proud NOT to be the first woman in our family with a doctoral degree. Mom and Dad – your love and support have meant so much to me, not just in this portion of my school career but throughout the whole thing. All the papers you read and assemblies you attended; I always knew you were in my corner rooting for me. Thank you. I also have the best parents-in-law that one could ever ask for, who have been so incredibly supportive of me completing this crazy process. Gordon and Elva – thank you for always being generally compassionate, generous, and encouraging; it has truly helped fuel me.

Finally, this project is dedicated to the students I taught during my first two years in the world of education. The questions I am exploring in this paper were directly inspired by our time in the classroom together. I learned so much from you that I had to go do something with it! I humbly hope that this paper is one step in a long march towards better training and information for teachers like me who are teaching amazing students like you. Thank you.

ACKNOWLEDGMENTS

I would first like to acknowledge the dedication and hard work of Joel Meyers, who is one of the kindest and humblest people I have ever known. Dr. Meyers – you use your extensive experience and expertise not to boost your own reputation, but to encourage others and to make positive change. Your approach to your work is inspiring to me, and I am honored to have gotten the opportunity to work with you and learn from you for the past six years. Thank you for guiding me through the doctoral process with gentleness and great care. You encouraged me to pursue this degree in the first place, and so I would not be here without you! Thank you.

I also want to acknowledge the rest of the school psychology faculty at Georgia State for your guidance, dedicated training, and encouragement. Thank you to my committee members Kenneth Rice, Don Davis, and Emily Graybill for your diligence, time, and expertise. I would like to specifically thank Dr. Rice who spent several hours helping me figure out this data and the best analytical approach (all with his usual humor to keep things light-hearted). Thank you to the Project AWARE team and the Center for Research on School Safety for allowing me the opportunity to explore these important questions through research, and for generally being open to students' passions and ideas. I am grateful to the students who completed the survey and to school and grant personnel who helped ensure the data was collected.

I absolutely must acknowledge the other half of team KEmily, Emily Srisarajivakul. Emily, you are my sister and I cannot imagine school or life without you. Thank you for your friendship, critical feedback and for being a safe space for me and many others. Thank you also to other GSU students for their encouragement, support, and their “bearing with” me – DJ, Junaid, Jo, Kiriar, Erin, Kiarra, and Dustin stand out in my memory, but many, many others inspire and encourage me. I am proud to be among you.

TABLE OF CONTENTS

LIST OF TABLES	IV
LIST OF FIGURES	V
1 CULTURAL HUMILITY IN TEACHER PROFESSIONAL DEVELOPMENT	1
GUIDING QUESTIONS	15
METHOD	16
RESULTS	18
DISCUSSION	40
REFERENCES.....	50
2 CULTURAL HUMILITY IN THE CONTEXT OF TEACHER-STUDENT	
RELATIONSHIPS.....	69
METHOD	79
RESULTS	88
DISCUSSION	99
REFERENCES.....	107
APPENDIX.....	134

LIST OF TABLES

Table 1.1. Attributes of Cultural Humility Integrated with Culturally Responsive Teaching.....	60
Table 1.2. Sixteen studies of the Effects of Inservice Interventions to Promote Culturally Responsive Practices.....	61
Table 1.3. Evidence of Credibility Indicators in Seven Qualitative Studies of Inservice Interventions to Promote CRT	66
Table 1.4. Evidence of Credibility Indicators in Nine Qualitative Studies of Inservice Interventions to Promote CRT	67
Table 1.5. Evidence of Encouraging Cultural Humility in all 26 Studies	68
Table 2.1. Demographics of Participants	118
Table 2.2. Teacher Demographics (As Reported by Students)	119
Table 2.3. Participants' Most Salient Cultural Identities	120
Table 2.4. Item Means and Factor Loadings of CHS-S Final Model	121
Table 2.5. Item Means and Factor Loadings of Conduct Problems Model	122
Table 2.6. Item Means and Factor Loadings of IT-SR Final Model	123
Table 2.7. Fit Statistics for Measurement Models	124
Table 2.8. Measurement Invariance Model Fit Indices	125
Table 2.9. Correlations of Latent Variables	127
Table 2.10. Interactions and Main Effects on Teacher-Student Relationship Quality	128
Table 2.11. Wald Chi-Square Values for Subgroup Model Tests	129

LIST OF FIGURES

Figure 2.1 Theoretical Moderation Model of Cultural Humility’s Role in Teacher-Student Relationship Quality	130
Figure 2.2 Interaction of Cultural Humility and Conduct Problems on TSRQ	131
Figure 2.3 Three-way Interaction Between Gender, Conduct Problems, and Cultural Humility on Trust	132
Figure 2.4 Three-way Interaction Between Gender, Conduct Problems, and Cultural Humility on Communication	133

1 CULTURAL HUMILITY IN TEACHER PROFESSIONAL DEVELOPMENT

Changing demographics in the US student population indicate that classrooms are now more racially and socioeconomically diverse than ever (Musu-Gillette, de Brey, McFarland, Husar, Sonnenberg, & Wilkinson-Flicker, 2017), and that the proportion of students from racial minority groups in schools is growing at an unprecedented rate (Musu-Gillette et al., 2017). Despite this student population shift, teacher demographics remain primarily White and female (USDOE, 2016). This change in student demographics is troubling because researchers have identified clear and persistent disproportionality in discipline practices and special education referrals across the country that affect children from racial/ethnic minority and/or low socioeconomic backgrounds (e.g. Skiba, Horner, Chung, Rausch, May, & Tobin, 2011; Smolkowski, Girvan, McIntosh, Nese, & Horner, 2016; Sullivan & Bal, 2013). In order to correct this disproportionality and provide educational equity for all student populations, teachers will likely need to change their pedagogical practices to provide effective education to students from a variety of cultural backgrounds (Ladson-Billings, 1995b, 2014).

Cultural humility, or an attitude of openness towards others of different backgrounds, has proven to be a promising concept in multiple helping fields (e.g., Foronda, Baptiste, Reinholdt, and Ousman, 2016; Mosher, Hook, Davis, Deblaere, Captari, & Owen, 2017) but is only beginning to emerge as a topic of interest in education (Brown et al., 2016; Lund & Lee, 2015; Tinkler & Tinkler, 2016). Cultural humility is part of a multicultural orientation framework that describes a general unassuming and unoppressive way of being with others (Owen, Tao, Leach, & Rodolfa, 2011). Practitioners with cultural humility tend to form better relationships with their clients, as they are perceived as more trustworthy and commit fewer microaggressions during

service provision (Davis et al., 2018). Though the role of cultural humility has been minimally explored in education, this paper proposes that, like practitioners in other helping fields, teachers can be more effective in their practice if they work from a culturally humble orientation. This paper will integrate the concept of cultural humility with prevailing theories of culturally responsive practice and pedagogy to strengthen current thinking about effective teacher practice and to evaluate the state of teacher professional development around these topics.

To examine how cultural humility can theoretically help enhance effective teacher practice, it is necessary to understand what practices are currently being encouraged for working with racially and socioeconomically diverse students. Thus, through this paper we will also conduct an examination of the research on professional development for teacher culturally responsive teacher practice. Though culturally responsive teaching has been discussed frequently in the professional education literature, few studies have rigorously examined the effects of professional development programs on teachers' cultural responsiveness in the classroom (Bottiani, Larson, Debnam, Bischoff & Bradshaw, 2018; Parkhouse, Lu, & Massaro, 2019; Sleeter, 2012). Bottiani et al. (2018) conducted a systematic review of this literature from 1998-2014 and found that no articles met standards of generalizable, evidenced-based practice. Understanding more about how these studies, as well as any that have been published since 2014, encouraged cultural humility as part of their training efforts will help provide more guidance on how professional development can be improved in the future to help teachers serve children with more cultural humility. This paper not only extends Bottiani et al.'s work but also expands it to examine whether cultural humility is being encouraged in current inservice teacher professional development programs.

Cultural Humility and Culturally Responsive Teaching

Cultural humility may provide a productive way to strengthen culturally responsive practice and pedagogy. A helping professional with cultural humility is one who is open to others' ways of knowing the world, is self-aware, is devoted to self-reflection and critique, and is willing to dismantle power imbalances in helping relationships (Foronda et al., 2016; Tervalon & Murray-Garcia, 1998). Cultural humility is a promising concept for education, as research in other helping fields has demonstrated more effective outcomes for clients who are being served by culturally humble practitioners (Hook et al., 2013). Cultural humility has been adopted as an improved alternative to multicultural competence in fields such as medicine (Chang, Simon, & Dong, 2012; Tervalon & Murray-Garcia, 1998), social work (Fisher-Borne, Cain, & Martin, 2015), and counseling (Hook, Davis, Owen, & Deblaere, 2017; Hook, Davis, Owen, Worthington, & Utsey, 2013). However, it has only been preliminarily considered as a way to conceptualize best teacher practice in elementary and secondary education settings (Brown et al., 2016; Lund & Lee, 2015; Tinkler & Tinkler, 2016).

Cultural humility may fill a hole in the education research literature about multicultural competence when it comes to fleshing out how teachers should approach working with students who are different from them. Guo, Arthur, & Lund (2009) found that training programs for teacher cultural competence were ineffective with teachers because they emphasized a finite set of knowledge and skills that one can attain to meet the needs of all students. This was problematic as it encouraged an "other" view of someone with difference and primarily focused on integrating students into the mainstream culture instead of validating and empowering the diversity of their students' cultural experience. Parkhouse et al. (2019) reviewed the literature about multicultural teacher education and found that many studies failed to sufficiently challenge ethnocentric beliefs and attitudes that teachers held, sometimes increasing stereotyped and/or essentialist

views of students from racial and socioeconomic minority groups. Teacher training for cultural responsiveness should be shifted away from a set of practices and more to a fluid way of thinking, acting, and adapting during interactions with people who are different (Lund & Lee, 2015). Cultural humility is consistent with Lund and Lee's (2015) description, as it is an attitude that enables practitioners to engage in a variety of situations and skills from a more open, sensitive, and reflective position (Davis et al., 2018; Owen et al., 2011). Though a culturally humble professional should engage in some regular practices to enhance their service provision, cultural humility is not a skill or practice-based concept (Davis et al., 2018; Owen et al., 2011). It should be thought of as a precursor to multiculturally competent practice; a philosophical position that influences the way one carries out specific skills and interventions. Thus, cultural humility can improve theoretical and practical thinking about preparing teachers to work with a wide-range of students from an overarching and open multicultural orientation.

The components of cultural humility were operationalized by Foronda et al. (2016), who analyzed the published research on cultural humility in order to generate common attributes and consequences of the term. The resulting components of cultural humility were openness, self-awareness, egolessness, supportive interactions, and self-reflection/self-critique. Cultural humility is a subskill of overall relational humility. Whereas relational humility describes an individual's ability to be other-oriented, express positivity towards others, regulate self-oriented emotions, and accurately view himself/herself (Davis et al., 2010), cultural humility focuses these qualities one's attitude towards their own and others' cultural experiences. The consequences (outcomes) associated with culturally humble practice were mutual empowerment for helping professionals and their clients, respect, optimal care, and lifelong learning for practitioners (Foronda et al., 2016 p. 212). These components align very well with tenants of culturally

responsive practice and pedagogy for educators, as defined by the most widely used conceptualizations of cultural responsiveness (i.e., Gay, 2002; 2010; Ladson-Billings, 1994; 1995a; 1995b; 2014). Gay (2010) defined culturally responsive *practice* as “using the cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning encounters more relevant to and effective for them” (p. 31). Ladson-Billings (1994) defined culturally responsive *pedagogy* (it has been recently renamed, ‘culturally sustaining pedagogy’; Paris, 2012) as teaching that “empowers students intellectually, socially, emotionally, and politically using cultural references to impart knowledge, skills, and attitudes” (pp. 16-17). Table 1.1 summarizes the alignment of attributes of cultural humility with principles of culturally responsive practice and culturally responsive pedagogy. Hereafter, both culturally responsive practice and culturally responsive pedagogy will be combined and referred to as culturally responsive teaching (CRT; for a more detailed description of the similarities and differences between practice and pedagogy, see Aronson & Laughter, 2016).

Cultural humility requires that a practitioner be aware of his/her own cultural biases and experiences and commit to an ongoing reflection of how these biases and experiences influence his/her practice with clients (Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013). Similarly, CRT calls for teachers to build their own ongoing socio-political consciousness and grow to a full understanding of why culture and difference are vitally important to address explicitly in their classrooms (Ladson-Billings, 2014; Gay, 2010). It is a key feature in CRT theories that practitioners never expect to arrive at a fixed level of competence about the way cultural experience affects their clients’/students’ identities; instead, helpers must be constantly open to learning and improving their practice for the benefit of those they are helping (Ladson-Billings, 2014; Foronda et al., 2016).

Culturally humble practice also means that practitioners should be open to hearing others' cultural experiences instead of assuming they can be easily understood (Foronda et al., 2016). Congruent with this idea, Gay (2010) writes that CRT requires educators to examine and replace deficit perspectives of students' cultures and engage in cultural knowledge building, while teaching students to do the same. Ladson-Billings (2014) notes that constant openness is necessary for teachers to remain adept at how best to know and connect with their students.

Culturally humble practice is also characterized by supportive relationships between helpers and those whom they are helping (Foronda et al., 2016). CRT requires teachers to set high expectations for students and to commit to helping them academically, socially, and emotionally so that they can succeed (Gay, 2002, 2010; Ladson-Billings, 1995a, 1995b). Cultural humility leads to mutual empowerment of the practitioner and client as the practitioner exposes and challenges perceived power structures in the helping relationship (Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013). Similarly, CRT calls for teachers to empower students by helping them connect diverse home cultures with what is often a monolithic school culture so that students can better succeed in systems with unspoken cultural norms (Ladson-Billings, 1995a, 1995b). This can be accomplished by pointing out student strengths and helping students develop critical thinking skills and socio-political consciousness about education and their place in society, thereby enabling students to take pride in their cultures and better advocate for themselves at school.

Finally, the purposes for practicing with cultural humility and CRT are shared, as both aim to enhance service provision for all cultural groups. Cultural humility calls for practitioners to work from an unassuming standpoint in order to encourage optimal and equitable care for all populations (Foronda et al., 2016; Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013).

The goal of CRT is also to promote the socially just, optimal, and equitable education for all students (Gay, 2002, 2010; Ladson-Billings, 2014; Paris, 2012; Villegas-Lucas, 2002).

Thus, it is clear that cultural humility is well aligned with major tenants of CRT. The point of this paper is not to suggest a replacement of CRT theory, about which a significant literature has been thoughtfully established. Instead, this paper argues that cultural humility can be a helpful addition to CRT theory for a number of reasons. Cultural humility is particularly helpful because it is not primarily a set of competencies that are specific to a classroom or office, but instead it is a way of being within a diverse environment. This makes cultural humility even more useful for schools because it is relevant to all school staff as it provides general outlines for the attitudes of a culturally humble helping professional; teachers, administrators and support staff can all do their jobs informed by the same tenants of cultural humility. Additionally, culturally humble practice has applications at the individual and systemic level, as the tenants apply to one-on-one interactions between people within an organization and for how organizations as a whole (such as schools) serve diverse populations and interact with their communities in general (Tervalon & Murray-Garcia, 1998). In order to propose exactly how cultural humility can enhance CRT, it needs to be known what is currently being encouraged and found effective in teacher preparation for cultural responsiveness.

Inservice Professional Development for Culturally Responsive Teaching

There is a need for empirical studies of professional development programs aimed to increase culturally responsive practice and pedagogy (e.g., Bottiani et al., 2018; Parkhouse et al., 2019; Sleeter, 2012). Theoretically, CRT professional development efforts should be able to help teachers meet the diverse needs of all their students (e.g., Ladson-Billings, 1995b; Gay, 2002; Villegas & Lucas, 2002). Whether through increasing teachers' culturally competent skills or

changing their attitudes towards student diversity, it is expected that training teachers to be more culturally responsive will help close gaps in discipline/special education disproportionality and student achievement between demographic groups (Ladson-Billings, 1995b; Gay, 2002). Yet, few studies of teacher professional development programs have demonstrated these outcomes (Bottiani et al., 2018; Parkhouse et al., 2019; Sleeter, 2012). Even further, few studies of teacher professional development programs to increase cultural responsiveness have met standards of evidenced-based practice that would encourage generalizability for other settings and populations (Bottiani et al., 2018). Aronson and Laughter (2016) reviewed 45 studies of culturally responsive educational practices and concluded there was evidence across all school subjects demonstrating that CRT improves student achievement, engagement in school, and lifelong learning practices. This would suggest that there are culturally responsive instructional strategies that benefit student outcomes. However, the lack of empirically-validated inservice professional development programs to increase teachers' CRT is troubling.

There have been two recent reviews of the literature around professional development for CRT. Parkhouse et al. (2019) reviewed 40 studies of professional development programs for inservice teachers, administrators, and support staff from around the world. Their findings suggested that there is a significant variance in the type of research available evaluating such programs. In fact, no conclusions could be drawn about the comparative effectiveness of the studies because they varied so widely in terms of focus, methodology, and theoretical approaches. Parkhouse et al. recommended that future studies of teacher professional development should build on the small amount of research that has been conducted so that comparisons can be made about the effectiveness of certain training techniques. Additionally, Parkhouse et al. raised questions about the overall approaches to professional development for teachers, suggesting that new

theoretical frameworks may need to incorporate humility, and a specific focus on educational equity and teacher reflection.

Bottiani et al. (2018) published a review of the literature about inservice teacher professional development to increase culturally responsive teacher practice. Their review suggested that only ten studies of professional development for culturally responsive teacher practice were published between 1998 and 2014, and none of these studies met rigorous standards of generalizable, evidenced-based practice. Interestingly, though the Parkhouse et al. (2019) paper and the Bottiani et al. paper review studies about similar constructs during the same time frame, none of the studies included in their reviews overlapped. This would suggest that there needs to be better consistency in methodological review procedures and inclusion criteria so that a full picture of the literature can be established for future researchers and school personnel attempting to make research-based decisions. However, both reviewers arrived at similar conclusions: First, the paucity of research in this area makes it impossible for schools to implement empirically-validated professional development; both studies concluded that more rigorous research needs to be conducted to strengthen theoretical and practical knowledge in the field. Second, both papers recommend the development of consistent operational definitions of culturally responsive practices, suggesting a better understanding of these factors would likely allow for a better assessment of the effects of CRP training for teachers (proximal outcomes) and could yield programs/practices that are better prepared for dissemination and generalizability. Bottiani et al. also recommended the development of psychometrically valid measures to assess professional development regarding CRT and a thorough exploration of the distal outcomes of culturally responsive teacher training. This would include an examination of the effects that professional development should have on student-level (i.e., impacts on discipline or achievement data), and school-level variables (i.e.,

perceptions of school climate). Enhancing understanding of these factors would allow for rigorous experimental designs that could more robustly assess the effectiveness of professional development for culturally responsive practice in schools. This paper proposes that incorporating cultural humility will help researchers address some of the problems with current CRT training research that Bottiani et al. and Parkhouse et al. outlined.

As the necessity of culturally responsive practice within schools has become more widely realized, it is possible that more recent research has produced studies that would meet the guidelines for evidenced-based practice used in the Bottiani et al. (2018) study, which only reviewed studies published through 2014. The Parkhouse et al. (2019) study evaluated articles published through December 1st, 2017, but they did not evaluate the studies for whether they met standards for evidence-based research. Both studies asked important questions that should be followed up with reviews of more recent years of research. Their methodology differed in a few ways. Parkhouse et al. used a larger range and combination of search terms and included articles from any country. This was in contrast to Bottiani et al., who utilized a shorter string of search terms and only included articles from the United States. In addition to their differing search terms, the level of detail described in the methodology was different between both papers. The inclusion criteria for the Bottiani et al. paper was clear and systematic, described in a narrative and summarized in a table. The Parkhouse et al. paper described the inclusion criteria in narrative format, but it was not as specific as the Bottiani et al. criteria, making it open to more subjectivity in replication. Additionally, Bottiani et al. followed a detailed procedure during the elimination phase of the review, which was described more clearly than the procedure used in the Parkhouse study. Thus, Bottiani et al. was much more explicit and systematic with their methodology, making it possible to replicate with greater reliability. Thus, in order to evaluate the use of cultural humility in

teacher professional development practices, this paper will describe an extension of their review from 2014-2018 that may help inform education professionals and researchers who are hoping to improve the cultural responsiveness of teacher practice and pedagogy.

Proposed components of professional development to increase teachers' cultural humility

Using cultural humility to help provide answers to the questions raised by Bottiani et al. (2018) requires operationalizing the components of a professional development program that might increase teachers' cultural humility. A few preservice education programs have begun to analyze the development of cultural humility within teacher candidates (Brown et al., 2016; Lund & Lee, 2015; Tinkler & Tinkler, 2016), and have outlined practices that have preliminary evidence of effectiveness. Additionally, researchers in other fields have discussed best practices needed to encourage cultural humility in helping professionals (e.g., Chang et al., 2012; Fisher-Borne et al., 2015). A synthesis of those five publications, along with guidelines on the general goals and intended outcomes of practicing with cultural humility (Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013) has resulted in a list of four proposed elements of a professional development programs to increase cultural humility in teachers.

1. The first proposed element of cultural humility development is that the intended outcome of the PD is attitudinally transformative. By transformative, we mean that researchers should expect for the professional development program to have robust impacts along multiple levels of outcomes – including teacher self-reported change (beliefs/attitudes/knowledge), observable teacher behavioral change, and improvements in student behavior and/or achievement. Embracing the beliefs of a culturally humble worldview is a prerequisite to incorporating specific CRT skills into teachers' current practices. Learning to live from a culturally humble orientation requires participants to do meaningful introspective work into their own cultural identity and their

corresponding beliefs/actions. Lund and Lee (2015) suggested that participants should “move beyond superficial understandings of diversity toward critical reflections about their own identities and experiences as they are nestled within complex hierarchies of social class, sexual orientation, gender, and racialization” (p. 363).

The goal of transforming teachers’ mindsets is consistent with the highest recommendations for CRT training. For example, Aronsen & Laughter (2016) argued that, “to be effective, [CRT] must be embraced more fully as a guiding ethos for every aspect of the classroom, a highly attuned pedagogy that drives effective teaching” (p. 198). If a teacher fully embraces a culturally humble attitude, the effects of professional development for CRT should be observable in many different aspects of the classroom (including teacher beliefs/knowledge, teacher behavior, and student outcomes). Additionally, collecting comprehensive outcome data from multiple sources allows for quality feedback for teachers consistent with researchers recommendations for quality evaluations (Darling-Hammond, 2014). A recent study found that teachers consider specific, frequent, and evidence-based feedback to be the most useful for their professional growth (Liu, Visone, Mongillo, & Lisi, 2019). Thus, collecting comprehensive data that can be used to provide specific feedback may further transformational CRT training.

Current goals for teacher change in professional development for CRT vary widely. Though changing knowledge/beliefs and developing skills are two common purposes for professional development (Bottiani et al., 2018; Aronson & Laughter, 2016), many studies oversimplify the conceptualization of cultural responsiveness and in turn undermine the extent to which teachers have to fully adopt CRT beliefs in order to be effective in CRT practices (Aronson & Laughter, 2016; Ladson-Billings, 2014). It is important that training to increase cultural humility be focused on authentically challenging participants’ ways of thinking about culture, and thus

imparting meaningful skills of deep reflection and a commitment to more effective interactions with diverse students. We, along with others (Chang et al., 2012, Fisher-Borne et al., 2015) argue that the goal of transformation is a crucial first step in increasing participants' openness to and positive views about cultures different from their own.

2. Another proposed facet of training to increase levels of cultural humility is the presence of explicit teaching about privilege, power, oppression and its effects in public education (e.g., disproportionate discipline practices, etc.; Lund & Lee, 2015; Tinkler & Tinkler, 2016). Partridge et al. (2019) found that previous studies vary on their inclusion of such training, some tending to avoid the tension and defensiveness that can arise from such topics. However, this explicit teaching is necessary to give teachers knowledge and awareness of issues that can inspire critical self-reflection about their own cultural identities (Lund & Lee, 2015). Additionally, cultural humility calls for helping professionals to correct power imbalances where none ought to exist (Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013). Being aware of the power imbalances in education due to cultural differences is the first step to correcting those imbalances (Ladson-Billings, 1995b, 2014). This requires that training explicitly inform participants about the inequities that exist within the US education system, and that trainees become aware of how privileged teacher identities can augment these inequities in the classroom (de Freitas & McAuley, 2008).

3. A proposed third element of professional development to increase cultural humility is providing time for participants to engage in critical self-reflection of their current beliefs about and actions towards people who are different from them, as a commitment to ongoing critical self-reflection is an important component of cultural humility (Foronda et al., 2016; Tervalon & Murray-Garcia, 1998). These continual reflective opportunities should be individual (written

and/or with a mentor) and collective (group discussion) (Fisher-Borne et al., 2015; Lund & Lee, 2015; Ross, 2010). Fisher-Borne et al. (2015) suggests that reflective opportunities should occur at institutional and personal levels in order to bring awareness of systemic and personal contributions to service disparities or clients from diverse backgrounds. Parkhouse et al. (2019) conclude that, “PD providers may want to consider Marshall et al.’s (2012) recommendation to have teachers not only reflect but also ‘unpack contradictions between their professed beliefs, their teaching practices, and their explanations for those teaching practices’” (p. 450). Structured time of reflection that addresses these issues will likely help participants achieve a deeper understanding and personal commitment to practice with cultural humility, as well as provide examples of the types of the reflections that will be most useful to them outside of the training environment.

4. Part of being culturally humble is being aware that one cannot do all the work towards equity by oneself (Waters & Asbill, 2013). Thus, the fourth proposed tenant of training for increased cultural humility is that the training program encourages collaborative action outside the classroom, such as advocacy on behalf of students and school-family-community partnerships. Tervalon & Murray-Garcia (1998) write that outside work with the public and community agencies ensures that the professional has continued awareness of salient issues in the field that can spark self-reflection about the context of diverse populations and increase humility and equity in practice. It also demonstrates an orientation towards social justice, which is implicit and necessary for a culturally humble practitioner (Lund & Lee, 2015; Tervalon & Murray-Garcia, 1998; Waters & Asbill, 2013). As a means of training, developing family and community partnerships may help develop practitioners’ cultural humility. Brown et al., (2016) describe a university course for preservice teachers that is completely focused on increasing cultural humility through developing stronger family partnerships, suggesting that humility can be expanded through an

exposure to family structures and cultures that are different from one's own. Experience with working with communities and families while also undergoing CRT training may help teachers develop the critical consciousness necessary to combat the "savior" complex that can develop in privileged practitioners (e.g., Straubhaar, 2015). Additionally, it has been theorized that altruism is closely related to the development of true humility (Ashton, Lee, & de Vries, 2014). Thus, professional development efforts for increasing cultural humility should encourage educators to advocate for and collaborate with families and communities to work to achieve justice.

Guiding Questions

For the purpose of providing recommendations for how cultural humility can improve teacher preparation for CRT, this paper extends the Bottiani et al. (2018) systematic review to include articles published between 2014-2018. To examine the growing literature related to in-service training for CRT that has been produced in that time frame, this paper will follow the same inclusion criteria and evaluative processes established by Bottiani et al. Additionally, the articles included in this review and in the Bottiani et al. review will be examined for the presence of the four proposed methods for increasing teachers' cultural humility. Thus, this paper will examine if and how cultural humility is being encouraged in the empirical literature around professional development for CRT. Implications and recommendations for future research and training processes will be discussed.

This review has two guiding research questions:

- 1) What empirically valid studies of professional development for CRT have been published between 2014 and 2018 and how are they characterized using the evaluation procedures outlined by Bottiani et al. 2018?

2) What (if any) strategies are being incorporated in professional development programs/practices for CRT that encourage educators' cultural humility?

Method

To answer research question 1, the search, screening and evaluation methods outlined by Bottiani et al. (2018) were followed as closely as possible with one notable exception: the search was limited to articles published between January 1st, 2014 and December 31st, 2018. The inclusion criteria were identical to Bottiani et al. and defined using a PICO framework (Population, Intervention, Comparators, and Outcomes; Schardt, Adams, Owens, Keitz, & Fontelo, 2007). The population of interest was defined as inservice teachers or school administrators who practice in US schools with a general education population. The interventions of focus were those intended to promote teachers' cultural responsiveness and related student outcomes with inservice teachers and/or school administrators. Comparators is used to describe designs that have a control group; Bottiani et al. did not require the interventions to have a control group, though they did require that all included articles report qualitative or quantitative results (as opposed to just describing the professional development program). Finally, the outcomes of interest included both proximal outcomes (e.g., teacher attitudes and CRT practices), and distal outcomes (e.g., disparities in achievement and discipline between students).

Identical to Bottiani et al. (2018), 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”” (Bottiani et al., 2018 p. 369). Additionally, the centralized EBSCOHost search function was used to search Academic Search Complete, ERIC, PsychINFO, and The Sociological Collection. The Sociological Collection was substituted for the Social Sciences Full Text database

used by Bottiani et al. because University Library Services were unable to access Social Sciences Full Text. However, none of the final included articles were indexed in this database, so it is likely that this substitution made little difference in the overall search process.

The search results were reviewed, and articles were systematically eliminated if they did not meet the inclusion criteria specified in the Bottiani et al. (2018) paper (summarized above). On the few occasions when it was unclear how to apply the inclusion criteria, the first author consulted with one other member of the research team who had extensive research experience to reach a consensus about whether the article should be included. This consultation was necessary for four articles at the last stage of elimination; one article of the four was ultimately included in the review.

After a final list was generated, the articles were classified by their study design, sample and setting, definition of CRT, intervention characteristics, staff and/or student outcome measures, and findings. Following the procedure of Bottiani et al. (2018), the quality of the studies was then further evaluated using guidelines set forth for quantitative (Gottfredson et al., 2015) and qualitative (Creswell & Miller, 2000) research. For quantitative studies, the following aspects were examined: 1) *statistical conclusion validity*, which refers to the use of appropriate statistical analysis to relate the intervention to the outcome; 2) *internal validity*, which refers to whether causal inferences can be made about the intervention on the outcome variable; 3) *construct validity*, which refers to the validity of measurement approaches; and 4) *external validity*, which refers to whether it can be concluded about how the findings might generalize to other sample populations. These characteristics were compared with best practice outlined by Gottfredson et al. (2015) and the What Works Clearinghouse (WWC) standards for evidence-based research (What Works Clearinghouse, 2017). For qualitative and mixed-methods studies, the

following aspects of qualitative study validity were examined: *triangulation, disconfirming evidence, researcher reflexivity, member checks, external auditors, peer debriefing, audit trail, prolonged field engagements, thick detailed description* and *particularizability*. These characteristics were compared with best practice outlined by Creswell and Miller (2000) and the APA Publications and Communications Board Working Group on Journal Article Reporting Standards (Levitt, Bamberg, Creswell, Frost, Josselson, & Suarez-Orozco, 2018).

Finally, the list of articles included in this study, as well as the list of articles generated by Bottiani et al. (2018) were analyzed for the inclusion of practices to encourage the development of cultural humility within teachers and school administrators. The described interventions were read closely to determine if: 1) the goals of the intervention were transformative. Specifically, we were looking for articles to explicitly state the importance of teacher mindset, behavior, and student outcomes for successful CRT and distinct steps to change/influence/measure all three outcomes; 2) There was education about education inequity as part of the training program; 3) There was time for critical reflection in an individual and/or group setting; and 4) the program encouraged collaboration with outside/community agencies. The articles were read closely for the presence of these proposed practices, and, if the practice was present, its application was described.

Results

The search from 2014-2018 initially yielded 415 articles. Abstracts were reviewed for the above inclusion criteria (e.g., the research was conducted in the United States, the articles described an intervention to improve CRT), and 60 articles were retained from the original 415. The abstracts for these 60 articles were read again, and 24 were further excluded because the article did not report results or because the intervention was not completed directly with in-service

teachers or administrators (interventions done with preservice teachers or other school support staff were excluded). The remaining 36 articles were read in their entirety to assess for the inclusion criteria. Twenty articles were further excluded after this close read due to one or more of the exclusion criteria that was not previously clear from the abstract, leaving 16 peer-reviewed articles eligible for inclusion from the 5-year period of 2014-2018. The key aims and findings from these articles are described below and summarized in Table 1.2. What follows are the findings from the review which include descriptions of the features of each intervention, the methodological characteristics of each intervention, a description of the impact of each article, and a review of the aspects of cultural humility in each article.

Features of Inservice Interventions from 2014-2018

This section provides an overview of the characteristics of each intervention. The first section, *target for change*, reviews each intervention's targeted population and general procedures. The sections *mode of delivery*, *dosage*, and *implementation* review the method of training delivery, the frequency, intensity and duration of each intervention, and the person responsible for implementing each intervention. The last sections describe whether the intervention was voluntary or mandatory, and whether it was directed towards a specific cultural group or generalizable for all student groups.

Target for change. All but two reviewed articles reported results from different interventions directed at improving CRT with inservice teachers. Two studies (Gregory et al., 2016 and Gregory et al., 2014) reported results from a trial of the My Teaching Partner – Secondary (MTP-S); the first article is from the initial year of the intervention and the second article reports results for the two years following. Three other studies used previously developed intervention or training programs, including Instructional Conversations (Portes, Canche, Boada, & Whatley,

2018), Double Check/Classroom Check Up (Bradshaw, Pas, Bottiani, Reinke, Herman, Rosenberg, 2018), the Culturally Responsive Instruction Observation Protocol (Powell, Cantrell, Malojuvera, & Correll, 2016), and Invitational Education practice (Allen & Fitzgerald, 2017). The other 11 studies evaluated the implementation of an intervention that was of original design.

Ten of the included studies focused on improving both teacher knowledge and skills for working with culturally diverse students. Bradshaw et al. (2018) worked in 12 schools with 158 teachers to reduce disproportionality in office discipline referrals and special education referrals through professional development combined with coaching. McKenney et al. (2017) did consultation with three elementary school teachers to increase their knowledge of and skills in classroom management and cultural responsiveness. Allen and Fitzgerald (2017) conducted a professional learning community with five educators in one elementary school aimed at encouraging reflective practice and broadening perspectives of diversity. Parker et al. (2017) summarized the themes from end-of-semester projects from a university course for inservice teachers about cultural responsiveness. Powell et al. (2016) aimed to increase 27 teachers' use of culturally relevant instruction through professional development and one year of coaching. Brown and Crippen (2016) conducted six Saturday workshops to help high school science teachers collaborate for more culturally responsive lesson planning. Cavanagh et al. (2014) worked with Latino parents and students to design a school-based intervention to help better connect them to the school environment. Sigman et al. (2014) conducted three different weeklong workshops to connect science teachers in rural Alaska with local scientists and native experts. Bequette (2014) described an intervention that invited Ojibwe artists to teach 50 art teachers in the Midwest about how to incorporate indigenous art styles and culture into their classrooms. Dorado et al. (2016) used training

and consultation to implement trauma-informed care to a Californian school district with a large population of trauma-impacted students.

Four studies (Fallon et al., 2018, Gregory et al., 2014; Gregory et al., 2016; Portes et al., 2018) primarily focused on improving teacher skills in the classroom. Fallon et al. (2018) aimed to increase three middle school teachers' use of culturally and contextually relevant strategies in the classroom. Both Gregory et al. (2014, 2016) papers describe the effects from different years of the same intervention aimed at reducing disproportionality in office discipline referrals for African American students. Portes et al. (2018) trained teachers of English Language Learner students to use the Instructional Conversations intervention in order to help improve achievement for all students (including ELL students). Two studies (Ezzani, 2014; McKoy et al., 2017) focused on improving teacher knowledge/attitudes only. Ezzani (2014) targeted improving discourse around diversity issues among adults in one Texas school district. McKoy et al. (2017) trained 18 music teachers in the Southeast in a weeklong workshop to incorporate diverse styles of music in their classrooms in a culturally responsive manner.

Mode of delivery. The two most common approaches to training were workshop-type professional development (7 studies) and a workshop-type professional development accompanied with ongoing coaching and/or consultation (7 studies). The professional development activities varied from study to study and included didactics, lesson plan development, group discussion, and individual written reflection. Parker et al. (2017) described the use of an assignment in a university course to help teachers process their behavioral expectations for one child, and then to design a culturally appropriate intervention for that child. Three studies helped participants develop culturally responsive lesson plans by bringing in experts from the community or including training content to help participants broaden their understanding of local culture (Bequette et al.,

2014; McKoy et al., 2017; Sigman et al., 2014). Ongoing coaching and consultation relied on feedback from observations and/or video clips to help teachers understand their strengths and areas for growth in CRT (Bradshaw et al., 2018; Fallon et al., 2018; Gregory et al., 2014, 2016; Portes et al., 2018). McKenney et al. (2017) was the only study that did not incorporate specific workshop-type professional development and only used the consultation process to improve teachers' classroom management and cultural responsiveness. The three teachers described in the study all received different types of one-on-one professional instruction depending on their specific strengths and weaknesses. One study (Cavannagh et al., 2014) did not specify the type of professional development the teachers experienced as part of the intervention; it is described as "training in restorative justice," and did not describe the mode of delivery, dosage, those responsible for the implementation, or whether participation was voluntary or specific.

Dosage. The interventions varied widely in their frequency (e.g., number of sessions per week/month), intensity (e.g., how long each session met), and duration (e.g., how long the total intervention time was). Most articles reported at least one of these measures, but few studies reported all three of them. Three studies (Bradshaw et al., 2018; Cavanagh et al., 2014; Parker et al., 2017) did not report any dosage information, though Bradshaw et al. (2018) did refer to another article that describes the intervention in depth. The Parker et al. (2017) study reported findings from four semesters of a university course project but did not detail the time students spent on the project or the length and frequency of classes. Three interventions (Dorado et al., 2016; Gregory et al., 2016; Portes et al., 2018) had a fairly extensive duration (over 1 year), and three interventions (Ezzani et al., 2014; McKoy et al., 2017, & Sigman et al., 2014) were more limited (5 days or less). Ezzani et al. (2014) had the least dosage of the described interventions, with professional development sessions for teachers occurring 1 hour per semester for 3 years. The

remaining interventions reported durations between 6 days and 1 year of biweekly coaching.

Overall, there was enough dosage information to note the considerable variability between intervention efforts.

Implementation. All but two articles (Brown & Crippen, 2016; Cavanagh et al., 2014) noted who was responsible for implementing the intervention. Four studies (Bradshaw et al., 2018; Fallon et al., 2018; Parker et al., 2017; Powell et al., 2016) noted that the researchers conducted the training, coaching sessions and/or observations. Three studies (Bequette, 2014; Dorado et al., 2016; McKoy et al., 2017) added outside experts to the research team, all of whom completed the training efforts together. Five studies (Gregory et al., 2014, 2016; Sigman et al., 2014; Portes et al., 2018; McKenney et al., 2017) noted that outside experts alone were responsible for implementing the intervention efforts. In the Portes et al. (2017) study, participants tracked their own data during the year of intervention implementation. Finally, two studies noted that the participants were responsible for leading the intervention themselves: in the Ezzani (2014) study, school staff led local trainings after attending a one-day workshop conducted by district administration. In the Allen and Fitzgerald (2017) study, teachers participated in a professional learning community (PLC) used the I-CORT framework to guide themselves through a conversation each week. However, the principal investigator in the Allen and Fitzgerald study was involved in the PLC and was considered the lead facilitator.

Voluntary or mandated participation. Seven studies did not report whether participation in the intervention was mandatory for school staff. Eight studies noted that participation in the intervention was voluntary, and one study (Dorado et al., 2016) indicated that participation for school staff at the schools participating in the intervention was mandatory.

Specific to cultural group or generalizable. Ten studies did not specify a target student population or school subject matter as a focus for the intervention, which suggested that these interventions were generalizable to CRT practices with all student groups. Two studies focused on indigenous populations; one targeted rural Alaskan communities (Sigman et al., 2014), and one targeted incorporating local Ojibwe traditions into art classes (Bequette, 2014). Two studies focused on CRT in specific subjects: Brown and Crippen (2016) focused on high school science and McKoy et al. (2017) focused on music. Cavanagh et al. (2014) was interested in improving the experience for Latino students and their families, but the training offered was not specific to that population group. Similarly, the intervention in the Portes et al. (2017) study was intended to improve instruction for English Language Learners, but the reported benefits were not limited to that group of students.

Methodological Characteristics of Interventions from 2014-2018

After describing the features of each article, they were then sorted for whether they were quantitative, qualitative, or mixed methods, and the validity of each study's methodology was evaluated. The following section describes first the quantitative articles, each of which were analyzed to evaluate their methodological validity and to determine whether they met What Work's Clearinghouse's (WWC's) standards for evidence-based research. The quantitative articles are ordered by methodology: first randomized controlled trials (RCTs) (3), then quasi-experimental studies (1), then single-case designs (SCDs) (2), then program evaluation (1). Next, the results of the analysis of qualitative and mixed methods articles are reported collectively by the eight validity characteristics for qualitative research outlined by Creswell and Miller (2000). We also examined the qualitative and mixed-methods articles for what Bottiani et al. (2018) called

particularizability, which refers to the details of the sample that would enable a reader to make reasonable conclusions about the potential generalizability of the outcomes described in each article.

Quantitative Analysis. Seven studies used quantitative methods to evaluate the effectiveness of their intervention programs. Regardless of the methodology used, each study was evaluated for the four types of validity specified by Gottfredson et al. (2015; *statistical conclusion validity, internal validity, construct validity, and external validity*). Additionally, studies were evaluated for whether they met the criteria for WWC's guidelines for evidence-based research. Four studies (3 RCTs and 1 SCD) met WWC's guidelines for evidence-based research, as they used random assignment, control groups with baseline equivalence or appropriate manipulation of the independent variable, and appropriate statistical measures (WWC, 2017). However, all four of these studies were evaluating interventions that were designed by one or more of the authors of the paper, and some lacked detail of the exact procedures of training and intervention with teachers. These weaknesses threaten Gottfredson et al.'s description of *external validity*, as it makes it hard to assess the reliable replication of interventions with other populations. This suggests that more research is needed overall for even the most rigorous studies included in this review. The findings from the analysis of the quantitative section are summarized in Table 1.3.

The Bradshaw et al. (2018) study used an RCT design that meets all of the qualifications for evidence-based practice specified by WWC, as the groups were randomly assigned and statistically equivalent. The attrition rate was low and likely did not affect outcomes, suggesting that the requirements for *internal validity* were satisfied. Multilevel growth curve modeling was used to control for school- and teacher-level variables, and a two-level Hierarchical linear model was used to assess the difference between coached and non-coached teachers on the outcome

variables. These procedures were appropriate and met the requirements for *statistical conclusion validity*. Acceptable reliability and validity statistics were reported for all the observational and self-report measures, suggesting that the study met requirements for *construct validity*. The Bradshaw et al. study does fall under one Gottfredson exception to *external validity*, which is that the intervention utilized (the Double Check Coaching system) was designed by the authors of the paper. Additionally, the article lacks some procedural detail which would make it hard to replicate, though it does refer to another article about the intervention process that is much more detailed. However, *external validity* is not mentioned by the WWC's evidence-based practice guidelines. So, based on our application of the WWC and Gottfredson et al. (2015) guidelines, it appears that this study meets WWC standards for evidence-based research, but does not fully meet guidelines for *external validity* specified by Gottfredson et al.

Both Gregory et al. (2014, 2016) papers describe the same intervention study, so they will be described together. The Gregory et al. (2014) paper describes findings from the first year of an intervention study, while the Gregory et al. (2016) paper describes findings from the second year of the intervention and the third year follow up on intervention maintenance. The study followed an RCT design with equivalent groups. Attrition was not a significant issue and did not exceed limits specified by WWC guidelines. This suggests that both papers met qualifications for *internal validity*. Statistical methods used in both studies were hierarchical growth curve modeling controlling for teacher and student covariates with office discipline referrals (ODRs) as the outcome measure. Thus, requirements for *statistical conclusion validity* were satisfied. In the 2016 paper, Gregory et al. also tested a mediation model to determine which teaching practices were most predictive of reduced ODRs. This required that findings from the CLASS observation tool were used in the second study, and acceptable interrater reliability was reported, meeting the

standards of *construct validity*. While all of these characteristics meet WWC standards for evidenced based practice, there are questions of *external validity* because this is the first RCT evaluating the MTP-S intervention. Replications of this study by other research teams and with other populations will be necessary to determine the extent of *external validity* of the MTP-S.

The paper by Portes et al. (2017) used an RCT design to evaluate their intervention. However, they failed to provide statistical evidence of baseline equivalence between control and intervention groups, noting that the students in one grade-level's treatment group were statistically significantly more advanced than the control group. Thus, the findings in the Portes et al. study are considered quasi-experimental, and standards for *internal validity* were not met. The statistical methods used (stepwise regression) were appropriate, but there was no mention of adjusting alpha levels for the number of tests conducted, which is a limitation to *statistical conclusion validity*. Additionally, no reliability or validity information was reported about the observational measures or the academic outcome measures, so the requirements for *construct validity* were not met. Finally, the unequal groups and lack of overall intervention description do not allow for a meaningful assessment of *external validity*. Thus, the Portes et al. study did not meet qualifications of evidence-based practice as specified by WWC or Gottfredson et al. (2015).

Two quantitative studies used single case, multiple baseline design (Fallon et al., 2018; McKenney et al., 2017). Both studies had three cases, meeting standards of *internal validity* for the replication of effect specified by WWC (Kratochwill et al., 2010). The McKenney et al. (2017) study had three phases of manipulation to the teacher training conditions, which is below the required minimum of four phases in order to meet WWC standards (Kratochwill et al., 2010), but the Fallon et al. (2018) study had four phases with three data points per phase, indicating that the study met WWC standards with reservations (Kratochwill et al., 2010). To fully meet WWC

evidence standards, each phase would need five data points (Kratochwill et al., 2010). Both studies met standards for *construct validity*, as the observed change was measured systematically overtime by at least two assessors for more than 20% of the time in both studies. Gottfredson et al. (2015) does not outline guidelines for assessing *statistical conclusion validity* in small samples. Regarding *external validity*, the McKenney et al. study described the topics of consultation with the three teachers in detail, but because there was not a standardized protocol it is unclear if their approach would have the same effects in other settings. In the Fallon et al. study, little information is given about the training the teachers were given, which makes it unclear if this intervention could be replicated reliably in other populations. So, it appears that the Fallon et al. study meets WWC's standards for evidence-based practice with reservations, but, as with many of the other articles, there is a lack of detail about the specific training procedures that would allow for replication by another researcher. Thus, the Fallon et al. study does not satisfy Gottfredson et al.'s (2015) standards of *external validity*.

The Dorado et al. (2016) study used simple statistical procedures (paired t-tests of retrospective surveys, student engagement and discipline data) to evaluate the implementation of their trauma-informed care program. Thus, there was no control group, and no reporting of reliability or validity statistics of the measures used for each outcome. While the intervention suggested positive results, the requirements for *internal validity*, *construct validity*, *statistical conclusion validity*, and *external validity* were not met, and thus WWC standards for evidenced-based practice were not met.

Qualitative analysis. Nine studies utilized qualitative or mixed methods to evaluate the success of their intervention program. These studies yielded important information, examples of CRT training and, in some cases, rigorous qualitative methods when examined using Creswell

and Miller's (2000) standards for valid qualitative research. However, WWC does not review qualitative studies, so there was no way to evaluate for WWC evidence standards. The methodological characteristics (*triangulation, disconfirming evidence, researcher reflexivity, member checks, collaborative work, audit trail, external auditors/peer debriefing, prolonged field engagement, thick description, particularizability*) of the included articles are detailed below and summarized in Table 1.4.

Triangulation. Triangulation refers to the use of multiple data sources to search for convergence of information, or themes, in the study (Creswell & Miller, 2000). This was the most common source of methodological validity, as six of the nine qualitative studies demonstrated evidence of triangulation between data sources. As an example, the Powell et al. (2016) study utilized a particularly thorough method of triangulation, as they collected observational data of teachers' classrooms, interview data with teachers, and student academic outcome data. Three different types of data drawn from three different reporters allowed a robust analysis of the effectiveness of their intervention program.

Disconfirming evidence. Searching for disconfirming evidence means that after generating themes researchers look for evidence that opposes those themes in order to strengthen the validity of their findings (Creswell & Miller, 2000). No studies included in the review mentioned the search for disconfirming evidence. This is problematic as ignoring disconfirming evidence threatens the methodological integrity of qualitative research (Levitt et al., 2018). For example, in the Parker et al. (2017) study that evaluated student final projects in a graduate course, projects were only selected for analysis if the student demonstrated a perspective change in his/her written reflections. Thus, 28 projects of students who did not demonstrate a perspective change were left out of the study. This is precarious, as it suggests that potentially meaningful

disconfirming evidence was excluded from the study. If these projects were included, analysis of the unchanged perspectives could have been useful to hypothesize the mechanism of change, as well as provide a thorough challenge to the findings presented as a result of engaging in the project.

Researcher Reflexivity. A key validity procedure in qualitative research occurs when the research team examines their own biases, assumptions and beliefs in relation to the research subject, participants, and findings (Creswell & Miller, 2000). Reflexivity is particularly powerful when it is done systematically (e.g., bracketing, diary logs, journals or field notes; Levitt et al., 2018). Bequette (2014) was the only author to note researcher reflexivity by explicitly stating the research team's expectations of positivist outcomes and the researchers' intentional power sharing with the indigenous community throughout the project. More overt and/or systematic ways of ensuring reflexivity would have strengthened the validity of the articles included in the review.

Member checks. Member checks refers to giving the participants a chance to review the findings and confirm accuracy/representation of the participants' experiences (Creswell & Miller, 2000). Lincoln and Guba (1985) wrote that member checks are the best way to establish credibility of qualitative findings, and the recent Levitt et al. (2018) article on qualitative research reporting standards noted that member checks are an important supplemental support for qualitative analysis. Four of the articles included the use of member checks. Allen and Fitzgerald (2017) gave their participants the raw observation notes as well as copies of their responses to interview questions. Participants had 10 days to review and comment on the accuracy of the material. A different approach to member checks was done by Cavanagh et al. (2014), who invited feedback from participants on the findings and themes determined after data analysis.

Collaborative work. Collaboration with participants throughout the research process helps ensure that participant views are accurately represented in the findings (Creswell & Miller, 2000). Four articles noted the use of collaborative work with participants in their study. One example is in the Allen and Fitzgerald (2017) paper, which described participatory action research and demonstrated several instances of collaboration, including working with the school to develop a new survey of school climate, facilitating the professional learning community and member checks.

Audit Trail. The goal of an audit trail is to examine the process and product of the findings by including individuals outside the project to determine its credibility (Creswell & Miller, 2000). It is mentioned as a key process to support methodological integrity in the journal article reporting standards for qualitative inquiry (Levitt et al., 2018). Only two studies mention the use of an audit trail during their analytic stages. One example is in the Powell et al. (2016) paper; an appendix is included by the authors giving examples of the coding and theme identification process, allowing all readers the chance to determine the validity of their analytic choices.

External auditors/peer debriefing. Peer debriefing refers to regular review of the data by someone who is familiar with the project or the subject being studied (Creswell & Miller, 2000). It adds credibility to the study by spreading the interpretation of data to others, therefore reducing overall bias in the results. Five articles mentioned the use of peer debriefing. Both Sigman et al. (2014), McKoy et al. (2017), and Powell et al. (2016) described team approaches to confirming the accuracy of data analysis. Allen and Fitzgerald (2017) described peer review as an important part of the data analysis, as one principal investigator was responsible for the majority of the coding process. Bequette (2014) utilized input from the Ojibwe members of the research team throughout the data analysis process.

Prolonged field engagement. This validity procedure encourages researchers to spend a substantial amount of time at the research site to understand the full context of the phenomenon being studied (Creswell & Miller, 2000). Though Creswell and Miller do not acknowledge a set duration for prolonged engagement, they suggest that at least 4 months could be used as a potential guideline. Three studies noted prolonged field engagement. The Ezzani (2014) paper notes that the principal investigator was present with the project through three years of implementation. The Bequette (2014) researchers spent one year designing the approach of the intervention and then three years in the implementation phase. The Allen and Fitzgerald (2017) paper noted that while the intervention lasted for seven weeks, the principal investigator had worked in the district for 18 years.

Thick description. Researchers help readers assess the credibility of research through the use of vivid description of the research site and participants (Creswell & Miller, 2000). Only the Bequette (2014) paper provided a thick description of their research process. Details included in the Bequette paper include interactions between the facilitators and researchers, impressions of the participants' experience during the sessions, and values and goals of the Ojibwe traditions the intervention was attempting to impart in participants.

Particularizability. In addition to the nine strategies of validity described by Creswell and Miller, we followed Bottiani et al.'s (2018) procedures and assessed the studies for whether they provided enough information for readers to determine the potential generalizability of their findings. Bottiani generated five dimensions that would provide sufficient information about the sample: population density/geographic region, sample size, school grade level, school sector (i.e., private or public), and teacher and/or principal demographics. Four studies reported information in at least four of the five categories, suggesting enough information for

particularizability. Four studies (Ezzani, 2014; McKoy et al., 2017; Parker et al., 2017; Powell et al., 2016) specified that their sample taught in suburban or rural schools. Allen and Fitzgerald's (2017) study was conducted in an urban school, and the majority of participants in the Sigman et al. (2014) study taught in rural schools. The remaining three studies did not specify the urbanicity of their geographic region. Three studies (Brown & Crippen, 2016; Ezzani, 2014; McKoy et al., 2017) were conducted in the South/Southeast region of the United States, one in the Northeast (Allen & Fitzgerald, 2017), two in the Midwest (Bequette, 2014; Powell et al., 2016), two in the west (Cavanagh et al., 2014; Parker et al., 2017), and one in Alaska (Sigman et al., 2014). Grade levels varied; three studies (Ezzani et al., 2014; McKoy et al., 2014; Sigman et al., 2014) intervened with teachers of all grade levels. Two studies focused on elementary schools (Allen & Fitzgerald, 2017; Powell et al., 2016), and two focused specifically on high schools (Brown & Crippen, 2016; Cavanagh et al., 2014). Two studies did not specify the grade level taught by participants (Bequette, 2014; Parker et al., 2017). All studies that specified the school sector were in public school settings, but six studies did not specify whether teachers were in public or private school settings. Sample sizes varied and ranged from 5-58 participants. In the studies that reported demographics, participants were mostly White and female. However, five studies (Bequette, 2014; Brown & Crippen, 2016; Cavanagh et al., 2014; Ezzani, 2014; Sigman et al., 2014) did not specify the demographics of their teacher/administrator participants. This variability in reporting information about study samples is somewhat concerning, as the context of the intervention likely impacts its effects and their interpretation. In general, reporting more information about samples will help better inform other researchers and practitioners about the potential effectiveness of an intervention.

Overall, the validity of the qualitative articles varied significantly. These results are summarized in Table 1.4. Allen and Fitzgerald (2017) and Bequette (2014) papers used the most validity strategies (both included 7 out of 10 strategies) and Parker et al. (2017) used the fewest (including no strategies mentioned by Creswell & Miller (2000)). *Triangulation* and *peer debriefing* were used most commonly. It is concerning that no studies utilized *disconfirming evidence*, as this kind of information is useful both for the evaluation of the validity of the research and also for theorizing the mechanism of the intervention. It is difficult to evaluate the impacts of qualitative research without the use of standardized validity checks to assess the integrity of the research process (Levitt et al., 2018). While some of the articles included in this review utilized multiple methods to ensure qualitative research validity, others did not. This suggests that there is still more room to grow in establishing rigorous qualitative methods in the field of professional development for CRT.

Impact of Articles

In the following section, we briefly review how each article defined CRT and how the findings were described in order to connect what each article concludes about each intervention with the extent to which it conceptualized its potential influence.

Definitions of CRT. The definitions of CRT varied in focus, but articles can be grouped into similar themes. One group of articles specifically mentioned incorporating students' backgrounds into the classroom and/or curriculum, which would imply that teachers must take time to know their students well enough in order to make these adjustments to their practice (Brown & Crippen, 2016; Ezzani, 2014; Fallon et al., 2018; McKenney et al., 2017; McKoy et al., 2017; Powell et al., 2016; Sigman et al., 2014). However, few of those articles explicitly outlined the importance of teacher relational efforts. One article described CRT as an integrated practice –

meaning that no specific features of CRT can be reliably indicative of a culturally responsive teacher, but it is instead a disposition on part of the teacher that flows into all aspects of practice (Parker et al., 2017). A third group of articles conceptualized CRT through the student outcomes it would inspire, such as a reduction in disproportionality and students feeling welcomed because their “funds of knowledge” are acknowledged in the classroom (Bequette, 2014; Fallon et al., 2018; Gregory et al., 2014, 2016; Portes et al., 2018; Powell et al., 2016). Finally, a few articles conceptualized CRT as caring for students in explicit and specific ways that were foundationally relational (Allen & Fitzgerald, 2017; Bradshaw et al., 2018; Cavanagh et al., 2014; Dorado et al., 2016).

Findings. All articles reported some sort of positive change in *teacher beliefs*, *teacher behaviors* and/or *student outcomes* as a result of their intervention. Two articles (Bradshaw et al., 2018; Powell et al., 2016) reported a positive change in all three of those outcomes. Another five articles (Dorado et al., 2016; Fallon et al., 2018; Gregory et al., 2016; McKenney et al., 2017) reported positive change in *teacher behaviors* and *student outcomes* as a result of their intervention. The rest of the articles reported a positive change in either *student outcomes* (Gregory et al., 2014; Portes et al., 2018), *teacher beliefs* (Allen & Fitzgerald, 2017; Brown & Crippen, 2016; Cavanagh et al., 2014; Ezzani, 2014; McKoy et al., 2017; Parker et al., 2017; Sigman et al., 2014), or *teacher behaviors* (Bequette, 2014). Examples of positive change in *teacher behaviors* included an increase in culturally-responsive instructional strategies (e.g., Powell et al., 2016), higher rates of positive reinforcement (e.g., Bradshaw et al., 2018), and lowered use of exclusionary discipline practices (e.g., Gregory et al., 2016). Examples of positive change in *teacher beliefs* included a greater awareness of their own cultural identities and how culture can impact learning (e.g., Brown & Crippen, 2016), as well as a better understanding of CRT and its

importance (e.g., McKoy et al., 2017). Examples of positive change in *student outcomes* included higher academic achievement (e.g., Portes et al., 2018), lower rates of disruptive behavior (e.g., McKenney et al., 2017), and better emotional adjustment (e.g., Dorado et al., 2016). The Bequette (2014), Cavanagh et al. (2014), and Ezzani (2014) papers reported their positive findings with some hesitancy, suggesting that the interventions were less influential than they were originally designed to be. Nearly all articles reported limitations in regard to their sample, geographic location, method of data analysis, or some combination of those.

Cultural Humility

The 16 interventions studies included in this review as well as the 10 intervention studies included in the Bottiani et al. (2018) paper were reviewed for the proposed components of cultural humility specified above. These components were operationally defined as follows for this review: 1) Transformational goal: if the article made a statement about changing teacher attitudes and behaviors that would also influence student outcomes, and the study had measures for each of those outcomes, the goal of the intervention was considered sufficiently transformative; 2) Explicit equity education: the intervention had to specify that there was teaching of privilege and inequity and its effects on school performance/experience in some part of the training for teachers or school staff; 3) Structured time for reflection: if the intervention specified some amount of reflection with their participants (whether collectively or individually), this component was satisfied; 4) Encouragement of family/community partnerships and advocacy: the intervention had to mention family/community collaboration as either a goal or a byproduct of the intervention. Table 1.5 summarizes the cultural humility findings for each article.

Transformational goal. Of the total 26 articles, many articulated a goal for the intervention that sounded transformational. For example, Ezzani (2014) cited a framework from Terrel

and Lindsay (2009) that suggested, “an inside out approach to obtain knowledge of self, of others, and to recognize their crucial role in student success” (p. 2). However, only four studies (Bequette, 2014; Bradshaw et al., 2018; McKenney et al., 2017; Powell et al., 2016) both articulated a transformational goal for the intervention and measured outcomes to suggest they expected the intervention to affect teacher beliefs/attitudes/knowledge, teacher behaviors, and student academics or behaviors. Five studies (Dorado et al., 2016; Fallon et al., 2018; Gregory et al., 2016; Jones, Caravaca, Cizek, Horner, & Vincent, 2006; Vincent, Swain-Bradway, Tobin, & May, 2011) collected data on the student and teacher level, but did not include measures of both teacher self-reported change and teacher observed behavioral change (1 study reported only self-reported change and 4 studies reported only observable behavioral change). For example, while Fallon et al. (2018) measured changes in teacher and student behavior, they did not assess for changes in teacher attitudes and beliefs as a result of the intervention. Five studies collected and analyzed teacher artifacts from the professional development sessions (such as lesson plans or class projects), but it was unclear how these artifacts might influence their actual behaviors in the classroom. Most commonly, studies utilized only teacher self-reported outcomes of either behavior, attitudes/beliefs, knowledge, or some combination of those to assess the intervention effectiveness. These self-reports were collected through surveys (7 studies) or interviews (8 studies). Two studies only assessed student outcomes (e.g., Gregory et al., 2014; Portes et al., 2018). Overall, the articulated goals of each intervention sounded more transformational than the researchers’ selection of outcome variables suggested. Study designs with a range of outcomes were able to better examine the impact of the interventions compared with designs that only captured one level of outcomes.

Explicit teaching about equity and privilege in education. Three out of twenty-six studies (Bequette, 2014; Bradshaw et al., 2018; Hammerness & Matsko, 2013) explicitly mentioned that teachers were instructed about equity, privilege and their effects in the education system. Five studies (Dorado et al., 2016; Eberly, Joshi, Konzal, and Galen, 2010; Fickel, 2005; McKenney et al., 2017; Powell et al., 2016) described a training experience that might have included a discussion about privilege and equity, but it was not specifically mentioned. For example, Eberly et al. (2010) had a panel of parents from different minority cultures speak to the teachers about their experience with the US education system; it is likely that there were themes of equity and privilege in this discussion, but the study does not describe it in enough detail to be certain of that. The remaining 18 studies did not specifically mention this as part of their training subjects/materials or did not describe any part of the training in enough detail to gauge whether these topics were covered with participating teachers/school staff. The absence of the mention of this kind of training is very concerning, considering that teachers were being prepared for practicing in a space that educational equity was likely a concern.

Structured time for reflection. Nine studies (Bradshaw et al., 2018; Brown & Crippen, 2016; Eberly et al., 2010; Fickel, 2005; Gregory et al., 2016; McAllister & Irvine, 2002; McCormick, Eick, Womack, 2013; Parker et al., 2017; Portes et al., 2018) specifically mentioned providing participating trainees with time for individual or collective reflection. Methods of reflection included individual reflective writing prompts (3 studies), group discussions during meetings (3 studies), reflection on teaching performance with a coach (2 studies), and/or participation in reflective online discussions (1 study). Seven studies (Allen & Fitzgerald, 2017; Bequette, 2014; Fallon et al., 2018; Gregory et al., 2014; Hammerness & Matsko, 2013; McKenney et al., 2017; Powell et al., 2016) described activities that would likely include a reflective

component, but reflection was not specifically mentioned as part of the intervention process. For example, in interventions that had a coaching element (Fallon et al., 2018; Gregory et al., 2014; Hammerness & Matsko, 2013; McKenney et al., 2017), it seems that teachers would certainly reflect on their own practices in response to feedback. Another example is from the Allen and Fitzgerald (2017) and Bequette (2014) studies, because during each intervention there were group discussions, which likely provided time for collective reflection. However, this mechanism was not specifically described. The remaining ten studies did not mention reflection as part of their intervention processes or did not sufficiently describe their intervention procedures enough to infer what was done during the professional development time. The use of reflection in the training process in these studies is encouraging, as critical reflection is possibly the most important regular practice for a culturally humble professional.

Encouragement of family and community partnerships/advocacy. Nine studies (Bequette, 2014; Cavanagh et al., 2014; Dorado et al., 2016; Eberly et al., 2010; Fickel, 2005; Powell et al., 2016; Shriberg et al., 2012; Sigman et al., 2014; Thompson & Byrnes, 2011) specifically mentioned that encouraging family and community partnerships and/or advocacy was part of the goals of the intervention or addressed in some part of the professional development. For example, the Sigman et al. (2014) paper focused on connecting rural Alaskan science teachers with local expert scientists to help inform lesson planning and build future relationships, and the Shriberg et al. (2012) study was focused on helping teachers build better relationships with parents and families. Two studies (McCormick et al., 2013; McKenney et al., 2017) noted outcomes that involved family partnerships, though it was not mentioned as a goal for the intervention. These studies encouraged the interaction between teachers and families/outside agencies to facilitate relationships between teachers and their students' communities, which were designed to

enable teachers to understand the strengths in diverse cultures and be aware of the resources from their students' communities. The remaining 15 studies did not mention family/community partnerships or advocacy or did not describe their intervention methods well enough to assess if it was present in the intervention procedures. Though less than half of the interventions focused on family and/or community partnerships, this was the second most commonly used strategy to promote cultural humility in the reviewed studies. While it is encouraging to find that promoting family and community partnerships is present in some training programs, more exploration of this strategy and its impacts would be helpful in future research.

Discussion

The purpose of this study was to review the recent literature for empirical studies of professional development programs that target CRT, and to examine how and if teacher cultural humility is being encouraged through these training programs. Overall, it does seem that research examining CRT training programs is progressing. However, the field can benefit from growth, specifically in quantitative methods, replication studies, and more instances of rigorous qualitative and quantitative research designs. Regarding cultural humility, the findings from this review suggest that elements to encourage cultural humility could be increased throughout professional development efforts. The significance of these findings and their implications for research and practice are discussed below in more detail.

Systematic Review

Compared with Bottiani et al. (2018), the findings from our review suggested that research evaluating CRT professional development programs has been published much more frequently in the past four years compared with the previous sixteen years. This could suggest that the importance of this research has become more widespread in recent years, resulting in the

publication of more studies. It is interesting that this body of literature has taken so long to develop, as the theoretical importance of CRT was established between 20 and 25 years ago (e.g., Gay, 2000; Ladson-Billings, 1994). Notably, the Parkhouse et al (2019) review found 29 studies conducted in the United States that empirically evaluated CRT training programs between 2000 and 2017 (8 of these were published between 2014 and 2017, 2 of which were included in the current review). This could suggest that Bottiani et al.'s review is not truly representative of the literature, as the Parkhouse et al. paper found 21 articles that were not included in the Bottiani et al. review, though they covered similar topics and the same time period. However, the Parkhouse et al. review did not include any of the same articles as the Bottiani et al. paper nor the majority of articles included in the current review, which suggests limitations in their search procedures as well. In the future, researchers should use a wide range of search terms to ensure that all articles with findings relevant to the research question are examined in the review process. It is also important that researchers continue to examine the full range of effects that professional development for CRT can have on teachers, students, and schools, as culturally-responsive strategies will only become increasingly more important as the US student population becomes more racially and ethnically diverse.

Another promising finding from this review is that four publications met our application of the WWC standards for evidence-based practice. Four studies (three RCTs and one SCD) utilized rigorous quantitative methods that supported their internal validity, construct validity, and statistical validity. Additionally, these studies demonstrated positive findings, including a reduced ODR rate for Black students, a higher rate of positive behavior for students in the classroom, more consistent positive reinforcement, and increased relational practices by teachers. Gregory et al. (2016) demonstrated a maintenance of these effects after one year without

coaching. The training programs in each of these studies utilized ongoing coaching of teachers as their primary method of professional development, suggesting that it is an effective training method. However, as there were no other equally rigorous studies that utilized a different method of training, it is not possible to compare coaching with other types of professional development, such as workshops, professional learning communities, and graduate coursework. More research that is rigorous in these ways would greatly benefit the field so that schools and LEAs can discern effective CRT training programs. Regarding external validity, the four studies that met WWC standards for evidence-based practice were all authored by the intervention designer; it is important that these studies be replicated by researchers who are not incentivized to report positive outcomes and with populations that are different from those in the original studies.

With the exception of the four articles discussed above, this review found that the recent literature evaluating CRT teacher training has similar strengths and weaknesses to those identified by Bottiani et al. (2018). One notable difference is that a greater proportion of articles were quantitative, with seven out of 16 articles in the current review being quantitative, compared with two out of 10 in the Bottiani et al. paper. Also, 10 out of 16 articles in the current review focused on improving teachers knowledge and skills instead of just one or the other, whereas in the Bottiani et al. review only four out of 10 studies focused on improving both teacher knowledge and skills related to CRT. The qualitative and mixed method studies included in the current review also shared similar strengths to those identified in the Bottiani et al. paper. Specifically, the researchers incorporated various strategies to ensure the validity of the research, some researchers using quite a few different methods. This review also confirmed what Bottiani et al. (2018) and Parkhouse et al. (2019) concluded, that there is still room for progress in the qualitative and quantitative research methodology for studies of PD for CRT. For the majority of the reported

articles, the problems identified by Bottiani et al. (such as inconsistent definitions of CRT, inconsistent use of proximal and distal outcome measures, lack of/poor construct validity, and underreporting of training procedures) are still present in recent research.

Bottiani et al. (2018) identified a few directions for future research that these articles did and did not address. First, Bottiani recommended that multidimensional measures be created and validated to accurately measure teachers' cultural responsiveness; studies used a wide variety of measures of cultural responsiveness, but very few were instruments that had been previously examined for structural integrity, reliability, and/or construct validity. Those measures that did meet these requirements were all self-report. Only one study (Bradshaw et al., 2018) used self-report measures that controlled for social desirability bias. There is clearly still room to develop valid and reliable multidimensional measures of CRT that come from multiple reporters. Bottiani et al. also recommended that distal outcomes be clarified, so that the effects of CRT training could be accurately assessed and compared across studies. The studies that examined student outcomes used either academics, observations of student behavior, or office discipline referrals. However, those that utilized observational measures of student behavior defined problematic behavior differently, and there were not enough studies examining academics to make comparisons across age, grade, subject matter, or demographic groups.

Regarding specific intervention techniques, nearly half the interventions reviewed in our study utilized ongoing coaching and/or consultation in order to help teachers bridge the gap between CRT workshop training and implementation in the classroom. This is a higher percentage than the 2 out of 10 studies that used ongoing coaching in the Bottiani et al. (2018) review. This could suggest that researchers are more interested in effecting long-term teacher changes in addition to the growth in knowledge that is a likely outcome of workshops that do not provide this

additional support. A higher percentage of studies in this review reported the demographics of their participants. Bottiani et al. suggested that teacher demographics should be considered in the intervention because it is possible that some teachers need differing training techniques depending on their background and life experiences. No studies explicitly addressed the differing participant backgrounds in their study designs, although McKenney et al. (2017) did assess their participants' initial knowledge and skills and then adjust the practices of consultation to address weaknesses. As teachers are trained to consider their students backgrounds more thoroughly in the classroom, trainers should similarly consider the backgrounds of teachers in training programs. There is still room for expansion in the literature in this area as well.

Cultural Humility

Examining the studies included in this review and the studies included in Bottiani et al.'s (2018) review for practices that could increase cultural humility has made an important contribution to the literature, given that, to our knowledge, no other systematic review has done so. In general, there is a lack of research about the training procedures utilized to improve any kind of educators' CRT – possibly due to the problems with inconsistency in intervention designs that prevent comparisons of the effectiveness of such strategies as was identified by Bottiani et al. and the current study.

The four proposed elements of training to increase cultural humility are theoretical, as no studies of interventions to improve cultural humility with inservice teachers have been conducted. However, this portion of the paper is intended to provide direction to future researchers who may want to analyze the effect of training on teachers' cultural humility. One concerning finding is that very few studies have measured multidimensional transformation in teacher practice. Though some studies' definitions of CRT sound transformational, there was a lack of

measurement of either teacher self-reported change, observable teacher change, or student outcomes. This could suggest that researchers did not expect their training programs to influence all three outcomes, or that researchers misunderstood the potential effects of CRT training. To assess teacher growth, measuring both self-reported and observational changes is necessary to adequately evaluate the effectiveness of inservice interventions for CRT. Leaving out the measurement of self-report mindset/attitudinal changes risks misrepresenting the mechanism of intervention or underestimating the importance of teacher mindset to influence teacher behavior. Ignoring teachers' potential attitude changes in response to an intervention decreases the importance of helping teachers develop an ongoing commitment to practicing with cultural humility (Lund & Lee, 2015), and risks reducing conceptualizations of CRT to superficial, behavioral changes (Aronson & Laughter, 2016; Ladson-Billings, 2014). At the same time, excluding an observational measure of teachers' behavioral changes risks overestimating the effectiveness of an intervention for actual classroom change. The most common evaluation of intervention effectiveness was teacher self-report. This is problematic, as self-reports are prone to social acceptability bias (e.g., Stanhope, Solomon, Pernell-Arnold, Sands, & Bourjolly, 2005), and previous research on professional development has suggested that training may change self-reported mindsets but have little to no influence on actual behavior (Joyce & Showers, 2002).

In examining the second proposed component of cultural humility, we found that few studies are addressing equity concerns that are specific to education, nor are they encouraging teachers to consider their role in these equity concerns. There have been several assertions from researchers that teachers (especially those with more privileged identities) must examine their own role in oppression in order to effectively serve students with more equity (e.g., Matais, 2013). Dismantling unjust power structures is a critical part of practicing with cultural humility

and thus explicit training to help teachers do that in their classrooms is necessary in training for CRT. It is unclear why this element is left out of so many trainings that are supposed to encourage equity among students. It is possible, as Ladson-Billings (2014) points out, even as educators are trying to implement culturally relevant strategies, many “seem stuck in very limited and superficial notions of culture” (p. 77). So, it is perhaps a limited understanding of these ideas prevents them from being incorporated into a training curriculum. Ladson-Billings (2014) also writes, “Even when people have demonstrated a more expansive knowledge of culture, few have taken up the sociopolitical dimensions of the work, instead dulling its critical edge or omitting it altogether” (p. 77). Thus, perhaps the topic is avoided because it is either too complicated or controversial. However difficult as it may be, expanding educators’ knowledge of privilege, equity and its impacts on the education system is necessary to inspire change in communities that are most disadvantaged by the current political systems of power. One example for how to do this is described in Hammerness and Matsko (2013), who designed their intervention to train teachers who were informed of the local context of their schools. They hired coaches who could adequately articulate the socio-political context of the schools and districts they were teaching in. It then followed that trainees gained perspective about the local context of their schools, which helped them be more effective in CRT efforts. In this example, the researchers started with an awareness of the need for sociopolitical consciousness on behalf of the interveners in order to achieve that outcome with their participants. Future researchers should consider this example and others as they plan to implement training around equity and privilege with education professionals.

Of the proposed methods to encourage cultural humility, reflection was most commonly utilized among included studies, with 16 studies either specifically mentioning reflection or

describing a scenario that very likely had reflective components. The forms of reflection varied, with some studies giving time for individual reflection and some giving time for collective reflection in groups. Though it has been recommended that training programs provide reflective opportunities individually and collectively (e.g., Fisher-Borne et al., 2015), only three studies (Brown & Crippen, 2016; Eberly et al., 2010; McCormick, Eick & Womack, 2013) mentioned using both mechanisms. Of these three studies, only Brown and Crippen suggested that reflection was a key mechanism to help increase teachers' implementation of culturally responsive practices in the classroom. A few studies have reported preliminary findings that suggest highly-reflective activities improve preservice teachers' cultural humility (Brown et al., 2016; Ross et al., 2010; Tinkler & Tinkler, 2016). It is likely that any amount of time for reflection encourages self-awareness skills for teachers about their classroom practices, but it would be better for future studies to examine the impact of each method on teacher's levels of cultural humility. More rigorous research designs that explicitly evaluate the impact of reflective activities during CRT training will be beneficial contributions to the literature.

The final proposed component of training for increasing teacher cultural humility is the encouragement of school-family-community partnerships and advocacy. Though it was mentioned as an intended outcome or element of the training process, very few studies actually evaluated partnerships or advocacy efforts as a sign of intervention effectiveness. So, though nine studies mentioned this as part of their training efforts, imparting attitudes and actions consistent with this component was clearly not viewed as an essential part of teachers' CRT. It is clear from the literature that practitioners with cultural humility should engage in activities that build up and draw strength from the community in which they work (Tervalon & Murray-Garcia, 1998; Water & Asbill, 2013). Thus, future studies of practices to encourage cultural humility, as well as

studies that investigate the best ways to encourage CRT in general, should also examine how school-family-community partnerships and advocacy impact teachers' cultural responsiveness.

Limitations

An important limitation to note is that the search terms utilized by Bottiani et al. (2018) resulted in different results from other very similar reviews of the same time frame of research (e.g., Parkhouse et al., 2019). This suggests that future reviews of the literature should include a broader scope of search terms that would be more inclusive of a larger range of articles. Three articles included in Parkhouse et al.'s (2019) review were published during the time frame of this review but yet did not populate as a result of Bottiani et al.'s search terms. Though the purposes of the Bottiani et al. review and the Parkhouse et al. review are slightly different, it is troubling that two similar articles could yield such different results. Thus, the limitation of specific search terms and/or inclusion criteria of this review must be acknowledged. Due to human subjectivity in the review process, it is also possible that some part of our application of Bottiani' et al.'s search methodology is incorrect in some way, though we did our best to follow the procedure outlined in their article.

Another limitation is that research on training for cultural humility is in a nascent stage, as there are no studies that suggest explicit training for cultural humility makes any difference in teacher performance. We believe that there is strong theoretical evidence to suggest that the proposed components of cultural humility are important for consideration in teacher professional development. However, future studies will need to explore empirically whether a focus on enhancing teachers' cultural humility has an influence on the effectiveness of professional development programs. Additionally, future studies will need to determine how translating cultural humility to education from other helping disciplines impacts the characteristics and measurement of the

construct. Finally, any study evaluating the implementation of some kind of cultural humility-targeted intervention will need to identify what school readiness factors need to be in place in order to make such an intervention successful. A number of factors, including staff readiness, training, capacity for coaching and consultation, and administrative support should be carefully considered when bringing any innovation into an existing school culture (Fixsen, Blase, Naoom, & Wallace, 2009).

Conclusion

While the integration of cultural humility into the theoretical education literature is promising, the current state of research investigating the impact of various methods of professional development for teacher cultural responsiveness is progressing at a slow rate – though it is improving. Similar to the findings from Bottiani et al. (2018), this review indicates there is still a great need for rigorous, controlled research designs that do more than generalize teachers' impressions immediately after the training program has concluded. This review identified several randomly controlled, quasi-experimental, and single-case design studies that examined outcomes on multiple levels and from multiple sources. However, because culturally responsive teaching has been a term in the literature for over twenty years, and because the necessity of CRT has become especially clear in the US, the literature is in desperate need of more research-based training for teachers to improve their CRT in the classroom. It is possible that encouraging cultural humility may be an effective way to impact teachers' classroom practice, and we recommend that researchers explore this variable more in future work.

References

- *Allen, E. J., & FitzGerald, A. M. (2017). Cultural Care and Inviting Practices: Teacher Perspectives on the Influence of Care and Equity in an Urban Elementary School. *Journal of Invitational Theory and Practice*, 23, 5–26.
- Aronson, B. & Laughter, J. (2016). The theory and practice of culturally relevant education: A synthesis of research across content areas. *Review of Educational Research*, 86(1), 163-206. doi: 10.3102/0034654315582066
- Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review*, 18(2), 139-152. doi: 10.1177/1088868314523838.
- *Bequette, J. W. (2014). Culture-Based Arts Education That Teaches against the Grain: A Model for Place-Specific Material Culture Studies. *Studies in Art Education: A Journal of Issues and Research in Art Education*, 55(3), 214–226. doi: 10.1080/00393541.2014.11518931
- Bottiani, J. H., Larson, K. E., Debnam, K. J., Bischoff, C. M., & Bradshaw, C. P. (2018). Promoting educators' use of culturally responsive practices: A systematic review of inservice interventions. *Journal of Teacher Education*, 69(4), 367-385. doi: 10.1177/0022487117722553
- *Bradshaw, C. P., Pas, E. T., Bottiani, J. H., Debnam, K. J., Reinke, W. M., Herman, K. C., & Rosenberg, M. S. (2018). Promoting Cultural Responsivity and Student Engagement through Double Check Coaching of Classroom Teachers: An Efficacy Study. *School Psychology Review*, 47(2), 118–134. <https://doi.org/10.17105/SPR-2017-0119.V47-2>

- *Brown, J. C., & Crippen, K. J. (2016). Designing for Culturally Responsive Science Education through Professional Development. *International Journal of Science Education*, 38(3), 470–492. <https://doi.org/10.1080/09500693.2015.1136756>.
- Brown, E. L., Vesely, C. K., & Dallman, L. (2016). Unpacking biases: Developing cultural humility in early childhood and elementary teacher candidates. *Teacher Educators' Journal*, 9, 75-96.
- *Cavanagh, T., Vigil, P., & Garcia, E. (2014). A Story Legitimizing the Voices of Latino/Hispanic Students and Their Parents: Creating a Restorative Justice Response to Wrongdoing and Conflict in Schools. *Equity & Excellence in Education*, 47(4), 565–579.
doi: 10.1080/10665684.2014.958966
- Chang, E., Simon, M., & Dong, X. (2012). Integrating cultural humility into health care professional education and training. *Advances in Health Sciences Education: Theory and Practice*, 17(2), 269–278. <https://doi.org/10.1007/s10459-010-9264-1>
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, 39(3), 124-130. doi:10.1207/s15430421tip3903_2
- Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4–13.
- de Freitas, E. & McAuley, A. (2008). Teaching for diversity by troubling whiteness: Strategies for classrooms in isolated white communities. *Race Ethnicity and Education*, 11(4), 429-442. doi: 10.1080/13613320802479018
- *Dorado, J. S., Martinez, M., McArthur, L. E., & Leibovitz, T. (2016). Healthy Environments and Response to Trauma in Schools (HEARTS): A whole-school, multi-level, prevention

- and intervention program for creating trauma-informed, safe and supportive schools. *School Mental Health: A Multidisciplinary Research and Practice Journal*, 8(1), 163–176. <https://doi-org.ezproxy.gsu.edu/10.1007/s12310-016-9177-0>
- Eberly, J. L., Joshi, A., Konzal, J., & Galen, H. (2010). Crossing cultures: Considering ethnotheory in teacher thinking and practices. *Multicultural Education*, 18, 25-32.
- *Ezzani, M. (2014). Ethical Leadership: A Texas School District's Efforts toward Cultural Proficiency. *Values and Ethics in Educational Administration*, 11(1).
- *Fallon, L. M., Cathcart, S. C., DeFouw, E. R., O'Keeffe, B. V., & Sugai, G. (2018). Promoting Teachers' Implementation of Culturally and Contextually Relevant Class-Wide Behavior Plans. *Psychology in the Schools*, 55(3), 278–294. <https://doi.org/10.1002/pits.22107>
- 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
- Fisher-Borne, M., Montana Cain, J., & Martin, S. L. (2015). From mastery to accountability: cultural humility as an alternative to cultural competence. *Social Work Education*, 34(2), 165–181. <https://doi.org/10.1080/02615479.2014.977244>
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531-540. doi:
10.1177/1049731509335549
- Foronda, C., Baptiste, D., Reinholdt, M. M., & Ousman, K. (2016). Cultural humility: A concept analysis. *Journal of Transcultural Nursing*, 27(3), 210-217. doi:
10.1177/1043659615592677

- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106-116. <https://doi.org/10.1177/0022487102053002003>
- Gay, G. (2010). *Culturally responsive teaching* (2nd ed.). New York, NY: Teachers College Press.
- 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
- *Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. (2014). Eliminating the Racial Disparity in Classroom Exclusionary Discipline. *Journal of Applied Research on Children*, 5(2).
- *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. doi: 10.17105/SPR45-2.171-191.
- Guo, Y., Arthur, N., & Lund, D. (2009). Intercultural inquiry with pre-service teachers. *Intercultural Education*, 20(6), 565-577. doi: 10.1080/14675980903448619
- 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.
- Hook, J. N., Davis, D. E., Owen, J., DeBlaere, C. (2017). *Cultural humility: Engaging diverse identities in therapy*. Washington, DC: American Psychological Association.

- Hook, J. N., Davis, D. E., Owen, J., Worthington, E. L. J., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology, 60*(3), 353–366. <https://doi.org/10.1037/a0032595>.
- Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- 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.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0 a.k.a. the remix. *Harvard Educational Review, 84*(1), 74-84. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>
- Ladson-Billings, G. (1995a). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice, 34*(3), 159-165.
- Ladson-Billings, G. (1995b). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465-491.
- Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children*. San Francisco, CA: Jossey-Bass.
- Levitt, H. M., Bamberg, M., Creswell, J. W., Frost, D. M., Josselson, R. & Suarez-Orozco, C. (2018). Journal article reporting standards for qualitative primary, qualitative meta analytic, and mixed methods research in psychology: The APA publications and

communications board task force report. *American Psychologist*, 73(1), 26-46.

<http://dx.doi.org/10.1037/amp0000151>.

Liu, Y., Visone, J., Mongillo, M. B., & Lisi, P. (2019). What matters to teachers if evaluation is meant to help them improve? *Studies in Educational Evaluation*, 61, 41-54.

<https://doi.org/10.1016/j.stueduc.2019.01.006>.

Matais, C. E. (2013). Check yo'self before you wreck yo'self and our kids: Counterstories from culturally responsive white teachers?...to culturally responsive white teachers! *Interdisciplinary Journal of Teaching and Learning*, 3(2), 68-81. doi:

<https://doi.org/10.1177/0042085913519338>.

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.

*McKenney, E. L. W., Mann, K. A., Brown, D. L., & Jewell, J. D. (2017). Addressing Cultural Responsiveness in Consultation: An Empirical Demonstration. *Journal of Educational & Psychological Consultation*, 27(3), 289-316.

<https://doi.org/10.1080/10474412.2017.1287575>.

*McKoy, C. L., MacLeod, R. B., Walter, J. S., & Nolker, D. B. (2017). The Impact of an In-Service Workshop on Cooperating Teachers' Perceptions of Culturally Responsive Teaching. *Journal of Music Teacher Education*, 26(2), 50-63.

- Mosher, D. K., Hook, J. N., Captari, L. E., Davis, D. E., DeBlaere, C., & Owen, J. (2017). Cultural humility: A therapeutic framework for engaging diverse clients. *Practice Innovations*, 2(4), 221-233. <http://dx.doi.org/10.1037/pri0000055>.
- Musu-Gillette, L., de Brey, C., McFarland, J., Hussar, W., Sonnenberg, W., & Wilkinson-Flicker, S. (2017). *Status and trends in the education of racial and ethnic groups*. (NCES 2017-051). Retrieved from: <https://nces.ed.gov/pubs2017/2017051.pdf>
- Owen, J. J., Tao, K., Leach, M. M., & Rodolfa, E. (2011). Clients' perceptions of their psychotherapists' multicultural orientation. *Psychotherapy*, 48(3), 274–282. <https://doi.org/10.1037/a0022065>
- *Parker, F., Bartell, T., & Novak, J. D. (2017). Affordances of the Cultural Inquiry Process in Building Secondary Mathematics Teachers' Capacity for Cultural Responsiveness. *North American Chapter of the International Group for the Psychology of Mathematics Education*, 39, 399-406.
- Parkhouse, H., Lu, C. Y., Massaro, V. R. (2019). Multicultural education professional development: A review of the literature. *Review of Educational Research*, 89(3), 416-458. doi: 10.3102/0034654319840359
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93-97. <https://doi.org/10.3102%2F0013189X12441244>
- *Portes, P. R., González Canché, M., Boada, D., & Whatley, M. E. (2018). Early Evaluation Findings from the Instructional Conversation Study: Culturally Responsive Teaching Outcomes for Diverse Learners in Elementary School. *American Educational Research Journal*, 55(3), 488–531. <https://doi.org/10.3102/0002831217741089>.

- *Powell, R., Cantrell, S. C., Malo-Juvera, V., & Correll, P. (2016). Operationalizing Culturally Responsive Instruction: Preliminary Findings of CRIOP Research. *Teachers College Record*, 118(1), 2-46.
- Ross, L. (2010). Notes from the field: Learning cultural humility through critical incidents and central challenges in community-based participatory research. *Journal of Community Practice*, 18, 315-335. doi:10.1080/10705422.2010.490161
- 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
- Skiba, R. J., Horner, R. H., Chung, C., Rausch, M. K., May, S. L. & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, 40(1), 85-107. <http://psycnet.apa.org/record/2011-07091-006>
- 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.
- *Sigman, M., Dublin, R., Anderson, A., Deans, N., Warburton, J., Matsumoto, G. I., ... Harcharek, J. (2014). Using Large Marine Ecosystems and Cultural Responsiveness as the Context for Professional Development of Teachers and Scientists in Ocean Sciences. *Journal of Geoscience Education*, 62(1), 25-40. <https://doi.org/10.5408/12-403.1>.
- 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

- Sleeter, C. E. (2012). Confronting the marginalization of culturally responsive pedagogy. *Urban Education, 47*(3), 562-584. doi: 10.1177/0042085911431472
- Stanhope, V., Solomon, P., Pernell-Arnold, A., Sands, R. G., & Bourjolly, J. N. (2005). Evaluating cultural competence among behavioral health professionals. *Psychiatric Rehabilitation Journal, 28*, 225-233. <http://dx.doi.org/10.2975/28.2005.225.233>.
- Straubhaar, R. (2015). The stark reality of the 'White Saviour' complex and the need for critical consciousness: A document analysis of the early journals of a Freirean educator. *Compare: A Journal of Comparative and International Education, 45*(3), 381-400. doi: 10.1080/03057925.2013.876306.
- Sullivan, A., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*(4), 475-494. <https://doi.org/10.1177/001440291307900406>
- Tervalon, M. & Murray-Garcia, J. (1998). Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education. *Journal of Health Care for the Poor and Underserved, 9*(2), 117-125. doi: 10.1353/hpu.2010.0233
- Thompson, J., & Byrnes, D. (2011). A more diverse circle of friends. *Multicultural Perspectives, 13*, 93-99. doi:10.1080/15210960.2011.571552
- U.S. Department of Education. (2016). *The state of racial diversity in the educator workforce*. Retrieved from: <https://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversity-workforce.pdf>
- Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education, 53*(1), 20-32.

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

Waters, A. & Asbill, L. (2013, August). *Reflections on cultural humility*. [Children youth and families newsletter]. Retrieved from: <https://www.apa.org/pi/families/resources/newsletter/2013/08/cultural-humility.aspx>.

What Works Clearinghouse (2017). *What Works Clearinghouse: Standards Handbook Version 4.0*. Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf.

*Indicates articles included in the systematic review

Table 1.1

Attributes of Cultural Humility Integrated with Culturally Responsive Teaching

Attributes of Cultural Humility (Foronda et al., 2016)	Culturally Responsive Practice (Gay, 2002, 2010)	Culturally Responsive Pedagogy (Ladson-Billings, 1994, 1995a, 1995b, 2014)
<ul style="list-style-type: none"> · Self-Awareness · Commitment to ongoing critical reflection 	<ul style="list-style-type: none"> · Teacher understands why culture and difference are essential to consider 	<ul style="list-style-type: none"> · Teacher builds ongoing socio-political consciousness
<ul style="list-style-type: none"> · Egoless · Openness 	<ul style="list-style-type: none"> · Teacher replaces deficit perspectives of student cultures · Teacher engages in cultural knowledge building 	<ul style="list-style-type: none"> · Commits to life-long learning about culture and what works best for students
<ul style="list-style-type: none"> · Supportive interaction 	<ul style="list-style-type: none"> · Teacher sets high expectations for students · Teachers commit to educate whole child 	<ul style="list-style-type: none"> · Teachers help students succeed within oppressive school systems
<ul style="list-style-type: none"> · Mutual empowerment · Dismantle power structures in the relationship 	<ul style="list-style-type: none"> · Teacher incorporates student strengths · Teachers and students critically question authority of scholarly truth 	<ul style="list-style-type: none"> · Teachers help students understand their own culture · Teacher helps students develop socio-political consciousness

Table 1.2

Sixteen Studies of the Effects of Inservice Interventions to Promote Culturally Responsive Practices.

Intervention Overview	Study Design	Sample and Setting	CRT Definition	Staff/school CRT outcome measure	Student Outcome Measure	Findings
Allen & Fitzgerald, 2017: 7, once-weekly professional learning community meetings with 5 educators from one school to help participants understand the perspectives of effective teachers, build capacity involving equitable practices, form positive relationships, develop greater cultural competence, and encourage reflective practices around equity and social justice.	Qualitative participatory action research	5 educator participants (teachers and mental health professionals) from one Western Pennsylvania Spanish magnet school – 1 man, 4 women, 3 white, 2 African American.	The combination of Invitational Education (IE) practices (defined as purposefully encouraging people to realize their unlimited potential in all aspects of life) and culture of care practice (displaying one’s interest in another person’s well-being while simultaneously recognizing culture as a significant part of that person’s identity).	Pre and Post interviews with participants, participant observations, artifact examination	Participants were asked about student achievement in their interviews	Participant post interviews indicated a more positive, peaceful outlook on the climate of the school and their job as teachers. Teachers reportedly engaged in more positive interactions with students. had a better understanding of their own race and culture and how it might influence their teaching practice. These practices reportedly better supported student learning and achievement.
Bequette, 2014: Week-long workshops plus 3-4 inservice trainings throughout the school year to help art educators be more competent towards indigenous Ojibwe students.	Qualitative	50 teachers in three cohorts	Art education that is framed by a critical pedagogy creates opportunities for students to engage in thinking and artmaking that consider vital questions about nature, place, culture, and ecology.	Survey data, observations of teaching sessions and teachers in the training program. Analysis of teacher lesson plans.	Collected achievement and attendance data but did not use it.	About 25% of participants excelled at designing and teaching culturally responsive lesson plans and 25% performed “perfunctory” at best.

<p>Bradshaw et al., 2018: voluntary participation in 5 60-minute workshops and follow-up coaching using the Classroom Check Up model. Training was focused on lowering overrepresentation of students of color in special education and office discipline referrals through improving teachers' classroom management practices.</p>	<p>Quantitative RCT</p>	<p>158 predominantly White, female teachers (100 experimental, 58 control) from 12 schools (6 middle and 6 elementary) in Maryland. Over half the students were Black or Hispanic, and 57% received free and reduced price meals.</p>	<p>5 core domains of culturally responsive practice, a.k.a. CARES domains: Connection to the curriculum, Effective communication, Authentic relationships, Reflective thinking, and Sensitivity to culture.</p>	<p>Observations in classrooms using the ASSIST global ratings instrument. Self-report scales: Multicultural Efficacy Scale; Culturally Responsive Teaching Self-Efficacy Scale; Classroom Behavior Management Self-Efficacy; Work-Related Stress Scale; Social desirability bias scale</p>	<p>Numbers of office discipline referrals as well as behavioral observations in classrooms using ASSIST global ratings</p>	<p>Authors concluded Double Check and CCU coaching provided "significant added value" to teachers' development of CRP, as coached teachers had the greatest decrease in ODRs for Black students and lowered student disruptive behavior (assessed by ASSIST obs), participated in more proactive practices and had improved self-report ratings. The results of the student outcome and obs differences were statistically significant but might not be meaningfully significant.</p>
<p>Brown & Crippen, 2016: 6, once-per-month Saturday workshops for high school science teachers (called STARTS) to collaborate in lesson planning, growing cultural awareness, and professional growth tasks.</p>	<p>Qualitative: matrix and inductive analysis</p>	<p>5 life science teachers (all female, experience ranging from 2-24 years) from one culturally diverse Southeastern school district.</p>	<p>Specific to science instruction: "learner-centered environments that 'pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring to the educational setting'" (p. 472).</p>	<p>Program artifacts, group interviews, video recordings during Saturday sessions, culturally responsive science units, and reflective writing prompts from every activity.</p>	<p>None.</p>	<p>The STARTS program changed science teachers' knowledge and practice. The most helpful training activities included critically reflecting on their practice and thinking about ways to improve as they learned more about their own cultural identities. However, researchers found limited implementation of CR practice.</p>
<p>Cavanagh et al., 2014: Intervention to implement changes to be more inclusive of Latino students and families according to their</p>	<p>Qualitative: typological analysis</p>	<p>Does not specify the number of teachers or students or parents interviewed. Does report that the</p>	<p>Caring for students well-being as a key role for teachers, culturally responsive pedagogy of relations (placing relationships</p>	<p>Group interviews with teachers</p>	<p>Group interviews with students & parents</p>	<p>Developing a culture of care was challenging for the school. Some steps were taken, and some improvements were made but they were inconsistent.</p>

feedback. Teachers engaged in restorative justice training.		schools equity team participated in data collection and interviews.	at the center of teaching and learning), and restorative practice.			
Dorado et al., 2016: School district-wide effort to implement trauma informed care in schools with high rates of trauma-exposed students. Training consisted of school-wide workshops and follow up coaching/consultation.	Mixed methods – program evaluation	3 elementary schools and one k-8 school. Demographics of students across schools: 47% female, 38% African-American, 34% Latino, 76% Free and reduced price lunch.	Understand trauma and stress; establish safety and predictability; foster compassionate and dependable relationships; promote resilience and social emotional learning; practical cultural humility and responsiveness; facilitate empowerment and collaboration	Pre/post assessments for staff about trauma knowledge and perceived student engagement	Longest implementing school's disciplinary data Child and Adolescent Needs and Strengths Scale for students receiving therapy	Significant changes in staff knowledge about trauma-informed care, significant changes for teacher perceptions of student engagement. Large decrease in exclusionary discipline after 5 years of implementation. Significant improvements for five CANS items.
Ezzani, 2014: District-wide attempt to increase conversations about diversity among school staff members through one-hour, once-per-semester workshops over 3 years.	Qualitative Case Study	7 schools (3 elementary, 2 middle, 1 high) from a school district in Texas that had influx of diverse students in the 1980s. Interviewees were school and teacher leaders in focus groups.	5 standards for culturally proficient organizations: 1) understand how culture of self/organization affects others; 2) acknowledge differences as diversity and not responses that are inappropriate; 3) learn strategies to resolve conflict that arises due to cultural difference; 4) change how things are done to recognize differences; and 5) integrate these adaptations through professional development	Focus group interviews with school staff, district artifacts, observations of PD sessions, and classroom walk-throughs.	None	Due to the schools' power to implement the intervention differently, it is unclear if there was a significant change in participant attitudes. However, because the main goal was to increase vocabulary and space to discuss difference, the author thinks that these efforts will be maintained and could lead to future, deeper change in the schools and district leadership.

Fallon et al., 2018: 3 teachers were trained in self-monitoring and then randomly exposed to performance feedback based on their self-reported intervention implementation fidelity.	Quantitative single case with randomized A-B-C-D multiple baseline design	3 teachers (one White male, one Black female, and one Latina female) in one middle school in an urban school district in the Northeast	Increasing positive, equitable interactions with students, setting high expectations, teaching social skills, and using students culture and language in instruction.	Observations of teachers' treatment fidelity	Observations of class-wide student behavior	Teachers implemented with more fidelity during the self-monitoring phase and even more in the self-monitoring with performance feedback phase. Students were more engaged during performance feedback phase for 2/3 teachers. Students were less disruptive during performance feedback phase for one teacher.
Gregory et al., 2014: 1-day workshop and one year of coaching (My Teaching Partner – Secondary) to decrease disproportionality in discipline referrals for African American students.	Quantitative RCT	82 Teachers (39 intervention & 43 control) and 979 students of the participating teachers (59% black, 30% white), all from one midsized city in a Southeastern state	Teaching in line with the domains of the CLASS observation system /Reduction in exclusionary discipline practices for African American children	Teacher demographic variables (gender, race, years of experience, course subject area)	Student demographic variables (race, gender, SES) and school records (ODRs, prior EOY exams)	African American Students in intervention teachers' classrooms were two times less like to be given exclusionary discipline than African American students in control classrooms.
Gregory et al., 2016: 1-day workshop followed by 2 years of biweekly coaching using the My Teaching Partner-Secondary program and the CLASS observation system to help reduce disproportionality in ODRs.	Quantitative RCT	86 teachers (56 female; 56% White, 33% Black) from 5 schools	Teaching in line with the domains of the CLASS observation system /Reduction in exclusionary discipline practices for African American children	Teacher demographics (gender, race, years of experience, course subject area) and CLASS observation domains	Student demographics (race, gender, SES), and school records (ODRs, prior EOY exams)	Teachers who participated in 2 years of MTP-S coaching had no significance difference of discipline referrals between black students and their peers (control teachers did have a difference). This was maintained into year 3 while the teachers were not being coached. Results also indicated that teachers who used high-level problem solving in their classrooms tended to issue less referrals in general.

<p>McKenney et al., 2017: 2 phases of voluntary consultation (total duration lasting between 25-40 weeks) focused on improving teachers' classroom management and culturally responsive practice.</p>	<p>Quantitative Case study: multiple baseline</p>	<p>3 elementary school teachers with 7-15 years of experience (all 3 female, 2 white and 1 black) who taught in diverse classrooms (defined broadly) in St. Louis metropolitan area around the time of Michael Brown killing and subsequent protests in Ferguson</p>	<p>The natural and integrated application of multicultural competencies/sought-out knowledge of students into the curriculum</p>	<p>Frequency of occasions teachers used labeled praises and provided opportunities to respond as well as pre/post self-report Cultural Responsiveness Questionnaire. Assessed for social acceptability with Treatment Evaluation Inventory-SF</p>	<p>Frequency of student disruptions (defined by each teacher) as measured by observation throughout the study</p>	<p>Disruptions steadily decreased and maintained at around 5 disruptions per 25 minutes after classroom management phase of consultation. Also maintained throughout cultural responsiveness phase so results of that could not be fully estimated (though teachers' used of culturally responsive strategies increased during CR phase.)</p>
<p>McKoy et al., 2017: Weeklong workshop to help music teachers be more culturally competent in their teaching practice and advising of student teachers.</p>	<p>Mixed methods</p>	<p>18 music educators (15 female, 3 male, 13 White, 5 African American) who were partners with UNC Greensboro.</p>	<p>Culturally responsive teaching incorporates cross-cultural competence, increased global perspective, and multicultural music through using the cultural characteristics, experiences, and perspectives of ethnically diverse students as conduits to teach them more effectively.</p>	<p>Pre-post surveys. Post surveys had open-ended questions.</p>	<p>None.</p>	<p>Participants were more familiar with culturally responsive teaching after the workshop but were still unsure of how to implement it. Participants came to a greater understanding of culture and the potential harm in making assumptions about students. Some teachers were willing to reexamine their own practices. Teachers came to understand that students' skills, knowledge and dispositions toward music are cultural.</p>

<p>Parker et al., 2017: A masters-level university course for practicing teachers called culture in the math classroom. Outcomes are from end of the semester projects intended to broaden teachers' understanding of culturally diverse students and to maximize these students' success</p>	<p>Qualitative -- narrative inquiry</p>	<p>58 practicing secondary mathematics teachers (60% women, 40% men) in the Rocky Mountain region who taught in schools with average of 27-35% minority students, 7-10% of ELLs and 26-43% students FRL status. Projects from only 30 teachers were included in analysis</p>	<p>"Dispositions grounded in cultural awareness to work to know, understand, and support the engagement and learning of all students" (p. 399).</p>	<p>Class Cultural Inquiry Process (CIP) project. Projects were included in analysis if teachers described a positive perspective change with regard to student culture.</p>	<p>None</p>	<p>Teachers indicated an increased willingness and confidence in reaching out to, understanding, and supporting students they initially found puzzling. However, the authors were unable to make statements about what mechanisms were behind this change (though they hypothesized it was due to the project in addition to the teachers' preexisting levels of empathy for their students).</p>
<p>Portes et al., 2018: 100 hours of training in Instructional Conversations and then year of a follow up coaching during the implementation. Tracked the achievement of ELL students in control and treatment classrooms.</p>	<p>Quantitative quasi-experimental design (attempted RCT but baseline equivalencies were not achieved through randomization)</p>	<p>74 teachers (40 treatment, 34 control) and 1521 3rd and 5th grade students across 14 Southeastern school districts.</p>	<p>CREDE pedagogical standards (IC is the last one). "Instructional practices centered on students' backgrounds and experiences [that] promote meaningful learning, student engagement and motivation, and access to high quality curriculum" (p. 491).</p>	<p>None noted.</p>	<p>Standardized test scores in reading, ELA, science, social studies, and math (focused on ELL students but collected data from all students)</p>	<p>The intervention positively influences ELA test scores after controlling for pretest scores. This was true for both ELL and non-ELL students.</p>
<p>Powell et al., 2016: 3 workshop sessions of professional development and one year of coaching to help elementary teachers increase their use of culturally responsive instruction in the classroom.</p>	<p>Mixed Methods (concurrent triangulation design)</p>	<p>27 teachers (all but 1 were white) from 4 elementary schools that were diverse (2 rural, 1 town, 1 city; free and reduced price lunch rates ranged from 60%-90%).</p>	<p>Instructional practices that connect learning to the cultural knowledge and experiences of students and that draws on students' cultural and linguistic strengths and frames of reference in instruction, thereby resulting in high</p>	<p>Observations using the Culturally Responsive Instruction Observation Protocol (CRIOP) Teacher interviews (using</p>	<p>Reading and Math EOY achievement tests from 2 of the 4 schools</p>	<p>Teachers did increase their use of CRI from fall to spring. Teacher interviews revealed that building positive relationships were most helpful in their implementation of CRI. Students in classes with high CRI implementing</p>

Sigman et al., 2014: 3 5-day workshops to help science teachers in rural Alaska plan lessons that are appropriate for the indigenous culture and unique local marine life; workshops included local scientists.	Qualitative	43 science teachers and 35 scientists from rural Alaskan communities (21 teachers were from states other than Alaska).	levels of student achievement. The inclusion of diverse cultural contexts as a means to address diversity and equity in science education. Students' inherent cultural fund of knowledge can be leveraged and transformed into science knowledge over time.	the CRIOP interview tool) Formative interviews throughout the workshop, pre/post surveys, lesson plans that teachers created (35 LPs were reviewed)	None.	teachers had significantly better math scores. Shift occurred for teachers from knowing science content to thinking of "science literacy." The interactive and collaborative atmosphere helped teachers feel empowered. Lesson plans ranged in the inclusion of culturally responsive practices.
---	-------------	--	--	--	-------	---

Table 1.3

Evidence of Credibility Indicators in Seven Quantitative Studies of Inservice Interventions to Promote CRT

In-Text Citation	Design	Control group	Internal Validity	Statistical Validity	Construct Validity	External Validity
Bradshaw et al., 2018	RCT	Y	X	X	X	
Dorado et al., 2016	Program Evaluation	N				
Fallon et al., 2018	SCD	N	X	X	X	
Gregory et al., 2014	RCT	Y	X	X	X	
Gregory et al., 2016	RCT	Y	X	X	X	
McKenney et al., 2017	SCD	N	X		X	
Portes et al., 2018	Quasi-Experimental	Y				

Table 1.4

Evidence of Credibility Indicators in Nine Qualitative Studies of Inservice Interventions to Promote CRT

In-Text Citation	Triangulation	Disconfirming Evidence	Reflexivity	Member checks	Collaborative Work	Audit Trail	External auditors/peer debriefing	Prolonged field engagement	Thick Description	Particularizability:
Allen & Fitzgerald, 2017	X			X	X	X	X	X		X
Bequette, 2014	X		X	X	X		X	X	X	
Brown & Crippen, 2016	X									X
Cavanagh, Vigil, & Garcia, 2014	X			X	X					
Ezzani, 2014				X				X		
McKoy, MacLeod, Walter, & Nolker, 2017							X			X
Parker, Bartell, & Novak, 2017										
Powell, Cantrell, Malo-Juvera & Correll, 2016	X					X	X			X
Sigman et al., 2014	X				X		X			

Table 1.5

Evidence of Encouraging Cultural Humility in all 26 Studies

	In-Text Citation	Transfor- mational Goal	Equity Content	Reflec- tion	Family/ Com- munity Part- ners	Total
	Allen & Fitzgerald, 2017					0
	Bequette, 2014	X			X	2
	Bradshaw et al., 2018	X	X	X		3
	Brown & Crippen, 2016			X		1
	Cavanagh et al., 2014				X	1
	Dorado et al., 2016		X		X	2
	Ezzani, 2014					0
Current Study	Fallon et al., 2018					0
	Gregory et al., 2016			X		1
	Gregory et al., 2014					0
	McKenney et al., 2017	X				1
	McKoy et al., 2017					0
	Parker et al., 2017			X		1
	Portes et al., 2018			X		1
	Powell et al., 2016	X			X	2
	Sigman et al., 2014				X	1
		Eberly et al., 2010			X	X
	Fickel, 2005			X	X	2
	Hammerness & Matsko, 2013		X			1
From Botti- ani et al. (2018)	Jones et al., 2006					0
	McAllister & Irvine, 2002			X		1
	McCormick et al., 2013			X		1
	Ryan et al., 2007					0
	Shriberg et al., 2012				X	1
	Thompson & Byrnes, 2011				X	1
	Vincent et al., 2011					0

2 CULTURAL HUMILITY IN THE CONTEXT OF TEACHER-STUDENT RELATIONSHIPS

The social practices between teachers and students in schools have been shown to be influenced by cultural differences between these two groups; for example, researchers have demonstrated that classroom discipline (e.g., Gregory et al., 2015; Skiba, Michael, Nardo & Peterson, 2002) and the relationship quality between teachers and students (e.g., Split & Hughes, 2015) can be different for students based on their cultural group (e.g., race, religious background, etc.). These differences often have resulted in systemic inequity for students from certain groups: for example, black students were more likely to receive harsher consequences in schools than white students (Skiba et al., 2002). This is a concerning factor for schools in the United States, as the student population is becoming increasingly diverse, but the teaching force continues to be overwhelmingly white and female (USDOE, 2016).

One response to this potential cultural gap in the classroom has been to help teachers and other public educators develop multicultural competence (MCC). However, criticism of this construct has grown, as MCC implies that it is possible for someone to be “competent” in all the cultures of those receiving their services (Hook & Watkins, 2015). An alternative construct to examine is Multicultural Orientation (MCO), which describes one’s general “way of being” with others from diverse backgrounds (Hook, Davis, Owen, Worthington, & Utsey, 2013; Owen, Tao, Leach, & Rodolfa, 2011). MCO has been discussed as a more useful construct compared to MCC, especially in populations of helping professionals such as physicians, nurses, and child welfare workers, as it describes a philosophical position that lends itself to a more generalizable set of interpersonal and intrapersonal communication skills (Marian, 2016).

Cultural humility, which is an aspect of one's MCO, has been identified as particularly important for service providers in caring professions (Hook et al., 2013; Hook & Watkins, 2015; Tervalon & Murray-Garcia, 1998). Cultural humility refers to a continual process of self-reflection and critique leading to an attitude of openness to others' experiences and an awareness of one's personal limitations in understanding the cultural backgrounds of others (Hook et al., 2013; Hook & Watkins, 2015; Tervalon & Murray-Garcia, 1998). It is a quality that can be considered from an intrapersonal and interpersonal perspective and has been theorized to influence the quality of relationships between people (Davis et al., 2013). Researchers have attempted to encourage and assess levels of cultural humility in many helping fields, including medical professions (e.g., Chang, Simon, & Dong, 2012; Tervalon & Murray-Garcia, 1998), social work (Fisher-Borne, Montana Cain, & Martin, 2015; Ortega & Coulborn, 2011) and counseling (e.g., Hook et al., 2013). There is evidence that helpers with high levels of cultural humility were able to develop a better working alliance with clients, which then led to more effective treatment (e.g., Davis et al., 2018; Hook et al., 2013). Thus, cultural humility appears to be an important concept in helping professions, especially when service providers are working with clients with a variety of cultural experiences. However, cultural humility has been explored minimally within the fields of research related primary and secondary educational settings. Because humility has been conceptualized as a relational process (Davis et al., 2013), one important area to investigate is how cultural humility could affect the quality of relationships between teachers and students.

The current study aims to explore if and how cultural humility is related to teacher-student relationship quality (TSRQ) in secondary school classrooms. High-quality teacher-student relationships have been shown to predict beneficial outcomes for students (e.g., Roorda, Jak, Zee, Oort, & Koomen, 2017), and previous research has demonstrated a cultural gap in

relationship quality between teachers and students of differing cultural backgrounds (e.g., Saft & Pianta, 2001; Split & Hughes, 2015). We hypothesize that cultural humility may moderate the relationship between students placed at risk for high-conflict teacher-student relationships and their perceived TSRQ.

Cultural Humility

Recent reviews of the literature (e.g., Davis et al., 2018; Foronda, Baptiste, Reinholdt, & Ousman, 2016; Mosher, Hook, Captari, Davis, DeBlaere, & Owens, 2017) regarding cultural humility identified this trait as a “process of openness, self-awareness, being egoless, and incorporating self-reflection and critique after willingly engaging with diverse individuals” (Foronda et al., 2016 p. 213). Cultural humility is likely necessary when there is a power imbalance present, and results in mutual empowerment, respect, partnerships, optimal care, and lifelong learning for the culturally humble individual (Foronda et al., 2016). In a conceptual analysis of cultural humility, Foronda et al. (2016) found that research and theoretical publications about cultural humility have primarily existed within the fields of medicine, nursing, social work, and ‘others’ (p. 215), including counseling and community psychology (e.g., Hook et al., 2013; Ross, 2010). Mosher et al. (2017) reviewed the exploration of cultural humility within the field of counseling psychology, finding that empirical studies of cultural humility suggested that it has been most influential in developing a healthy therapeutic relationship between two parties of different cultural backgrounds, in navigating cultural ruptures, and in facilitating continual multicultural development through processes of self-reflection and critique. Davis et al. (2018) reviewed the preliminary evidence to support the importance of components of MCO (including cultural humility) in psychotherapy settings and concluded that it supports two possible theories of cultural humility’s influence on relationships (Davis et al., 2018): the social bond and the social oil

hypothesis. The social bond hypothesis states that perceived humility in relationships helps protect against the damages done by microaggressions and other negative interactions; this hypothesis suggests that forgiveness is granted more easily to those with higher perceived cultural humility (Davis et al., 2013). The social oil hypothesis has been explored less often in the literature, but there is some support for the theory, which states that humility is a sort of lubricant in tense relational situations that can be calming to both parties and helps resolve cultural conflict more rapidly and more often (Davis et al., 2018).

These reviews identified literature on cultural humility as an emerging field of research that requires further empirical exploration to understand how it influences best practices in multiple caring professions and contexts (Davis et al., 2018; Foronda et al., 2016; Mosher et al., 2017). One challenge in expanding this research is measuring real levels of cultural humility within practitioners; measures of cultural competence and related constructs have often relied on self-reports, which can bias responses as a result of social acceptability or low self-awareness (Stanhope, Solomon, Pernell-Arnold, Sands, & Bourjolly, 2005). Measuring all types of humility has been especially complicated due to the presumed paradox of self-reported humility measures (Davis, Worthington, & Hook, 2010), which posits that truly humble individuals will likely underreport their levels of humility, and less-humble individuals will overestimate their own humility. An alternative reporting method has been used in the field of counseling psychology, where cultural humility has most often been measured using the Cultural Humility Scale (CHS; Hook et al., 2013) that asks clients to rate perceived levels of cultural humility in their counselors. However, to our knowledge, this scale has not been adapted outside of the field of counseling psychology. Other studies have used qualitative evaluations of cultural humility, such that personal reflections of practitioners have been analyzed for characteristics in line with how cultural

humility has been conceptualized in the research literature (e.g., Tinkler & Tinkler, 2016; Tormala, Patel, Soukup, & Clarke, 2018). More research to help understand empirical measurements of cultural humility, especially in multiple professional fields, will aid in understanding the impact of cultural humility in helping professions.

Teacher-Student Relationships

Positive teacher-student relationships have been described as those relationships that are high in emotional support, trust, and respect between the teacher and student (e.g., Pianta, Hamre, & Stuhlman, 2003; Ryan & Deci, 2000). Negative relationships have been considered to be high in conflict, low in trust, and low in respect (e.g., Pianta et al., 2003; Ryan & Deci, 2000). Strong teacher-student relationships have been correlated with students' academic success and positive social-emotional development (Roorda, Jak, Zee, Oort, & Koomen, 2017), whereas weak/negative teacher-student relationships have been correlated with poorer academic and social outcomes for students (e.g., Roorda et al., 2017). Improvements due to positive teacher-student relationships have been shown to be larger for students who are placed at risk for poorer academic and behavioral outcomes, such as students from low-income and/or minority backgrounds (McGrath & Van Bergen, 2015). Additionally, improving teacher-student relationship quality has eliminated racial gaps in discipline practices for students from diverse racial and socioeconomic backgrounds (Gregory et al., 2016). Thus, improving teacher-student relationships could be extremely beneficial for students from racially diverse backgrounds, as students with strong teacher-student relationships have tended to achieve more and to remain more connected to school than their peers who have poor teacher-student relationships.

The role of teacher-student relationships in student development has been conceptualized most commonly using extended attachment theory (Pianta et al., 2003) and social motivation

theory (Ryan & Deci, 2000). Extended attachment theory posits that students can form attachments to teachers that are similar to the attachments of children to parents (Pianta et al., 2003). Forming these attachments is thought to be due to previous attachments the adult and child have formed in their lives, as well as each party's mental model of the relationship (Davis, 2006; Pianta, 2003). For example, when a teacher views a child's behavior as primarily positive, that child will likely elicit more positive attention from the teacher. Similarly, when a child expects a teacher to respond in a positive, nurturing manner, that child then feels safe and will be more likely to trust the teacher and be more compliant. Social motivation theory suggests that students have three major, integrated needs in order to remain motivated in school: autonomy, competence and relatedness (Ryan & Deci, 2000). Positive teacher-student relationships have been thought to meet students' need for relatedness, which enhances their autonomy and competence to increase their engagement in school.

Interestingly, neither extended attachment nor social motivation theory have addressed the way cultural differences may function in teacher-student relationships. One theory, Relational-Cultural Theory (RCT; Miller 1976), posits that growth-fostering relationships (such as teacher-student relationships) must consist of mutual empowerment, mutual empathy, and authenticity. This is true especially in relationships where there is a power imbalance and/or cultural differences between each person (Jordan, 2008). RCT theorists have suggested that establishing growth-fostering relationships in these scenarios is difficult and thus specific strategies should be developed to overcome these potential differences (Jordan, 2008). Enhancing culturally-responsive teaching strategies has shown some evidence of increasing growth-fostering relationships between teachers and students (e.g., Chowela, Goodman, West-Olatunji, & Amatea,

2014). Understanding more about the connection between cultural humility and TSRQ could help further knowledge in this area.

Risk Factors Associated with Low Teacher-Student Relationship Quality (TSRQ)

Considering the clearly established importance of teacher-student relationships in schools, it is equally important to investigate what factors in a classroom may predict positive or negative relationships between teachers and students. Researchers have demonstrated that student factors such as race/ethnicity, gender, socioeconomic background, prior achievement, disability status, and behavioral history have predicted relationship quality with teachers (see McGrath & Van Bergen, 2015 for a review). In general, minority students from low-income backgrounds with a tumultuous behavioral history and low achievement have been shown to be at greater risk for developing low quality relationships with teachers (McGrath & Van Bergen, 2015). Some longitudinal studies (e.g., Jerome, Hamre, & Pianta 2009; Split, Hughes, Wu, and Kwok, 2012) have demonstrated that, while relationship trajectories can change over time, students matching this demographic description were generally overrepresented in groups of students that teachers rated as having higher levels of conflictual relationships, starting as early as pre-Kindergarten. Other studies have suggested that forming poor relationships with teachers in early elementary school negatively influenced TSRQ even through adolescence (Hamre & Pianta, 2001). Thus, the potential impact of one year of negative TSRQ could be quite detrimental to long-term student outcomes.

Students' emotional adjustment can affect the quality of relationships with their teachers. Students with high levels of externalizing behavior were more likely to have relationships with teachers that are high in conflict (Lei, Chu, & Chiu, 2016). In their 2016 meta-analysis, Lei et al. reviewed 57 studies that investigated the relationship between TSRQ and evidence of student

externalizing behavior problems. Findings suggested that positive TSRQ was significantly correlated with lower rates externalizing behavior problems, and that negative TSRQ was significantly correlated with higher rates of externalizing behavior problems. Other studies (e.g., Roorda, Verschueren, Vanraeyveldt, Van Craeyveldt, & Colpin, 2014; Zhang & Sun, 2011) of children's emotional adjustment and TSRQ have found that children's internalizing behaviors and prosocial behaviors could also influence TSRQ, though they were generally less predictive of relationship quality when compared to externalizing behavior problems (Zhang & Sun, 2011). One finding of Lei et al. (2016) was that culture significantly moderated the relationship between TSRQ and externalizing behavior problems – the correlation between TSRQ and emotional issues varied between Eastern and Western cultures. This could indicate that cultural expectations for behavior may influence the relationship between behavior problems and TSRQ. This is important, as certain populations of students (i.e., African American, male) have been shown to be more likely to be perceived by teachers as having behavior problems (Jerome et al., 2009; Matthews, Kizzy, Rowley, & Contina, 2010). It is possible that, similar to its influence on perceptions of externalizing behavior problems, cultural background could also influence teachers' perceptions of students with varying levels of internalizing behaviors and prosocial skills. Understanding the role of cultural expectations could increase helpful interventions for the most vulnerable groups of students.

Less research has been conducted to assess which teacher factors may contribute to relationship quality with students. Some teacher characteristics, such as experience, gender, and age have shown little to no effect on relationship quality (Hamre & Pianta, 2006). Teacher racial/ethnic background, especially when compared to student racial/ethnic background, has shown some prediction of relationship quality with students (e.g., Split & Hughes, 2015), suggesting that

racial/ethnic match is predictive of better relationships between teachers and students (e.g., Thijs, Westhof, & Koomen, 2011). Cherng & Halpin (2016) found that minority students had significantly different perceptions of teachers based on teacher racial/ethnic background, which suggested a relational effect from a student's perspective based on teacher characteristics. More research in this area may clarify the relationships between teacher race/ethnicity and TSRQ, especially when it is compared with student race/ethnicity.

The Current Study

Hook et al. (2013) found that the working alliance between a counselor and client significantly mediated the relationship between counselors' cultural humility and the effectiveness of their practice. Since this study, multiple investigations have demonstrated that cultural humility plays a significant role in the development of therapeutic working alliance (Davis et al., 2018). Using a theoretical model that is congruent with the findings of Hook et al. (2013), it is possible that cultural humility is significantly related to a teacher's ability to build positive relationships with his/her students. This relationship could help clarify and extend the conflicting/limited evidence about teacher/student racial/ethnic match, student emotional adjustment, and teacher-student relationship quality. It is possible that cultural humility is particularly important for teacher-student relationships when teachers are working in classrooms that serve students from diverse backgrounds (or with students who have different cultural backgrounds from the teacher).

We hypothesize that teachers with high levels of cultural humility may be able to build relationships with their students that are more effective. Both Hook et al. (2013) and Lei et al. (2016) provided useful models for exploring the role of cultural humility in TSRQ. Hook et al. found that cultural humility plays a significant role in working relationships, which then in turn affect practice. Lei et al. demonstrated that student externalizing behaviors predicted TSRQ, and

that cultural variables moderated this relationship. For the current study, it is predicted that teachers' cultural humility may moderate the relationship between students' self-reported emotional adjustment and their perceived relationship quality with teachers (see figure 1). Finally, to expand our exploration into the role of culture in TSRQ, the current study will examine the role of demographic variables (race/ethnicity, gender, etc.) in students' perceptions of their teachers' cultural humility and their perceptions of TSRQ.

The research questions we will answer in order to test this hypothesis are as follows:

1. Is cultural humility related to TSRQ?

We hypothesize that student-perceived teacher cultural humility will be related to their perceptions of TSRQ. Previous studies that examined cultural humility in the context of counseling have demonstrated that it is significantly related to the counselor-client working alliance (e.g., Hook et al., 2013), which is a relational construct. Thus, we postulate that TSRQ will function similarly to working alliance, and that it will be significantly related to student perceptions of their teachers' cultural humility.

2. Does cultural humility moderate the relationship between student emotional adjustment and TRSQ?

Lei et al. (2016) found that cultural factors significantly moderated the relationship between externalizing behavior problems and TSRQ; thus, we believe that teachers' cultural humility will significantly moderate the relationship between student emotional adjustment and their reports of TSRQ. Similar to the mechanisms described in the social oil and social bond hypotheses (Davis et al., 2018), we think cultural humility will act as a buffer in teacher-student relationships. Specifically, we hypothesize that high ratings of cultural

humility will be correlated with more positive relationships than they are predicted to be based on students' emotional adjustment. (see figure 1).

3. How does the relationship between cultural humility and TSRQ differ by demographic subgroups (e.g., race, gender, teacher-student racial/ethnic match)?

Past research on cultural humility has suggested that it is most important in relationships between individuals who come from different cultural backgrounds. Thus, while our examination of demographic factors on the role of moderation is exploratory, we expect that cultural humility will have a stronger effect when students and teachers are demographically different from each other (i.e., when a student-teacher cultural mismatch is present).

Method

Participants

Participants were high school students from one school district in the Southeastern United States that was participating with a university in a larger grant-funded research project regarding student mental health; data were collected from every high school student in the district ($N = 2002$). The sample population was majority-minority (45% black, 37% White, 7% Latino, 10% other) and roughly half male (46%) and female (52%). Full student demographics are reported in Table 1. In the district overall, 77% of students are eligible for free- and reduced-price lunch.

Measures

All variables were student-reported. Evidence from prior studies has suggested that high school students' responses to measures of teacher-related constructs were both reliable and valid (Follman, 1992; Peterson, Wahlquist, & Bone, 2000), and were not more susceptible to bias when compared to ratings by college students and other adult groups (Worrell & Kuterbach,

2001). Additionally, though student reports of TSRQ have previously been found to have low correspondence with teacher reports (e.g., Hughes, 2011), student reports have been shown to be more predictive of important student constructs including school belonging and feelings of academic competence (Hughes, 2011; Murray, Kotsky, & Hauser-McLean, 2016).

Teacher-student relationship quality. TSRQ was assessed with the 17-item Inventory of Teacher-Student Relationships (IT-SR; Murray & Zvoch, 2011). Items (e.g. “My teacher understands me”) are rated on a 4-point scale, ranging from 1= *almost never or never true* to 4 = *almost always or always true* and load onto three latent factors: Communication, Trust, and Alienation. It was adapted from the Inventory of Parent and Peer Attachments (IPPA; Arnsden & Greenberg, 1987), which is based in attachment theory and has been shown to be a reliable and valid measure of adolescents’ attachments to individuals other than parents (Muris, Meesters, Marion, & Zwambag, 2001). Items on the Communication and Trust subscales assess student perspectives of teachers’ understanding, responsiveness, and sensitivity; Items on the Alienation subscale assess the degree to which students feel disconnected from teachers. The IT-SR has shown evidence of moderate to strong internal consistency among diverse adolescents, with Cronbach’s alpha coefficients estimated at .72 for the Alienation subscale, .84 for the Trust subscale, and .89 for the Communication subscale (Murray & Zvoch, 2011). Concurrent validity was also demonstrated in a diverse sample of adolescents (Murray & Zvoch, 2011).

Student Emotional Adjustment. Students reported their own levels of risk for externalizing and internalizing mental health and behavioral issues using the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). The SDQ is a 25-item measure of social, emotional, and behavioral strengths and difficulties in children and adolescents. There are parent, teacher, and self-report versions. The self-report version has 5 latent factors each with 5 indicators. The scales

are entitled Hyperactivity (items refer to levels of activity and impulsivity), Emotional Symptoms (items refer to internalizing symptoms of anxiety and depression), Conduct Problems (items refer to behavioral difficulties like fighting and tantrums), Peer Problems (items refer to problematic peer interactions), and Prosocial (items refer to positive interactions with others). Four of the five factors load onto two subscales: Externalizing (made up of Hyperactivity and Conduct Problems), and Internalizing (made up of Emotional Symptoms and Peer Problems). A total difficulties score is calculated from the internalizing and externalizing subscales. The SDQ is very widely used not only in the United States but also around the world. It has been translated into over 80 languages and has been used in over 4000 research articles (Uses of the SDQ, 2018). The SDQ has been shown to provide acceptable levels of validity and acceptable to good internal consistency in measuring self-reported emotional and behavioral strengths and difficulties and in adolescent samples from multiple countries, with Cronbach's alphas of .76 - .80 (e.g., Dickey & Blumberg, 2004; Van Roy, Veenstra, & Clench-Aas, 2008). However, recent literature has shown that the 5-factor model of the SDQ was a poor fit with racial/ethnic minority adolescent samples from the United States (e.g., Twyford, Buckley, Moffa, & Dowdy, 2018). Thus, in our exploration of the SDQ, multiple factor structures were considered to ensure that student emotionality was being measured with sufficient reliability and validity.

Demographic variables. Demographic information was collected about both teachers and students, though all information was student-reported as the anonymity of the teachers and students was protected throughout the data collection process. Students provided basic information about their teacher's demographic background (race/ethnicity, gender, age bracket). Student-reported teacher demographics suggested that the teacher population is roughly half White (53%) and female (52%). Students also reported the difficulty of the class taught by the teacher

they were rating and suggested that the majority of students (74%) found their classes easy. Full student-reported teacher demographics can be found in Table 2.2.

Cultural humility. Cultural humility was assessed using an adaptation of the 12-item Cultural Humility Scale (Hook et al., 2013), the Cultural Humility Scale for Students (CHS-S). The original CHS was developed to obtain ratings of counselors' cultural humility. Items (e.g., "Regarding core aspects of my cultural background, my counselor assumes he/she already knows a lot") are rated on a 5-point Likert-type scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree*. This version of the scale has demonstrated evidence of convergent validity when compared with measures of counselor multicultural competence (though cultural humility was more predictive of working alliance than multicultural competence). Additionally, the CHS has demonstrated excellent reliability, with a Cronbach's alpha of .93 (Hook et al., 2013). Hook et al. (2013) found that the CHS had two factors, one that suggested more positive ratings of cultural humility, and one that suggested more negative ratings of cultural humility. In addition to these subscales, the measure provides an overall rating of counselors' cultural humility from the perspective of their clients.

Srisarajivakul et al. (2019) adapted the CHS for adolescents in classroom settings to rate their teachers' cultural humility, renaming the scale the Cultural Humility Scale for Students (CHS-S). Compared to the CHS, the CHS-S was revised in vocabulary and structure to be appropriate for the setting and age of the respondents. Students were asked to complete the scale for their third period teacher in order to control for students' selection bias. Additionally, the beginning portion of the scale that asked participants to write the most central aspect of their cultural background and rate its importance was converted to a drop-down menu format. The areas of possible cultural background were presented in this menu using adolescent-friendly language

(e.g., “neighborhood” instead of “socioeconomic status”). The largest portion of students chose Race/Ethnicity as the most central aspect of their cultural background (32%), followed by Nationality (20%), Gender (15%), Neighborhood (12%), and Language Spoken at Home (9%). See Table 2.3 for the full report of students’ most salient cultural identities. Appendix A contains a completely revised form of the CHS-S.

In a sample of racially and socioeconomically diverse middle school students, Srisarajivakul et al. (2019) found that an 11-item bifactor factor structure of the CHS-S was the best fit to the data, with one factor representing positive ratings of cultural humility, one factor representing negative ratings of cultural humility, and one general factor representing an individuals’ total cultural humility. Adequate reliability of each factor was reported (McDonald’s omega (ω) = .85 for the overall factor, .85 for the positive factor, and .79 for the negative factor). In addition to the research questions specified above, this study attempted to confirm the fit of the factor structure of the CHS-S with a diverse sample of high school students.

Procedures

Data Collection. Data collection efforts were coordinated by the grant director in the district. The SDQ was administered to all the high school students in the district as part of a universal screening process that was required as part of the grant. The CHS-S and IT-SR were added to that universal screening process. The survey was created using an online survey software that was sent via a link to the grant director. The grant director distributed the link to both high schools in the district. The schools ensured that all students took the survey in a supervised computer lab. The responses were sent to a private account only accessible by the grant director. Once the survey data were collected, the grant director compiled the responses, removed student identifying information, and deposited the data into a private and password-protected online data

management account that is accessible only by the principal investigators and an IRB-approved research team. This data collection process took place in September 2018.

Informed Consent. Student assent was obtained before students completed the IT-SR and CHS-S. Passive parental consent was obtained for students who were willing to participate in the study. The data collection was anonymous and these procedures, including the use of passive parental consent, received IRB approval. Details of the project were distributed to parents who had the right to remove their children from participation at any time. The data collection was administered through online software that required no direct interaction with university researchers and participating students; administration of the survey was coordinated by the grant director in the district. The survey was completely anonymous. Though some demographic information was collected from students, no identifying information was collected. The anonymity of the survey minimized the risk to students, as the university researchers were not able to gain access to any specific identifying information per student. Finally, though participants could experience feelings of discomfort related to answering survey items about their feelings, attitudes, and experiences, they were given the option to skip items without penalty and could stop participating at any time.

Ethical Considerations. The primary ethical risks involve issues of confidentiality. All anonymous response data were stored in a password-protected Box account, so risk of identification was extremely low. Electronic copies of the survey were kept on firewall- and password-protected computers without any data that identified any individual student. Participants could have experienced feelings of discomfort related to answering survey items about their feelings, attitudes, and experiences. Participants were told that they could skip items without penalty, and they could stop participating at any time. In addition, parents were given contact information for

the research team if they had questions, concerns, or complaints about this study. They were also given contact information for a contact in the Georgia State University Office of Research Integrity if they wanted to talk to someone who was not part of the study team.

Analysis

The analysis procedure consisted of several steps. First, the data were examined for trustworthiness and any problematic data points (e.g., statistical outliers and inconsistent or insufficient effort responders) were removed from subsequent analyses. To answer the first, second, and third research questions, it was necessary to compare latent factors that indicate cultural humility, student emotional adjustment, and TSRQ. This required estimation of measurement models for each scale that appropriately accounted for the variability in students' raw responses. Thus, the initial measurement models for each scale were then estimated and adjusted for optimal fit and each scale was evaluated for measurement invariance. The latent factors could then be compared through structural modeling to answer research questions 1 and 2. The hypothesized structural model was estimated from the factor scores of each measurement model. To answer research question 3, the influence of teacher and student demographic factors on the final structural model was assessed by testing if there was a significant difference in moderation effects between demographic subgroups. All measurement and structural analyses were performed in *Mplus* Version 8 (Muthén & Muthén, 1998-2017).

Data trustworthiness. Data were initially evaluated for trustworthiness using several post hoc methods of detecting inconsistent responders (Meade & Craig, 2012). Responders were removed from analysis if they met one or both of the following criteria: response pattern indices including LongString (Johnson, 2005) and response variance that were substantially different from average, and outlier indices including Mahalanobis distances that were significantly

different from normal (Osborne, 2010). After the removal of inconsistent responders, missing data were inconsequential (missing data patterns on each measure ranged from 26 to 45) and were thus deleted listwise during analysis.

Measurement modeling. Confirmatory Factor Analyses (CFAs) were performed on each measure in order to account for the maximum amount of measurement error in the structural model (Kline, 2011). For each scale, initial CFAs were run using the predetermined factor structure based on previous literature. Then, if necessary, theoretically based adjustments were made to the models in order to improve model fit. Criteria for good model fit included CFI > .90, root mean square error of approximation (RMSEA) values less than .08, and standardized root mean square residual (SRMR) values less than .08 (Marsh, Hau, & Wen, 2004; Meade, Johnson, & Braddy, 2008). The chi-square significance test was not considered a part of model fit indices because chi-square has been shown to be overly sensitive to sample size (Bentler & Bonnet, 1980). Because each scale provided Likert-style response options, the three CFAs were modelled as ordered categorical indicators using a WLSMV estimator and theta parameterization. McDonald's omega (ω) was used to assess internal reliability for each factor, as it accounts for measurement error more realistically than Cronbach's alpha in psychological survey measures, especially with complex models (Dunn, Baguley, & Brunsten, 2014).

Measurement Invariance. Measurement invariance was also assessed for each scale to ensure that reliable comparisons could be made in the structural model between individuals from different demographic subgroups. It was tested through a series of model comparisons that increased restrictions on certain parameter estimates. When evaluating measurement invariance with a categorical model, the only empirically supported index of model comparison is the chi-square difference tests computed in *MPlus* using the DIFFTEST command (Muthén, 2004). This

was problematic for fit comparisons because chi-square is sensitive to large sample sizes, complex models, and minor misspecifications (Sass, Schmitt, & Marsh, 2014), and there is evidence that this sensitivity produces statistically significant results that are not meaningfully significant (e.g., Meade et al., 2008; Sass et al., 2014). Thus, we followed the example of others (e.g., Caci, Morin & Train, 2015; Rutkowski & Svetina, 2017) and evaluated each model comparison for meaningful measurement differences (e.g., less than .010 Δ CFI, less than .015 Δ RMSEA). Configural invariance was tested first to determine if the overall model fit is the same between two groups. Next, metric invariance was tested to determine if the factor loadings were the same between two groups. Finally, scalar invariance was tested by restraining item thresholds to be equal between groups. Because the structural model was estimated using the factor scores from each CFA for each scale, the presence of metric invariance provided enough support to make further inferences from the relationships between constructs measured by each scale. If the model fit was found to be worse after imposing restraints on the factor loadings, attempts were made to locate the sources of non-invariance and free problematic items, while constraining all others to be equal. Thus, partial metric invariance models were considered acceptable if freeing less than 20% of the factor loadings produced minimal differences between metric and configural models (Dimitrov, 2010).

Due to population imbalance between certain demographic groups, invariance was assessed on only the largest groups in the population (Yoon & Lai, 2018). Thus, the presence of invariance was assessed for each scale between the following subgroups: race (Black vs. White students), gender (male vs. female students), the presence of teacher-student racial match (match vs. non-match), and student self-selected salient cultural identity differences that were indicated on the CHS-S (race/ethnicity, vs. nationality).

Structural modeling. Once the factor structure was determined and verified for each measure, the factor scores from the final measurement model of each scale were saved and utilized to estimate the hypothesized structural model using path analysis. Factor scores were utilized instead of latent variables or scale scores from the preliminary measurement models because the process of estimation with multiple complicated models became impossible for the software to operate. Using factor scores instead of latent variables still controls measurement error, especially when factor scores from the most invariant models are used in analysis (Morin et al., 2016). The results of the structural model were used to answer the first and second research questions, as they enabled the exploration of relationships between externalizing behavior, teacher-student relationship quality, and cultural humility.

Multigroup SEM. After the full structural model was estimated using factor scores, a model test was conducted between demographic subgroups to assess the three-way interactions between the effect of cultural humility and student and teacher demographic variables. The MODEL TEST option in *MPlus* tests the constraints to one or more parameters in the model using the Wald chi square difference test. Thus, two subgroups were specified, and the moderation effect was constrained to be the same between groups. If the Wald test suggested a significant difference, it would indicate that the moderation effect varied between subgroups. We were most interested in whether the effect of cultural humility varied in magnitude for Black students compared with White students, or when there was a teacher-student racial/ethnic mismatch. The effect of other demographic characteristics including student gender and student self-selected cultural identities were tested as well.

Results

Preliminary Analysis

Missing Data. Data were examined for inconsistent/insufficient effort responders and eliminated based on recommendations by Craig and Meade (2012), and Huang et al. (2012). Of the 2002 respondents, 1,612 (80.5%) were retained in analysis. One fifth of respondents is a large proportion to remove from analysis based on the assumption of insufficient effort, but it is not outside the realm of possibility for certain samples (Meade & Craig, 2012). Risk factors for a high proportion of unreliable responders include online and anonymous survey modalities and younger samples (Meade & Craig, 2012). Thus, it is reasonable that 20% of responses could be the result of insufficient or inconsistent effort.

Raw Data Screening. Examination of variance inflation factors for each scale did not suggest the presence of excessive multicollinearity, and the scores for most items were symmetrically distributed. For each scale, the Kaiser-Meyers-Olkin test of sampling adequacy was acceptable (CHS-S = .916; SDQ = .868; IT-SR = .934) and Bartlett's test of sphericity was statistically significant (all $ps < .001$).

Raw score responses were summed for each measure to preliminarily evaluate the correlations and trends of raw response data. The SDQ was scored using a preexisting macro utilized by the school district to plan services for at-risk students (Center for Leadership and Disability, 2016). This macro yields a score for each subscale as well as a score for internalizing problems, externalizing problems, and a students' total difficulties. Correlations between the SDQ subscales were as expected, with small to moderate, significant, and positive correlations between the four difficulties subscales, and small to moderate, significant, and negative correlations between prosocial and the other subscales.

The CHS-S raw scores were summed for each factor based on previous analysis – the Positive Cultural Humility (CH) total included seven items and the Negative CH total included

four items, based on the model described in the Srisarajivakul et al (2019) study. Means for these composites were 28.8 for Positive CH and 11.0 for Negative CH. Consistent with previous research using the CHS (e.g., Choe, Srisarajivakul, Davis, DeBlaere, Van Tongeren, & Hook, 2018; Srisarajivakul et al., 2019), the correlation between Negative CH and Positive CH was essentially zero ($r = .06$).

The IT-SR raw scores were summed into three subscales in accordance with the previous factor structure reported by Murray and Zvoch (2011), with 8 items on the Communication factor, 5 items on the Trust factor, and 4 items on the Alienation factor. The mean for Communication was 20.8, Trust was 16.2, and Alienation was 11.1. Correlations between these factors were unexpectedly different from previous reports, with significant and slightly positive correlations between Alienation and the other two factors ($r = .13$ for Alienation and Trust, $r = .30$ for Alienation and Communication). This suggests that some portion of the students in the sample found it possible to feel alienated from their teachers but also to communicate with them, as well as feel alienated from their teachers but also consider them trustworthy. This finding is inconsistent with past studies using the IT-SR (e.g., Murray et al., 2016), and thus seems specific to the sample in this study. Though the IT-SR has demonstrated strong psychometric properties with adolescent samples from racial minority groups (Murray et al., 2016; Murray & Zvoch, 2011), these studies were conducted with students who were slightly younger and from more urban school districts than those in the current sample. It is possible that these overall demographic differences influenced the performance of the measure. The correlation between Trust and Communication was strong and statistically significant ($r = .72$).

Thus, there were some limitations in the raw data, including the relatively high proportion of inconsistent responders and the correlations between Alienation and the other IT-SR

subscales. However, basic assumptions were met for structural equation modelling. To answer the three primary research questions with minimal measurement error, CFAs and invariance analysis were run to determine the best fit and reliability of each scale.

Measurement modeling

CHS-S. The 11-item bifactor structure proposed by Srisarajivakul et al. (2019) was used to conduct the CFA on the CHS-S. Srisarajivakul and colleagues used a continuous model for the CHS-S. Prior literature has demonstrated that polytomous data with 5 or more response choices can be modelled as continuous, but that a categorical estimator generally performs better when assessing model fit (Kozioł & Bovaird, 2018; Rhemtulla, Brosseau-Liard, & Savalei, 2012; Tarka, 2017). Thus, we modelled the CHS-S categorically using the weighted least squares mean and variance adjusted (WLSMV) estimator and the theta parameterization in *MPlus*. The fit was found to be quite good (see Table 2.4 for factor loadings and Table 2.7 for fit statistics), and that model was selected for use in the structural analysis. Loadings on the Positive and Negative factors were strong and statistically significant, with standardized loadings ranging from $\lambda = .540$ -.793 for the Positive factor and $\lambda = .704$ -.859 for the Negative factor. The total factor loadings were more variable, with one negative and one weak loading from items that also loaded onto the Negative factor (item 12 $\lambda = -.263$, item 11 $\lambda = .212$). Because these items loaded very well onto the negative factor and because the overall fit was quite high, the model suggested by Srisarajivakul et al. was not altered. The reliability of all three factors ranged from good to excellent ($\omega_s = .86$ for the general factor, .92 for the positive factor, and .82 for the negative factor).

SDQ. A variety of models were fit to the SDQ based on prior research. The first model was Goodman's (1997) originally proposed 5-factor structure of the self-report measure, which has shown mixed results with a variety of populations (see Caci, Morin & Train (2015) for a

review). The 5-factor model resulted in a poor fit when used with the current sample's data (see Table 2.7 for full details). Caci et al.'s (2015) study found with a large sample of European adolescents that the best fitting model was a bifactor model with the five originally proposed specific factors and two general factors that account for the externalizing and internalizing items (the prosocial scale was left as its own specific factor). This model was also tested, with a poor overall fit, though modification indices suggested that allowing some of the specific factors to correlate would greatly improve the model fit. These very high intercorrelations between specific factors and general factors suggested that a second-order model might fit the data better. This model, with the Externalizing Behavior and Internalizing Behavior factors loading onto a Total Difficulties factor, was also found to have acceptable fit by Caci et al (2015) and others (e.g., Nicolaisen, Skovgaard, Andersen, Sømhøvd, & Obel, 2013). However, the second-order model was a poor fit to the current sample's data.

Alternative structures that were better fits for racial minority adolescent populations in the United States have been previously suggested (Ruchkin, Jones, Vermeien, & Schwab-Stone, 2008; Twyford et al., 2018). Ruchkin, Jones, Vermeien, & Schwab-Stone (2008) found that a three-factor model with two difficulties factors (Emotional Distress/Withdrawal and Behavior Reactivity/Conduct Problems) and one strengths factor (Prosocial Behavior/Peer Competence) was a good fit to a sample of adolescents from urban, racial minority backgrounds. This model fit slightly better than the previous models but did not meet the criteria for adequate fit with the current sample. Twyford, Buckley, Moffa, & Dowdy (2018) tested 3 previously suggested structures with a large sample of Latinx adolescent populations and found that none of them had adequate fit. Twyford et al. then conducted an exploratory factor analysis and suggested a two-factor model (one 'difficulties' and one 'strengths' factor) had the most meaningful and useful structure

in their sample. This 2-factor structure was also tested with the current sample but yielded a poor fit to the data. One additional model was tested that was originally suggested by Goodman, Lamping, & Ploubidis (2010), which is a simplified version of the bifactor model with one 10-item externalizing factor, one 10-item internalizing factor, and one 5-item prosocial factor. The item loadings in this model were the strongest and most consistent, but the overall fit was still quite poor. Exploratory structural equation modelling was then conducted to find a factor structure that worked adequately with the current sample. Results suggested a 4-, 5-, or 6-factor model would adequately fit the data. However, within all of these models, indicators loaded significantly onto more than one factor and made it impossible to determine a coherent structure that would provide useful predictor variables.

In an attempt to identify a set of items that would accurately measure a recognizable construct, each of the 5 subscales of items from Goodman's models were run in separate CFAs. It was found that three of the five subscales had very good fit (Conduct Problems CFI = .99, RMSEA = .05; Emotional Problems CFI = .99, RMSEA = .05; Prosocial CFI = .99, RMSEA = .04), and two had poor fit (Hyperactivity CFI = .88, RMSEA = .20; Peer Problems CFI = .82, RMSEA = .12). The indicators from the Conduct Problems, Emotional Problems, and Prosocial scales were then run in a combined, three-factor CFA that also had poor fit. Modification indices suggested this was due to multiple cross loadings between problematic items. The content of the items was examined, and the research team decided to use only the subscale Conduct Problems in further analysis. This decision was made because previous research suggests that externalizing behavior is a stronger predictor of TSRQ than internalizing behaviors or prosocial skills (Lei et al., 2016; Zhang & Sun, 2011), and because the indicators on the Conduct Problems subscale

describes behaviors that are overt (e.g., stealing, fighting, cheating) and thus less likely to be misinterpreted.

All standardized loadings and residual values for the Conduct Problems CFA can be found in Table 2.5. Item loadings were strong ($\lambda = .62 - .77$), with the exception of one moderate loading for item 7 (“I usually do as I am told”; $\lambda = .49$). McDonald’s omega for Conduct Problems was .83, indicating good overall reliability. The factor scores from this subscale were generated in *MPlus*, saved, and used in further analysis.

IT-SR. The previous three-factor model determined by Murray & Zvoch (2011) was fit to the data, and this model resulted in an acceptable but weak fit on several indices, however, the RMSEA value was elevated (RMSEA = .09; see Table 2.7 for full model fit indices). Modification indices suggested item 9 “I tell my teacher about my problems and troubles,” was a good indicator of all three factors (Communication, Trust and Alienation), though the Murray & Zvoch model included item 9 as part of the Communication factor. An alternative model identical to the previous model but without item 9 was tested, resulting in an increased CFI and reduced RMSEA. Thus, item 9 was excluded from the final IT-SR model used in structural analysis. Item loadings and residual values for the final IT-SR model can be found in Table 2.6. Standardized item loadings on each factor were very strong, ranging from $\lambda = .81$ to $.90$ for Communication, $\lambda = .82$ to $.93$ for Trust, and $\lambda = .70$ to $.88$ for Alienation. The reliability of each factor was also quite high ($\omega = .95$ for Trust, $\omega = .95$ for Communication, and $\omega = .88$ for Alienation). The correlation between Trust and Communication was very high ($r = .87$), so, to determine if there was discriminant validity between the two factors, another model was tested with the items from Trust and Communication combined onto one positive TSRQ factor. The RMSEA for this model was very elevated, so an Exploratory Factor Analysis (EFA) was run on just these items to

determine if a more parsimonious structure was possible. However, the EFA supported the original two-factor structure of the Trust and Communication items. Thus, the three-factor ITSR model without item 9 was selected for further analysis. Consistent with the findings during the raw data screening, the correlations between Alienation and the other two factors were significant and positive ($r = .31$ for Alienation and Communication, and $r = .15$ for Alienation and Trust).

Invariance. The final measurement model for each scale was assessed for measurement invariance between demographic groups. We were specifically interested in whether each measure met metric invariance (whether the factor loadings were similar enough between groups that they could be constrained without significantly worsening model fit) because the factor scores used in the structural model were calculated using the factor means of each responder. The presence of invariance would support the interpretation of any differences between demographic subgroups in later modeling (including the analysis done to answer research question 3), as it provides evidence that the differences are most likely true differences in the population and not due to measurement error.

For the CHS-S, configural, metric, and scalar invariance were supported for student gender, race (Black vs. White), and for the two most commonly selected cultural identities (race/ethnicity vs. nationality). This suggests that teacher cultural humility was understood similarly between students from different gender and racial backgrounds and between students who placed importance on different aspects of their cultural backgrounds. Configural, partial metric, and scalar invariance was supported between students who had similar and different racial backgrounds as their teacher. The statistics used for model comparisons for each scale are found in Table 2.8. It is notable that the *MPlus* chi-square difference test only supported invariance between a few of

the subgroups on the CHS-S. Otherwise, conclusions of invariance were drawn based on a lack of meaningful model change following the example of Caci et al (2015). Any results of the later structural model comparing demographic groups should be interpreted with caution, as differences may reflect differences in measurement instead of true differences in the latent variables.

For Conduct Problems, configural, metric, and scalar invariance was supported for the self-selected cultural identity subgroups and the racial match/nonmatch subgroups. For race subgroups, configural and metric invariance was supported. For gender subgroups, configural invariance was supported. Partial metric invariance was supported for gender after the mean of item 7 (“I usually do as I am told”) was allowed to vary freely between boys and girls. Neither scalar nor partial scalar invariance were supported between race or gender subgroups on the conduct problems subscale. This means that these subgroups did not rate each item with the same intensity. For example, boys may have endorsed lower or higher selections on certain items more frequently than girls. Ultimately, it suggests that there is a difference between boys and girls and between Black and White students in how they interpret and respond to questions on the Conduct Problems subscale. However, as metric and partial metric invariance were supported between these subgroups, and because factor scores are calculated from item means and not thresholds, this amount of noninvariance was thought to have little impact on the interpretation of later findings from the structural model.

For the IT-SR, configural invariance was supported for all groups, and metric invariance was supported for student gender and self-selected cultural identity groupings. Partial metric invariance was supported between race groupings when two item means were allowed to vary freely on the Trust factor (item 1 “my teacher respects my feelings,” and item 2 “my teacher accepts me as I am”). This indicates that Black and White students answered these two questions

differently and could suggest that the interpretation of these questions is inconsistent across racial subgroups. Partial metric invariance was also supported between the matched and non-matched teacher-student race groups when two items were freed on the communication factor (item 10 “my teacher encourages me to talk about my difficulties,” and item 11 “my teacher understands me”). This finding indicates that students answered these two questions differently depending on whether their reported race matched their teachers’ reported race. This could suggest that the underlying constructs represented in these two questions are inconsistent between demographic subgroupings. However, this presence of noninvariance was relatively small (Dimitrov, 2010), and thus it was not thought to impact the conclusions drawn from later structural models. Scalar invariance was supported for all demographic subgroups on the IT-SR.

Structural modeling

Factor scores were calculated in *MPlus*, saved, and then combined into a new dataset so that each individual respondent had a factor score for each latent variable determined by the measurement models. To answer research question one, the correlations between the factor scores were examined. The correlation between the Communication and Trust subscales of the IT-SR were moderately, positively and statistically significantly correlated with the Total Cultural Humility (CH) and Positive CH scores ($r_s = .38, .50$ and $.62, .66$, respectively). Consistent with the findings in the measurement model, the scores for IT-SR subscales Communication and Trust were so highly correlated that these constructs were essentially the same ($r = .92$). Total CH was significantly and negatively correlated with Conduct Problems ($r = -.15$), suggesting that more CH is related to lower problematic behavioral symptoms. Alienation was problematically and positively correlated with Trust and Communication. The full correlation matrix can be found in Table 2.9.

To answer research question 2 and to determine whether cultural humility moderated the relationship between student emotional adjustment and TSRQ, each subscale of the IT-SR was regressed on the Total CH scores, Conduct Problem scores, and the interaction term between CH and Conduct Problems. See Table 2.10 for a full report of moderation effects. Figure 2 depicts the simple slopes between the minimum and maximum point values of each interaction for each subscale of TSRQ.

The results suggested a statistically significant relationship between cultural humility and Conduct Problems on TSRQ. Total CH was a statistically significant, positive predictor of Trust and Communication, but not of Alienation. Conduct Problems predicted significantly lower Trust and significantly higher Alienation but was not a significant predictor of Communication. Additionally, Total CH positively and significantly moderated the relationship between Trust and Conduct Problems, suggesting that teacher cultural humility increases trust between teachers and students with conduct problems ($p < .05$, $R^2 = .26$). Total CH was not a significant moderator between Conduct Problems and Communication ($p > .05$). Unexpectedly, Total CH also positively and significantly moderated the relationship between Conduct Problems and Alienation, suggesting that students with more conduct problems feel more Alienation from teachers with higher Cultural Humility ($p < .001$; $R^2 = .10$). This finding could be due to the strange performance of the Alienation subscale in the current sample, but suggests more investigation is necessary about the role of teacher cultural humility and student feelings of alienation.

Multigroup SEM. To answer research question 3, model tests on each structural model were run to determine if the moderation effects of cultural humility were different between demographic subgroups. Overall, the Wald chi-square tests between models suggested few meaningful differences for each moderating effect between most demographic subgroupings (see

Table 2.11 for a full report of differences). There were no significant differences in the moderation effect comparisons between black and white students, students with differing salient cultural identities, or between students who did and did not have a racial match with teacher. However, a significant difference in the moderation effect was found between gender subgroups in both the Trust and Communication models (see Figures 2 and 3 for the simple slopes between the minimum and maximum values for each group). This suggests that student gender influences the moderation effect of Total CH between Conduct Problems and Trust as well as between Conduct Problems and Communication. Specifically, results indicated that the moderating effect of cultural humility was not significant for girls in either model, but it was significant for boys in both models. In the Trust model, boys with high conduct problems rated teacher cultural humility as extremely influential on their trust with teachers (moderator $\beta = .144, p < .001$). This relationship was so strong that results suggest that boys with high conduct problems have better trust with culturally humble teachers than boys with low conduct problems. Similarly, in the Communication model, boys with high conduct problems rated teacher cultural humility as a significantly positive influence on their level of communication with teachers (moderator $\beta = .134, p < .01$). Again, this relationship was so strong that boys with high conduct problems reported better communication with culturally humble teachers than boys with low conduct problems. Thus, gender was found to play an important role in the moderating effect of cultural humility on TSRQ.

Discussion

This is the first quantitative study of cultural humility in the K-12 educational context, and it is the first application of the CHS-S with high school students. The purpose of this study was to explore the role of cultural humility in teacher-student relationship quality. More specifically, this study intended to determine if cultural humility is related to teacher-student

relationship quality, if cultural humility moderates the relationships between student emotional adjustment and TSRQ, and if demographic factors influence the magnitude of the moderating effects. Because cultural humility has been shown to be important in other helping fields (e.g. Davis et al., 2018; Foronda et al., 2016), this study contributes unique and important findings about the role of teachers' cultural orientation in classrooms and has implications for the application of a new framework of relationship-focused culturally responsive teacher practice.

The bifactor model of the cultural humility scale proposed by Srisarajivakul et al. (2019) was supported in our sample. As the CHS-S is the first adaptation of the Cultural Humility Scale for adolescents in an educational context, it is significant that the structure of the measure was well supported in our sample. Notably, invariance testing suggested that the construct of cultural humility was measured with consistency between subgroups. This preliminarily suggests that the CHS-S is a valid and reliable scale for use by adolescents rating their teachers' levels of cultural humility. It is important for future research to continue to explore the significance of student-perceived cultural humility in classroom and school dynamics, and to determine its convergent validity with other approaches to measurement of cultural humility, such as observational and/or self-report measures.

Regarding research question one, our results suggested that cultural humility is significantly related to TSRQ. This is a significant finding, because TSRQ is predictive of long-term student academic and behavioral outcomes (Roorda et al., 2017). Researchers have demonstrated that minority students from lower socioeconomic backgrounds benefit most from high quality, supportive teacher-student relationships (McGrath & Van Bergen, 2015). Any contribution to understanding the mechanisms of building successful TSRQ can also help to support students from these vulnerable groups, as it provides information to facilitate teachers who work with

these students on a regular basis. Additionally, previous literature has lacked an in-depth exploration of teacher factors that influence TSRQ (Saft & Pianta, 2001). Though this study contributed to the empirical evidence that student conduct does predict TSRQ, it also provides further evidence that relationship quality is not determined by student characteristics alone. Instead, teachers share in the responsibility for building positive relationships with students. Future work should continue to explore the role that other teacher characteristics, especially those related to multicultural orientation, play in TSRQ.

Further, our results for research question two demonstrated that cultural humility is a statistically significant moderator between students' externalizing behavior and TSRQ. Previous research has shown that students' levels of externalizing behavior problems is predictive of TSRQ (e.g., Lei et al., 2016). This study confirms those findings and adds to the growing field of research about the role of multicultural orientation in these relationship dynamics. Specifically, our results demonstrate that students' perceptions of teachers' cultural behavior influence their perception of relationship quality with that teacher. Of particular importance are the findings from this study that indicate students with high conduct problems, whom research predicts should have poor relationships with teachers (e.g., McGrath & Van Bergen, 2015), report significantly higher levels of trust with culturally humble teachers. This shows that teachers who are open and sensitive to student cultural differences can build significantly better relationships with students who have conduct problems.

To our knowledge, this is the first study to quantitatively evaluate students' perspectives of any element of their teachers' multicultural orientation in the context of their relationship with that teacher. While studies have found teachers' perspectives of TSRQ are different along cultural lines (e.g., Split & Hughes, 2015; Thijs et al., 2012) and that students perceive their

teachers differently based on race (Cherng & Halpin, 2016), our study is the first to provide quantitative evidence that a teacher's cultural behavior influences TSRQ. This supports our hypothesis that cultural humility in classrooms functions in a similar way to cultural humility in therapeutic relationships, and this study adds to a growing body of literature about the role of humility in effective relationships in general (e.g., Davis et al., 2018).

Though some of our results for research question two were as we predicted, some of the findings regarding student feelings of alienation were unexpected. Specifically, the results suggested student feelings of alienation were higher with teachers who had high levels of cultural humility. It is possible that students with high conduct problems are prone to feeling alienated regardless of teacher characteristics, or that the construct of cultural humility was interpreted differently due to the experience of students who also had high ratings on the conduct problems scale. It is also possible that this unexpected finding is related to the Alienation variable in general, which was oddly correlated with Trust and Communication in this sample. Replication of this study with a new sample might help explore these topics in more depth and provide better explanations for these unexpected results.

Interestingly, and in conflict with our hypothesis for research question 3, the magnitude of the moderating effect of cultural humility only varied between gender subgroups, not between racial subgroups, self-selected cultural identity subgroups, or racial match/nonmatch subgroups. This is unexpected because other studies (e.g., Cherng & Halpin, 2016) have shown that demographic subgroups of students do differ in their perceptions of teachers. Additionally, and in conflict with our findings about racial match/nonmatch subgroupings, studies of cultural humility in psychotherapy with adults have suggested that the racial background of practitioners does impact the relationship between cultural humility and working alliance (Davis et al., 2018). More

research should assess the role of student demographic subgroups in concert with teacher cultural humility to discern which student groups might benefit most from teachers with high cultural humility.

The magnitude of the cultural humility moderation effect differed significantly between the levels of Trust and Communication reported by girls and boys. For both Trust and Communication, the moderating effect of Total CH was only significant for boys. This effect was very strong, suggesting that for boys with conduct problems, teacher cultural humility is incredibly important for developing strong positive relationships. This is meaningful, as boys with conduct problems are predicted to have very poor relationships with teachers (Jerome et al., 2009; Matthews et al., 2010). These students are much more likely to be recipients of exclusionary discipline, to be found eligible for special education, and to drop out of school (Jerome et al., 2009). This study shows that teacher cultural humility can potentially prevent these negative consequences. Research has shown that one strong, positive relationship with an adult can change the trajectory of such students' school experiences (Hamre & Pianta, 2001; Split, et al., 2012), and it is thus imperative that teacher cultural humility continue to be explored as a way to improve educational inequities for students who are marginalized by the education system.

A problematic finding was in the measurement model of the SDQ. A substantial amount of research has been conducted on the factor structure of the SDQ self-report with adolescents (see Caci et al., 2015 for a review), and with Western samples other than US students the 5-factor model has been found to be a good fit. Similar to other recent work exploring the SDQ with minority American samples (e.g., Twyford et al., 2018), no previously suggested models were an acceptable fit to the data. In particular, the subscales for peer problems and hyperactivity were both very poor fits with the current sample. Additionally, a large amount of cross loading items

suggests that even the factors that work well are still not as discriminating as Goodman's (1997) original model suggests. Future researchers should examine the wording of such items and clarify it in a way that is relevant and norm-supported in minority US adolescent samples. Those who are currently using the SDQ in research or practice in the US should be made aware of these findings and other more valid measures should be used, particularly in work with minority populations.

Finally, this study confirmed previous findings that the IT-SR is a reliable measure of TSRQ with diverse student groups (Murray & Zvoch, 2011), as we utilized the same factor structure as previous studies with the exception of excluding item 9. The wording of item 9, "I talk to my teacher about my problems and troubles," could be the reason for its loading on factors in addition to communication, as students who feel alienated from teachers or students who do not trust their teachers would likely not share their struggles with them. However, the trends of this item loading could also be sample-specific, especially because other characteristics of the structure such as the positive correlation between Alienation and the other two factors and the very high correlation between Trust and Communication are different from the findings of previous researchers as well (e.g., Murray et al., 2016). The study also revealed some small amounts of noninvariance that could be contributing to these unexpected findings. More research using the IT-SR would help clarify if these findings are sample-specific and/or illuminate the reason for the variability in its performance. Overall, the performance of the IT-SR in this study confirms it is a reliable and consistent measure of TSRQ in diverse adolescent samples.

Limitations and Future Directions

This study has several limitations that provide opportunities for the expansion of this critical work. The first is that the data were cross-sectional, preventing any conclusions about the

potential mediating effects of cultural humility that have been suggested by other studies (e.g., Hook et al., 2013). Additionally, the research design of keeping the students and teachers anonymous makes it difficult to conduct a follow-up assessment of these measures at a later date, as it would be nearly impossible to guarantee that the students would report data about the same teacher. Thus, future research should explore the possibility of collecting longitudinal data by adjusting the research design to make conclusions about the mediating role of cultural humility in TSRQ. Future research should also consider the use of multi-level modelling to assess how/if teacher cultural humility has impacts at the classroom level. Previous research has shown that aggregate therapist cultural humility has a stronger effect on therapy outcomes than cultural humility assessed at the individual level (Owen, Tao, Drinane, Hook, Davis, & Kune, 2016). It is important to assess if this finding holds true for teachers as well.

Additionally, though the sample in this study was relatively large and racially/socioeconomically diverse, it was limited to one school district in the Southeastern United States. The racial groups in this region were primarily black and white, with a large number of families living in lower socioeconomic neighborhoods. It is likely that the diversity of our sample facilitates exploration of the topics and constructs in this study, however the racial and socioeconomic groupings are somewhat limited. As a result, it is possible that performing the same analyses with other samples may produce different and equally important results. Further, due to the limitations in the factor structure of the SDQ, we only studied one predictor of TSRQ (students' conduct problems); there are other predictors, including internalizing behavior dimensions or prosocial skills, that might yield more information about the moderating role of cultural humility in TSRQ. In addition to these other dimensions of student emotional adjustment, there are other potential predictors of TSRQ that were not utilized in this study (e.g., student engagement, academic

achievement, teacher experience). Future studies with different samples and varying predictors of TSRQ should be conducted in order to help develop a robust understanding of the role of teacher cultural humility in TSRQ and other classroom dynamics.

Though the current study further demonstrates that the CHS-S is statistically reliable, the convergent validity of student-reported teacher cultural humility is unknown. Thus, examining teacher cultural humility from perspectives of other raters, such as observations or self-report, will help develop a better understanding of how cultural humility functions in schools. Future studies should compare student ratings of cultural humility with other measures of cultural humility. In addition to more studies of the role of cultural humility in other classroom and school dynamics, future research related to TSRQ should consider the importance of teachers' multicultural orientation in their interactions with students. It is possible that specific attention to teachers' cultural attitudes and actions could be helpful in training/coaching teachers to have more effective and nurturing relationships with students. Research evaluating these questions could potentially inform the use of cultural humility in professional development for teachers in classrooms and schools with diverse student bodies.

References

- Armsden, G. C., & Greenberg, M. T. (1987). The Inventory of Parent and Peer Attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence, 16*, 427-454.
- Asparouhov, T., & Muthen, B. (2006). *Robust chi square difference testing with mean and variance adjusted test statistics*. Retrieved from:
[http:// www.statmodel.com/download/webnotes/webnote10.pdf](http://www.statmodel.com/download/webnotes/webnote10.pdf)
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. doi:10.1037/0033-2909.88.3.588
- Caci, H., Morin, A. J., & Tran, A. (2015). Investigation of a bifactor model of the Strength and Difficulties Questionnaire. *European Child and Adolescent Psychiatry, 24*(10), 1291-1301. doi: 10.1007/s00787-015-0679-3.
- Center for Leadership in Disability (2016). *Strengths and Difficulties Questionnaire Auto-scoring Macro*. Georgia State University. URL: <https://disability.publichealth.gsu.edu/contact/>
- Chang, E., Simon, M., & Dong, X. (2012). Integrating cultural humility into health care professional education and training. *Advances in Health Sciences Education: Theory and Practice, 17*(2), 269–278. <https://doi.org/10.1007/s10459-010-9264-1>
- Cherng, H. S., Halpin, P. F. (2016). The importance of minority teachers: Student perceptions of minority versus white teachers. *Educational Researcher, 45*(7), 407-420.
- Cholewa, B., Goodman, R. D., West-Olatunji, C., & Amatea, E. (2014). A qualitative examination of the impact of culturally responsive educational practices on the psychological

- well-being of students of color. *Urban Review*, 46, 574-596. doi: 10.1007/s11256-014-0272-y.
- Choe, E., Srisarajivakul, E., Davis, D. E., DeBlaere, C., Van Tongeren, D., Hook, J. N. (2018). Predicting attitudes towards lesbians and gay men: The effects of social conservatism, religious orientation, and cultural humility. Submitted to *Journal of Psychology and Theology*, in press.
- Davis, D. E., DeBlaere, C., Owen, J., Hook, J. N., Rivera, D. P., Choe, E., Van Tongeren, D. R., Worthington, E. L., & Placeres, V. (2018). The multicultural orientation framework: A narrative review. *Psychotherapy*, 55(1), 89-100. <http://dx.doi.org/10.1037/pst0000160>.
- Davis, D. E., Worthington, E. L., & Hook, J. N. (2010). Humility: Review of measurement strategies and conceptualization as personality judgment. *The Journal of Positive Psychology*, 5(4), 243-252. doi: 10.1080/17439761003791672.
- Davis, D. E., Worthington, E. L., Hook, J. N., Emmons, R. A., Hill, P. C., Bollinger, R. A., & Van Tongeren, D. R. (2013). Humility and the development and repair of social bonds: Two longitudinal studies. *Self & Identity*, 12(1), 58-77. <https://doi.org/10.1080/15298868.2011.636509>
- Davis, H. A. (2006). Exploring the contexts of relationship quality between middle school students and teachers. *The Elementary School Journal*, 106(3), 193-223. doi: 10.1086/501483
- Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry*, 43(9).

- Dimitrov, D. M. (2010). Testing factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling Development*, 43(2), 121-149. doi: 10.1177/0748175610373459
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412. doi:10.1111/bjop.12046
- Fisher-Borne, M., Montana Cain, J., & Martin, S. L. (2015). From mastery to accountability: cultural humility as an alternative to cultural competence. *Social Work Education*, 34(2), 165–181. <https://doi.org/10.1080/02615479.2014.977244>
- Follman, J. (1992). Secondary school students' ratings of teacher effectiveness. *The High School Journal*, 75(3), 168–178.
- Foronda, C., Baptiste, D., Reinholdt, M. M., & Ousman, K. (2016). Cultural humility: A concept analysis. *Journal of Transcultural Nursing*, 27(3), 210-217. doi: 10.1177/1043659615592677
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the strengths and difficulties questionnaire (SDQ): Data from British parents, teachers, and children. *Journal of Abnormal Child Psychology*, 38, 1179-1191. doi: 10.1007/s10802-010-9434-x
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- Gregory, A., Allen, J., Mikami, A., Hafen, C., & Pianta, R. (2015). The promise of a teacher professional development program in reducing racial disparity in classroom exclusionary

- discipline. In D. J. Losen (Ed.), *Closing the discipline gap: Equitable remedies for excessive exclusion* (pp. 166–179). New York, NY: Teachers College Press.
- 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.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*(2), 625-638.
doi:10.1111/1467-8624.00301
- Hamre, B. K., & Pianta, R. C., (2006). Student-Teacher Relationships. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 59-71). Washington, DC, US: National Association of School Psychologists.
- Hook, J. N., Davis, D. E., Owen, J., Worthington, E. L. J., & Utsey, S. O. (2013). Cultural humility: Measuring openness to culturally diverse clients. *Journal of Counseling Psychology, 60*(3), 353–366. <https://doi.org/10.1037/a0032595>
- Hook, J. N., & Watkins, C. E. J. (2015). Cultural humility: The cornerstone of positive contact with culturally different individuals and groups? *American Psychologist, 70*(7), 661–662.
<https://doi.org/10.1037/a0038965>.
- Hughes, J. N. (2011). Longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic adjustment. *The Elementary School Journal, 112*(1), 38-60.
- IBM Corp. Released 2017. *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.

- Jerome, E. M., Hamre, B. K., & Pianta, R. C. (2009). Teacher-child relationships from kindergarten to sixth grade: Early childhood predictors of teacher-perceived conflict and closeness. *Social Development, 18*(4), 915-945. doi:10.1111/j.1467-9507.2008.00508.x.
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103–129.
doi:10.1016/j.jrp.2004.09.009
- Jordan, J. V. (2008). Recent developments in Relational-Cultural Theory. *Women & Therapy, 31*(2-4), 1-4. doi: 10.1080/02703140802145540.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd Ed.) London: The Guilford Press.
- Koziol, N. A., & Bovaird, J. A. (2018). The impact of model parameterization and estimation methods on tests of measurement invariance with ordered polytomous data. *Educational and Psychological Measurement, 78*(2), 272-296. doi: 10.1177/0013164416683754
- Lei, H., Cui, Y., & Chiu, M. M. (2016). Affect teacher-student relationships and students' externalizing behavior problems: A meta-analysis. *Frontiers in Psychology, 7*(1311), 1-12.
doi: 10.3389/fpsyg.2016.01311
- Marian, A., (2016). Cultural humility in multicultural educational context. *Scientific Annals of the "Alexandru Ioan Cuza" University of Iasi: Educational Sciences Series, 20*, 71–80.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.
doi:10.1207/s15328007sem1103_2

- Matthews, J. S., Kizzie, K. T., Rowley, S. J., & Cortina, K. (2010). African Americans and boys: Understanding the literacy gap, tracing academic trajectories, and evaluating the role of learning-related skills. *Journal of Educational Psychology, 102*, 757–771.
- McGrath, K. F., & Van Bergen, P. (2015). Who, when, why and to what end? Students at risk of negative student-teacher relationships and their outcomes. *Educational Research Review, 14*, 1-17. doi.org/10.1016/j.edurev.2014.12.001
- Meade, A. W. & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455. doi: 10.1037/a0028085
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592.
- Miller, J. B. (1976). *Toward a new psychology of women*. Boston: Beacon Press.
- Mosher, D. K., Hook, J. N., Captari, L. E., Davis, D. E., DeBlaere, C., & Owen, J. (2017). Cultural humility: A therapeutic framework for engaging diverse clients. *Practice Innovations, 2*(4), 221-233. http://dx.doi.org/10.1037/pri0000055.
- Morin, A. J. S., Meyer, J. P., Creusier, & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods, 19*(2), 231-254. doi: 10.1177/1094428115621148
- Muris, P., Meesters, C., Marion, M., & Zwambag, L. (2001). Self-reported attachment quality, and symptoms of anxiety and depression in young adolescents. *Personality and Individual Differences, 30*, 809-818.

- Murray, C., Kosty, D., & Hauser-McLean, K. (2016). Social support and attachment to teachers: Relative importance and specificity among low-income children and youth of color. *Journal of Psychoeducational Assessment, 34*(2), 119-135. doi: 10.1177/0734282915592537
- Murray, C. & Zvoch, K. (2011). The inventory of teacher-student relationships: Factor structure, reliability, and validity among African American youth in low-income schools. *Journal of Early Adolescence, 31*(4), 493-525. doi: 10.1177/0272431610366250
- Muthén, B. O. (2004) *Mplus Technical Appendices*. Los Angeles, Muthén and Muthén;
<http://www.statmodel.com/techappen.shtml>
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide (Eighth Edition)*. Los Angeles, CA: Muthén & Muthén.
- Niclasen, J., Skovgaard, A., Andersen, A.-M., Sømhovd, M., & Obel, C. (2013). A Confirmatory Approach to Examining the Factor Structure of the Strengths and Difficulties Questionnaire (SDQ): A Large Scale Cohort Study. *Journal of Abnormal Child Psychology, 41*(3), 355–365. doi: 10.1007/s10802-012-9683-y
- Ortega, R. M., & Coulborn, K. (2011). Training child welfare workers from an intersectional cultural humility perspective: A paradigm shift. *Child Welfare, 90*(5), 27–49.
- Osborne, J. W. (2010). Data Cleaning Basics: Best Practices in Dealing with Extreme Scores. *Newborn and Infant Nursing Reviews, 10*, 37–43. <https://doi-org.ezproxy.gsu.edu/10.1053/j.nainr.2009.12.009>.
- Owen, J., Tao, K. W., Drinane, J. M., Hook, J. N., Davis, D. E., & Kune, N. F. (2016). Client perceptions of therapists' multicultural orientation: Cultural (missed) opportunities and cultural humility. *Professional Psychology: Research and Practice, 47*, 30–37.
<http://dx.doi.org/10.1037/pro0000046>

- Owen, J. J., Tao, K., Leach, M. M., & Rodolfa, E. (2011). Clients' perceptions of their psychotherapists' multicultural orientation. *Psychotherapy, 48*(3), 274–282.
<https://doi.org/10.1037/a0022065>
- Peterson, K., Wahlquist, C., and Bone, K. Student Surveys for School Teacher Evaluation. *Journal of Personnel Evaluation in Education, 14*(2), June 2000, 135-153.
- Pianta, R.C., Hamre, B.K., & Stuhlman, M. (2003). Relationships between teachers and children. In W. Reynolds & G. Miller (Eds.), *Handbook of psychology (Vol. 7) Educational psychology* (pp. 199–234). Hoboken, NJ: John Wiley & Sons.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373. doi: 10.1037/a0029315
- Roorda, D. L., Jak, S., Zee, M., Oort, F. J., & Koomen, H. M. Y. (2017). Affective teacher-student relationships and students' engagement and achievement: A meta-analytic update and test of the mediating role of engagement. *School Psychology Review, 46*(3), 239–261.
<https://doi.org/10.17105/SPR-2017-0035.V46-3>.
- Roorda, D. L., Verschueren, K., Vancraeyveldt, C., Van Craeyevelt, S., & Colpin, H. (2014). Teacher-child relationships and behavioral adjustment: Transactional links for preschool boys at risk. *Journal of School Psychology, 52*, 495-510.
<http://dx.doi.org/10.1016/j.jsp.2014.06.004>
- Ross, L. (2010). Notes from the field: Learning cultural humility through critical incidents and central challenges in community-based participatory research. *Journal of Community Practice, 18*, 315-335. doi:10.1080/10705422.2010.490161

- Rutkowski, L. & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators and fit measure performance. *Applied Measurement in Education*, 30(1), 39-51. <http://dx.doi.org/10.1080/08957347.2016.1243540>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78. doi:10.1037/0003-066x.55.1.68.
- Ruchkin, V., Jones, S., Vermeiren, R., & Schwab-Stone, M. (2008). The Strengths and Difficulties Questionnaire: The self-report version in American urban and suburban youth. *Psychological Assessment*, 20(2), 175-182. doi: 10.1037/1040-3590.20.2.175.
- Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly*, 16(2), 125-141. doi.org/10.1521/scpq.16.2.125.18698.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34(4), 317-342.
- Split, J. L., & Hughes, J. N. (2015). African American children at risk of increasingly conflicted teacher-student relationships in elementary school. *School Psychology Review*, 44(3), 306-314. doi.org/10.17105/spr-14-0033.1
- Split, J. L., Hughes, J. N., Wu, J., & Kwok, O. (2012). Dynamics of teacher-student relationships: Stability and change across elementary school and the influence on children's academic success. *Child Development*, 83(4), 1180-1195. doi: 10.1111/j.1467-8624.2012.01761.x.

- Srisarajivakul, E. N., Choe, E., McPhee, K., Rice, K., Davis, D. E., Meyers, J., & Varjas, K. (2019). Structural analysis and validation of the Cultural Humility Scale for Students among middle school students. Manuscript in preparation.
- Stanhope, V., Solomon, P., Pernell-Arnold, A., Sands, R. G., & Bourjolly, J. N. (2005). Evaluating cultural competence among behavioral health professionals. *Psychiatric Rehabilitation Journal*, 28, 225–233. <http://dx.doi.org/10.2975/28.2005.225.233>.
- Tarka, P. (2017). The comparison of estimation methods on the parameter estimates and fit indices in SEM model under 7-point Likert scale. *Archives of Data Science*, 2(1), 1-16. doi: 10.5445/KSP/1000058749/10
- Tervalon, M. & Murray-Garcia, J. (1998). Cultural humility versus cultural competence: A critical distinction in definition physician training outcomes in multicultural education. *Journal of Health Care for the Poor and Underserved*, 9(2), 117-125. doi: 10.1353/hpu.2010.0233
- Thijs, J., Westhof, S., & Koomen, H. (2012). Ethnic incongruence and the student-teacher relationship: The perspective of ethnic majority teachers. *Journal of School Psychology*, 50, 257-273. doi.org/10.1016/j.jsp.2011.09.004
- Tinkler, A. S., & Tinkler, B. (2016). Enhancing cultural humility through service learning in teacher preparation. *Multicultural Perspectives*, 18(4), 192-201. doi: 10.1080/15210960.2016.1222282.
- Tormala, T. T., Patel, S. G., Soukup, E. E., & Clarke, A. V. (2018). Developing measurable cultural competence and cultural humility: An application of the Cultural Formulation. *Training and Education in Professional Psychology*, 12(1), 54-61. <http://dx.doi.org/10.1037/tep0000183>.

- Twyford, J. M., Buckley, L., Moffa, K., & Dowdy, E. (2018). The Strengths and Difficulties Questionnaire (SDQ): Factor structure of the self-report form in Latinx youth. *International Journal of School & Educational Psychology*, 1-14.
<https://doi.org/10.1080/21683603.2017.1414005>.
- U.S. Department of Education. (2016). *The state of racial diversity in the educator workforce*. Retrieved from: <https://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversity-workforce.pdf>
- Uses of the SDQ. (2018) *Youth in Mind USA*. Retrieved from: https://youthinmindusa.com/knowledge_base/uses-of-the-sdq/
- Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, 49, 1304–1312. doi: 10.1111/j.1469-7610.2008.01942.
- Worrell, F. C., & Kuterbach, L. D. (2001). The use of student ratings of teacher behaviors with academically talented high school students. *Journal of Secondary Gifted Education*, 14(4), 236–247.
- Yoon, M. & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 201-213.
<https://doi.org/10.1080/10705511.2017.1387859>
- Zhang, X., & Sun, J. (2011). The reciprocal relations between teachers' perceptions of children's behavior problems and teacher-child relationships in the first preschool year. *The Journal of Genetic Psychology*, 172(2), 176-198. <https://doi.org/10.1080/00221325.2010.52807>

Table 2.1

Demographics of Participants (N = 1612)

Characteristic	<i>N</i>	%
Gender		
Male	732	45.4
Female	842	52.2
Other/Prefer not to say	38	2.4
Grade		
9	543	33.7
10	400	24.8
11	381	23.6
12	288	17.9
School		
High School 1	827	51.3
High School 2	785	48.7
Race		
American Indian	22	1.4
Asian/Pacific Islander	20	1.2
Black, Non-Hispanic	738	45.8
Latino/a	115	7.1
Multi-Racial	113	7.0
White, Non-Hispanic	604	37.5

Note. Due to the elimination of data from inconsistent responders, the total *N* was reduced from 2002 to 1612.

Table 2.2
Teacher Demographics (As Reported by Students)

Characteristic	<i>N</i>	%
Gender		
Male	748	46.4
Female	830	51.5
Other	13	0.8
Missing	21	1.3
Age		
18-24	35	2.2
25-34	446	27.7
35-44	430	26.7
45-54	184	11.4
55-64	77	4.8
65+	40	2.5
I don't know	365	22.6
Missing	35	2.2
Race		
American Indian	14	0.9
Asian/Pacific Islander	8	0.5
Black, Non-Hispanic	458	28.4
Hispanic	78	4.8
Multi-Racial	5	0.3
White, Non-Hispanic	860	53.3
I don't know	160	9.9
Other	9	0.6
Missing	20	1.2
Class Difficulty		
Easy	1209	75.0
Difficult	395	24.5
Missing	8	0.5

Table 2.3

Participants' Most Salient Cultural Identities

Identity	<i>N</i>	%
Race/Ethnicity	515	31.9
Nationality	321	19.9
Gender	240	14.9
Neighborhood	200	12.4
Language spoken at home	133	8.3
Other	136	8.4
Missing	67	4.2

Table 2.4

Item Means and Factor Loadings of CHS-S Final Model

Factor	Item	<i>M</i>	<i>SD</i>	β	Residual ($1-R^2$)
Positive	1. Shows respect	4.43	1.00	.592	.135
	2. Is open	4.29	1.08	.627	.248
	5. Is considerate	4.04	1.14	.609	.611
	6. Shows an interest in learning more	4.12	1.10	.810	.087
	7. Tries to see my perspective	4.05	1.15	.788	.265
	9. Stays open-minded	4.12	1.09	.752	.236
	12. Asks questions when unsure	3.80	1.22	.703	.445
Negative	4. Is arrogant	3.67	1.39	.701	.282
	8. Makes assumptions	2.99	1.37	.785	.226
	10. Acts like a know-it-all	3.67	1.41	.839	.130
	11. Thinks he/she knows more than he/she does	3.64	1.39	.803	.310
Total	1. Shows respect	4.43	1.00	.717	.135
	2. Is open	4.29	1.08	.599	.248
	5. Is considerate	4.04	1.14	.133	.611
	6. Shows an interest in learning more	4.12	1.10	.507	.087
	7. Tries to see my perspective	4.05	1.15	.338	.265
	9. Stays open-minded	4.12	1.09	.446	.236
	12. Asks questions when unsure	3.80	1.22	-.246	.445
	4. Is arrogant			.476	.282
	8. Makes assumptions	3.67	1.39	.397	.226
	10. Acts like a know-it-all	2.99	1.37	.408	.130
	11. Thinks he/she knows more than he/she does	3.67	1.41	.212	.310

Note. β = Standardized factor loadings. All loadings significant, $p < .05$.

Table 2.5

Item Means and Factor Loadings of Conduct Problems Model

Factor	Item	<i>M</i>	<i>SD</i>	β	Residual ($1-R^2$)
Conduct Problems	5. I get very angry and often lose my temper	0.64	0.73	.621	.467
	7. I usually do as I am told	0.48	0.56	.485	.765
	12. I fight a lot. I can make other people do what I want	0.22	0.49	.766	.413
	18. I am often accused of lying or cheating	0.28	0.57	.740	.452
	22. I take things that are not mine from home, school or elsewhere	0.12	0.39	.757	.427

Note. β = Standardized factor loadings. All loadings significant, $p < .001$.

Table 2.6
Item Means and Factor Loadings of IT-SR Final Model

Factor	Item	<i>M</i>	<i>SD</i>	β	Residual (1- R^2)
Communication	4. My teacher can tell when something is upsetting me	2.87	1.10	.805	.352
	8. My teacher helps me understand myself better	2.78	1.12	.841	.293
	10. My teacher encourages me to talk about my difficulties	2.57	1.17	.817	.332
	11. My teacher understands me	2.78	1.09	.900	.190
	12. When angry, my teacher tries to be understanding	2.70	1.11	.836	.301
	15. I can count on my teacher when I need to get something off my chest	2.54	1.15	.858	.264
	17. If my teacher knows something is bothering me, they ask me about it	2.69	1.13	.833	.306
Trust	1. My teacher respects my feelings	3.35	0.90	.929	.137
	2. My teacher accepts me as I am	3.48	0.84	.925	.144
	3. I feel my teacher is successful as a teacher	3.43	0.88	.909	.174
	7. My teacher trusts my judgement	2.95	0.96	.820	.328
	13. I trust my teacher	3.08	1.03	.895	.199
Alienation	5. I get easily upset at school	2.13	1.14	.878	.229
	6. I get upset a lot more than my teacher knows about	2.23	1.19	.865	.252
	14. My teacher doesn't understand what I'm going through	2.24	1.13	.704	.504
	16. I feel that no one understands me	2.14	1.17	.775	.399

Note. β = Standardized factor loadings. All loadings significant, $p < .05$.

Table 2.7

Fit Statistics for Measurement Models

	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA	90% CI	SRMR
CHS-S							
Bifactor Model	501.31	117	<.001	.991	.068	.061-.076	.029
ITSR							
3-Factor Model full	1574.03	116	<.001	.969	.088	.084-.092	.057
3-Factor Model no item 9	1239.08	101	<.001	.977	.084	.079-.088	.046
2-Factor Model	1992.04	103	<.001	.962	.107	.103-.111	.069
SDQ							
5-Factor Model	3761.05	265	<.001	.762	.090	.088-.093	.108
Bifactor Model	3864.70	252	<.001	.754	.094	.092-.097	.109
Second Order Model	4003.32	268	<.001	.746	.093	.090-.096	.113
Ruchkin et al. Model	2973.63	272	<.001	.819	.078	.076-.081	.094
Goodman 3-Factor Model	4563.72	272	<.001	.712	.099	.096-.101	.119
Twyford et al. 2-Factor Model	3053.76	208	<.001	.791	.092	.089-.095	.100
Conduct Problems Only	21.43	5	<.001	.989	.045	.027-.066	.028

Note. CFI = Comparative fit index; RMSEA = Root mean square error of approximation. 90% CI = confidence interval for RMSEA. SRMR = Standardized root mean square residual.

Table 2.8

Measurement Invariance Model Fit Indices

	χ^2	<i>df</i>	$MD\Delta\chi^2$	Δdf	<i>p</i>	CFI	ΔCFI	RMSEA	90% CI	$\Delta RMSEA$	SRMR
CHS-S											
Boys (n = 732)/Girls (n = 842)											
Configural	379.116	66				.989		.078	.070-.085		.031
Metric	524.585	87	217.936	21	<.001	.985	-.004	.080	.074-.087	.002	.068
Scalar	764.878	132	284.460	44	<.001	.978	-.007	.078	.073-.084	-.002	.070
Black (n = 738)/White (n = 604)											
Configural	323.785	66				.989		.077	.068-.085		.030
Metric	368.415	88	136.874	22	<.001	.988	-.001	.069	.062-.077	-.008	.063
Scalar	465.743	132	135.219	44	<.001	.986	-.002	.062	.056-.068	-.007	.064
Race/Ethnicity (n = 515)/Nationality (n = 321)											
Configural	262.580	66				.987		.084	.074-.095		.032
Metric	154.659	88	33.982	22	.049	.995	.007	.043	.031-.053	-.041	.042
Scalar	199.902	132	54.470	44	.134	.995	.000	.035	.025-.045	-.008	.043
Race Match (n = 692)/ Non-Match (n = 920)											
Configural	441.442	66				.987		.084	.077-.092		.032
Metric	223.652	88	47.851	22	.001	.995	.008	.044	.037-.051	-.040	.041
Partial Metric	236.679	84	28.255	18	.058	.995	.008	.047	.040-.055	-.037	.035
Scalar	262.816	132	85.049	48	.001	.996	.001	.035	.029-.041	-.012	.042
SDQ											
Boys/Girls											
Configural	27.025	10				.988		.047	.026-.068		.034
Metric	76.795	14	38.118	4	<.001	.954	-.034	.075	.059-.092	.028	.060
Partial Metric	26.929	13	3.576	3	.311	.990	.002	.037	.016-.057	-.010	.036
Scalar	170.219	23	130.229	10	<.001	.892	-.098	.090	.078-.103	.053	.052
Partial Scalar	112.403	21	77.528	8	<.001	.933	-.067	.074	.061-.088	.037	.046
Black/White											
Configural	22.934	10				.988		.044	.020-.068		.032
Metric	30.902	14	9.344	4	.053	.985	-.004	.042	.022-.063	-.002	.039

Scalar	108.952	24	74.837	10	<.001	.922	-.063	.073	.059-.087	.031	.047
Partial Scalar	77.316	22	45.913	8	<.001	.949	-.036	.061	.047-.076	.012	.045
Race or Ethnicity/Nationality											
Configural	17.588	10				.989		.043	.000-.075		.037
Metric	28.633	14	10.353	4	.035	.978	-.012	.050	.023-.076	.007	.051
Scalar	42.083	23	22.444	10	.013	.971	-.007	.045	.022-.066	.005	.046
Race Match/Non-Match											
Configural	27.120	10				.989		.046	.026-.067		.031
Metric	23.131	14	2.495	4	.645	.994	.005	.028	.000-.048	.018	.033
Scalar	36.166	24	13.821	10	.181	.992	-.003	.025	.000-.041	.003	.036
IT-SR											
Boys/Girls											
Configural	1317.449	202				.978		.084	.079-.088		.050
Metric	1078.078	218	64.490	16	<.001	.983	.005	.071	.067-.075	-.013	.053
Scalar	1040.152	266	116.297	48	<.001	.984	.001	.061	.057-.065	-.010	.055
Black/White											
Configural	1147.176	202				.978		.084	.079-.088		.050
Metric	875.415	218	28.379	16	.029	.985	.007	.067	.062-.072	-.017	.051
Partial Metric	862.418	216	12.719	14	.549	.985	.000	.067	.062-.071	.000	.050
Scalar	908.391	264	139.692	48	<.001	.985	.000	.060	.065-.065	-.007	.052
Race or Ethnicity/Nationality											
Configural	738.616	202				.978		.080	.074-.086		.053
Metric	603.892	218	25.896	16	.056	.984	.006	.065	.059-.071	-.015	.055
Scalar	583.081	266	79.758	64	.088	.987	.003	.053	.047-.059	-.012	.055
Race Match/Non-Match											
Configural	1323.638	202				.978		.083	.079-.087		.049
Metric	984.948	218	31.801	16	.011	.985	.007	.066	.062-.070	-.017	.050
Partial Metric	1041.030	216	20.592	14	.113	.984	-.001	.069	.065-.073	.003	.049
Scalar	880.819	264	45.068	28	.594	.988	.004	.054	.050-.058	-.015	.049

Note. CFI = Comparative fit index; RMSEA = Root mean square error of approximation. 90% CI = confidence interval for RMSEA. SRMR = Standardized root mean square residual. $MD\Delta\chi^2$ = chi-square difference test in MPlus.

Table 2.9

Correlations of Latent Variables

	1	2	3	4	5	6	7
1. Total CH	1.00	--	--	--	--	--	--
2. Pos CH	.33	1.00	--	--	--	--	--
3. Neg CH	-.11	.13	1.00	--	--	--	--
4. Comm	.38	.62	.10	1.00	--	--	--
5. Trust	.50	.66	-.03 ⁿ	.92	1.00	--	--
6. Alien	-.02 ⁿ	.16	.34	.39	.23	1.00	--
7. Cond Prob	-.15	-.11	.13	-.09	-.17	.14	1.00

Note. Comm = CH = Cultural Humility; Communication; Alien = Alienation; Cond Prob = Conduct Problems.

ⁿindicates non-significance. All other correlations statistically significant $p < .05$

Table 2.10

Interactions and Main Effects on Teacher-Student Relationship Quality

	Model: Trust		Model: Communication		Model: Alienation	
	β	SE	β	SE	β	SE
Total CH	.46***	.02	.37***	.02	.04	.02
Conduct Problems	-.11***	.02	-.01	.03	.33***	.03
Total CH*Cond Prob	.06*	.02	.05	.04	.12***	.03

* $p < .05$; ** $p < .01$; *** $p < .001$

Table 2.11

Wald Chi-Square Values for Subgroup Model Tests

	$\Delta\chi^2$	Δdf	<i>p</i>
Outcome: Trust			
Conduct Problems*Gender	5.60	1	.018
Conduct Problems*Race	0.61	1	.436
Conduct Problems*Race Match	0.90	1	.343
Conduct Problems*Cultural Background	0.07	1	.788
Outcome: Communication			
Conduct Problems*Gender	6.09	1	.014
Conduct Problems*Race	0.25	1	.616
Conduct Problems*Race Match	0.97	1	.326
Conduct Problems*Cultural Background	0.21	1	.651
Outcome: Alienation			
Conduct Problems*Gender	1.16	1	.281
Conduct Problems*Race	0.16	1	.686
Conduct Problems*Race Match	0.17	1	.683
Conduct Problems*Cultural Background	0.02	1	.896

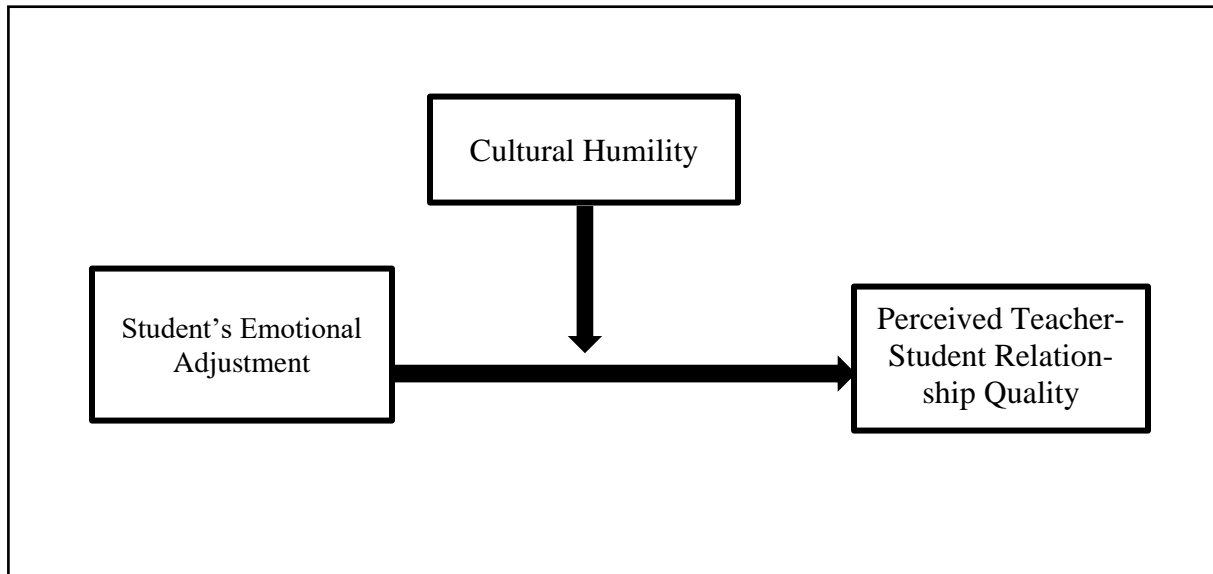
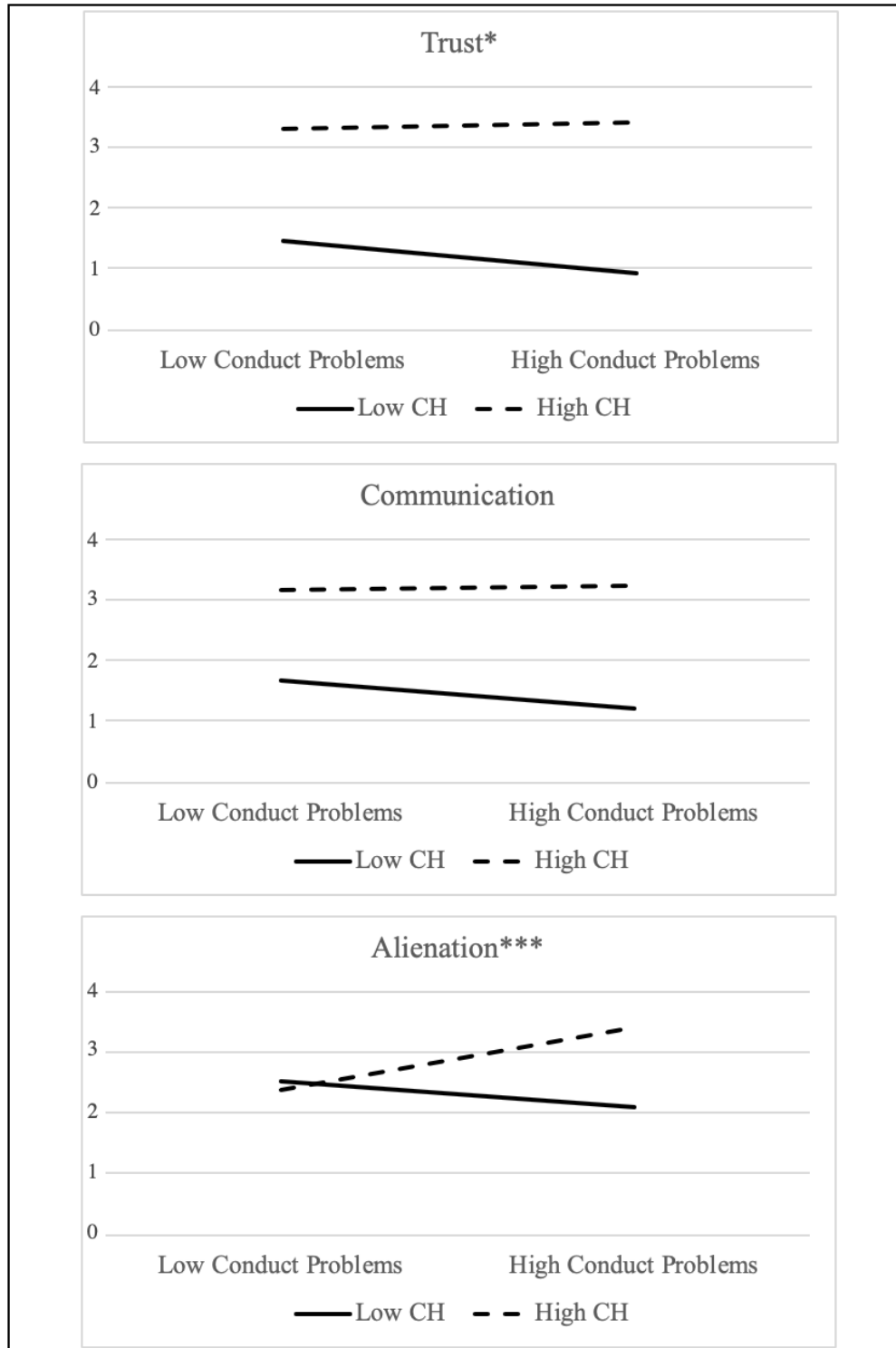
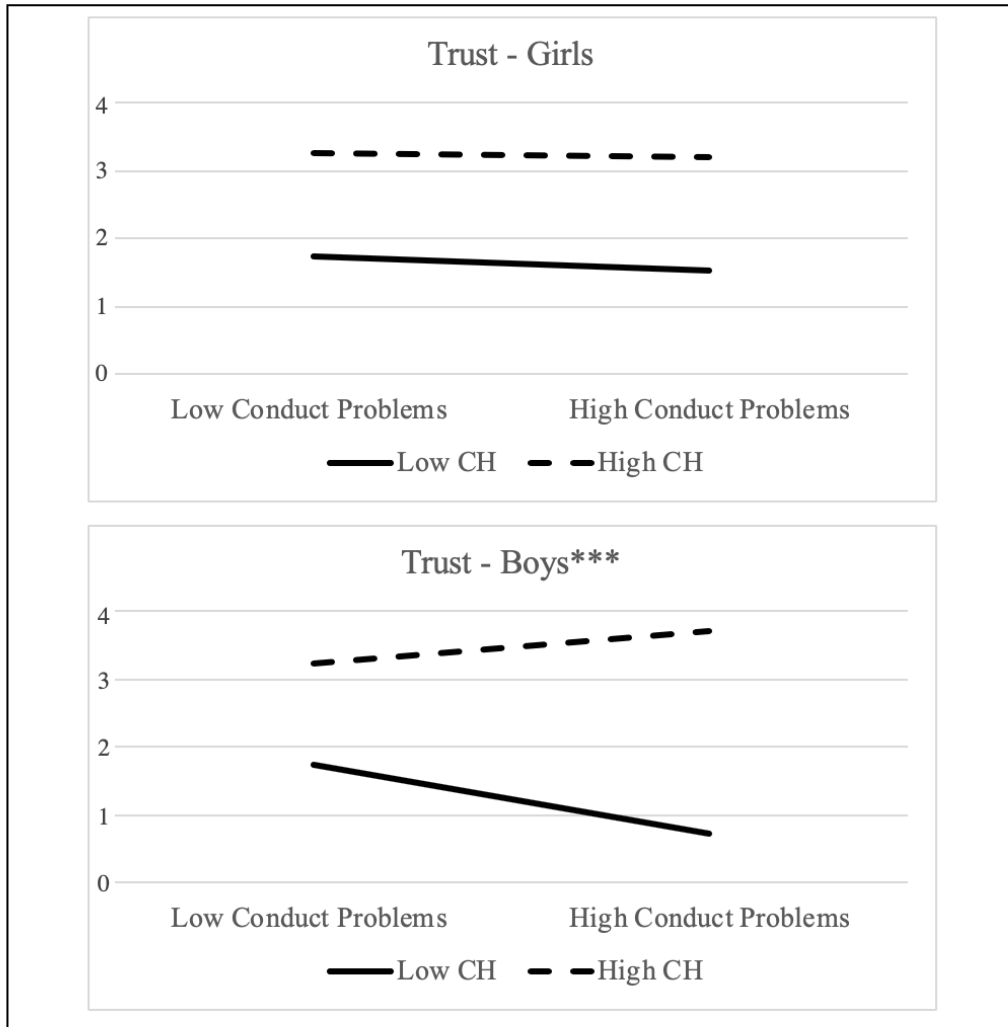


Figure 2.1. Theoretical Moderation Model of Cultural Humility's Role in Teacher-Student Relationship Quality.



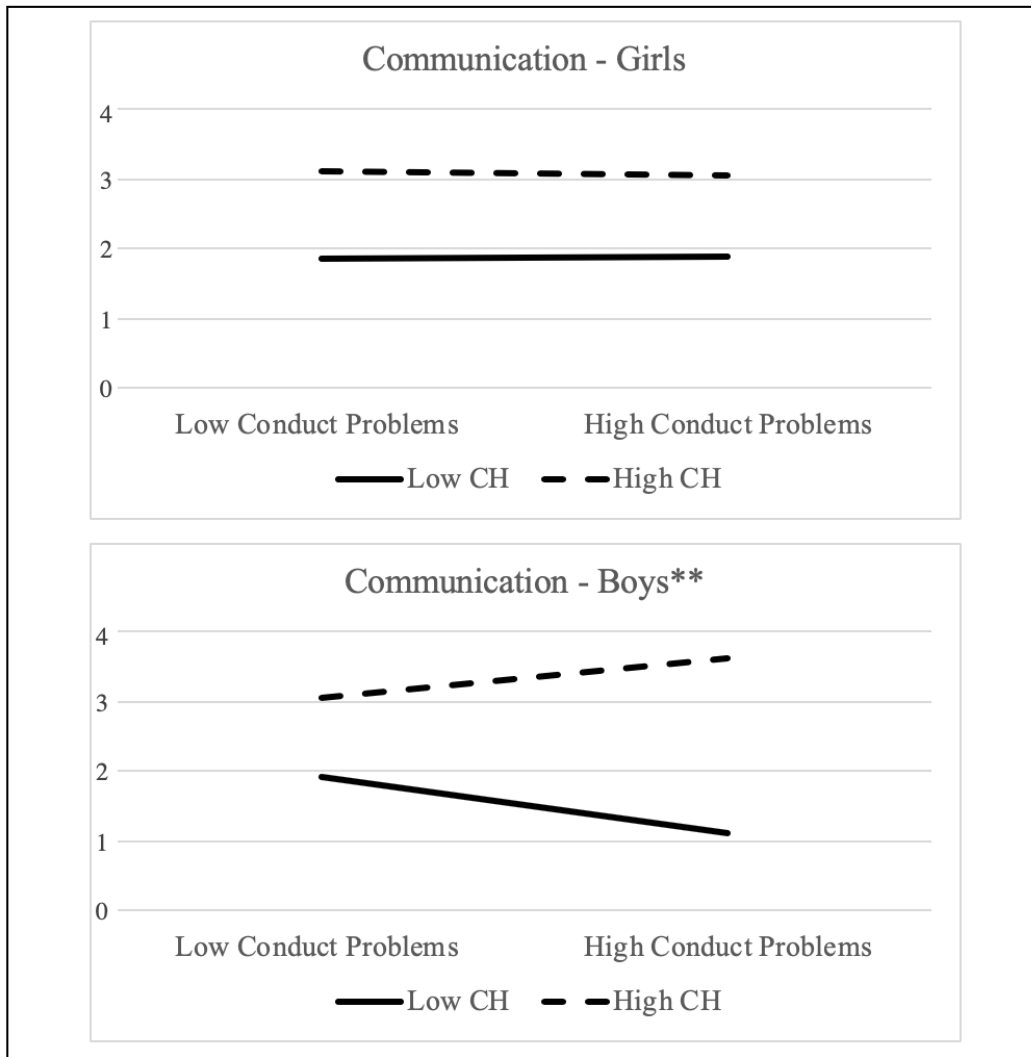
* $p < .05$; ** $p < .01$, *** $p < .001$

Figure 2.2 Interaction of Cultural Humility and Conduct Problems on TSRQ



* $p < .05$; ** $p < .01$, *** $p < .001$

Figure 2.3 Three-way Interaction Between Gender, Conduct Problems, and Cultural Humility on Trust.



* $p < .05$; ** $p < .01$, *** $p < .001$

Figure 2.4. Three-way Interaction Between Gender, Conduct Problems, and Cultural Humility on Communication.

APPENDIX

Cultural Humility Scale for Students (CHS-S)

Some parts of your culture could be your race or your gender. Other parts of your culture could be your nationality or the neighborhood you are from.

Please pick the part of your culture that is most important to you: [drop down menu including race, gender, nationality, neighborhood, language spoken at home, and other, which prompts a write-in option]

Think about you third period teacher. Answer the questions below:

Towards this part of my culture, my teacher...	Really Disagree (1)	Kind of Disagree (2)	Neutral (3)	Kind of Agree (4)	Really Agree (5)
1. Shows respect	1	2	3	4	5
2. Is open	1	2	3	4	5
3. Assumes he/she already knows a lot	1	2	3	4	5
4. Is arrogant	1	2	3	4	5
5. Is considerate	1	2	3	4	5
6. Shows an interest in learning more	1	2	3	4	5
7. Tries to see my perspective	1	2	3	4	5
8. Makes assumptions	1	2	3	4	5
9. Stays open-minded	1	2	3	4	5
10. Acts like a know-it-all	1	2	3	4	5
11. Thinks he/she knows more than he/she does	1	2	3	4	5

12. Asks questions when unsure	1	2	3	4	5
13. Acts like he/she is better than me	1	2	3	4	5
14. It willing to talk about it with me	1	2	3	4	5

This teacher's race is

This teacher's age is

This teacher's gender is

Positive subscale items: 1, 2, 4, 5, 7, 9, 12

Negative subscale items: 3, 6, 8, 10, 11