2014

DEVELOPMENT OF AN INSTRUMENT FOR ASSESSING CULTURALLY CONGRUENT SCIENCE TEACHING

Regina Christine Sievert

DEVELOPMENT OF AN INSTRUMENT FOR ASSESSING CULTURALLY CONGRUENT SCIENCE TEACHING

By

REGINA CHRISTINE SIEVERT

Master of Science in Education, Montana State University - Billings, Billings, MT, 1993
Bachelor of Science in Zoology, The Ohio State University, Columbus, OH, 1984
Associate of Natural Science in Wildlife Management, Hocking College, Nelsonville, OH, 1980

A dissertation study
presented in partial fulfillment of the requirements
for the degree of
Doctor of Education
in Curriculum and Instruction

The University of Montana
Missoula, MT

August 2014

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Dr. David Erickson, Co-Chair
Department of Curriculum and Instruction

Dr. Edith Gummer, Co-Chair
The National Science Foundation

Dr. Joan LaFrance
Mekinak Consulting

Dr. Bharath Sriraman
Department of Mathematical Sciences

Dr. Tammy Elser
Department of Curriculum and Instruction

Sievert, Regina, Ed.D., Summer 2014 Curriculum and Instruction

Development of an Instrument for Assessing Culturally Congruent Science Teaching

Co-Chair: David R. Erickson, Ph.D.

Co-Chair: Edith S. Gummer, Ph.D.

Demographers forecast that ethnic minority students will make up the majority of students in America's K-12 schools sometime in the next few decades. Yet most ethnic minority students continue to experience a lower level of achievement compared to their White peers. Emerging research indicates that culturally congruent instruction (CCI) is correlated with improved ethnic minority student achievement and so may be one means to close the achievement differential. Calls for more research in CCI are increasing, yet measuring CCI is challenging due to its context specific nature and abstract elements that are difficult to define and operationalize. This study responded to the need for improved assessment of CCI through the investigation of two research questions: What is a culturally congruent process for developing a valid instrument for assessing the use of CCI in teaching science with Montana American Indian students? and What is the technical quality of such an instrument? Investigating these questions resulted in (a) a culturally congruent instrument development model that utilized participatory methods and involved numerous and diverse stakeholders, (b) a model of CCI composed of three major elements (content, pedagogy, and environment), (c) a teacher self report survey known as the Revised Culturally Congruent Instruction Survey, and (d) a substantive body of evidence for the use of the instrument to draw valid inferences regarding CCI. While the context specific nature of CCI means that the Revised CCIS will likely require adaptation if used in contexts outside of the one for which it was designed, it holds significance to the research and education community in providing a template for the operationalization of CCI and its assessment. Likewise, the development process model, in demonstrating the use of culturally congruent practices to equitably engage stakeholders in instrument development, has potential value as a resource for guiding those looking to work with communities to develop a similar instrument. Both the instrument and development model have potential to move the research base forward regarding CCI, worthwhile goals that may assist in the attainment of equitable educational outcomes for all students.

This work is dedicated to my mother and father, Mary Ann and Lawrence Charles Long.

Acknowledgements

I would like to acknowledge the hard work of my mentors and colleagues, without whom the completion of this study would not have been possible. First and foremost, I owe a huge debt to Dr. David Erickson, whose perseverance and trust in me made the completion of this study a reality after many years. I consider myself very lucky to have been able to work with him. Next, I would like to acknowledge my friend and mentor Dr. Edith Gummer, whose intelligence and ability to push me to my limits have long been key to my growth in the field of education research. My thanks also go to Dr. Joan LaFrance, another of my most valued friends and mentors, whose support was indispensable in this and many other projects on which we have worked; the list of things she has given me as a professional and a friend is too long to describe here. I also owe a large debt to Dr. Elisabeth Swanson, who put me on this path to the doctorate and whose excellence in science education has enriched so many people in the field, including me. I would also like to thank Dr. Tammy Elser and Dr. Bharath Sriraman, who signed on for my committee when others would not take the risk; their support was invaluable and irreplaceable. My thanks also go out to my good friend Dr. Rodney Brod for the statistical analysis so important to the completion of this study. My long conversations about the mysteries of statistics and other wonders of the world with him over tea and soup challenged and delighted me.

Special thanks go to the tribal people, faculty, and teachers who participated in this study. Their contributions made the study a reality and the relationships that were fostered en route are long and priceless. Chris Strom, my coworker, stalwart supporter, and dear friend throughout this project, is also a part of the special thanks category. Words do not adequately capture my appreciation for her. Finally, I thank my husband David and daughter Grace for the many ways that they helped me through this long journey, including allowing my laptop in our campsite.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE – OVERVIEW OF THE PROBLEM

Ethnic Diversity and Achievement Gaps

The ethnic diversity of the Pre K-12 student population in schools in the United States is increasing, commensurate with the increasing diversity in the country's overall population. Recent figures indicate that 45% of Pre K-12 U.S. students are members of ethnic minorities, a percentage that continues to rise each year (Sable & Plotts, 2010). Conversely, the overwhelming majority of U.S. Pre K-12 teachers are White people of European descent. By 2012 figures, 81.9% of teachers reported that they are non - Hispanic Whites, a number that has increased slightly in recent years (National Center for Education Statistics, 2013b).

With few exceptions, ethnic minority students in U.S. schools are underachieving compared to their White peers on standard measures of achievement ranging from the National Assessment of Educational Progress, to state standardized tests, to high school and college graduation rates, and college entrance exams (ACT, 2012; Aud, Wilkinson-Flicker, Kristapovich, Rathbun, Wang, & Zhang, 2013; Education Trust, 2013; NCES, 2013b; NCES 2013c; Rampey, Dion, & Donahue, 2009). Factors that research has correlated with achievement gaps are many including school based factors such as teacher preparation and experience (Darling - Hammond, 2000), teacher practice (Wenglinsky, 2000), the rigor of the curriculum, and school safety; factors external to school and often associated with socioeconomic status such as student nutrition, enrichment activities, and student mobility (Barton & Coley, 2009; Institute of Medicine and National Research Council, 2010) and those factors that connect school and home, like parent participation in their student's education (Barton & Coley, 2009). Adverse conditions that have been correlated with lower student achievement are disproportionately more commonly experienced by ethnically diverse students (Almy & Theokas, 2010; Annie E. Casey

Foundation, 2006; Chicago Teachers Union, 2012; Lankford, Loeb, & Wyckoff, 2002; Olson, 2003; Schmidt, Cogan, & McKnight, 2010).

Another potential set of factors hypothesized as contributing to the underachievement of ethnically diverse students is rooted in the cultural incongruities that commonly exist between minority students' home cultures and those of their teachers and schools (Barnhardt & Kawagley, 2005; Gay, 2010; Lee & Buxton, 2010; Lipka & Adams, 2007; Parsons 2008; Skinner, 1999). Cultural incongruities encompass a suite of factors such as a lack of curriculum content relevant to ethnically diverse students' lives (Barnhardt & Kawagley, 2005; Gilbert, 2010); incompatibility between the behavioral norms of schools, classrooms and students' home cultures (Boykin, Tyler, Watkins-Lewis, & Kizzie, 2006; Deyhle, 1995; Gay, 2010; Tyler, Uqdah, Dillihunt, Beatty-Hazelbocker, Connor, Gadson, et al., 2008); differences between the language of instruction and students' home language (Lee & Buxton, 2010; Yazzie, 1999); and a disconnect between the pedagogy used in typical American classrooms and the traditional teaching methods familiar to ethnically diverse students (Barnhardt & Kawagely, 2005; Hilberg & Tharp, 2002; Hilberg, Tharp, & Degeest, 2000; Lee, 2002). Many of these scholars have hypothesized that reconciling the cultural incompatibilities between students' home cultures and schools through the use of more culturally congruent instruction (CCI) will improve the academic achievement of ethnically diverse students.

Calls for Equitable Educational Outcomes

Today in the U.S. there is an increasing push for equity in educational opportunity and outcomes for all students, a push that extends across many parts of society and involves a variety of stakeholders, both public and private. The Elementary and Secondary Education Act (ESEA), commonly referred to as No Child Left Behind (NCLB), is a prominent example of federal

policy that has a primary focus on equitable educational outcomes. NCLB has greatly increased the accountability of K-12 schools for closing achievement gaps and imposes sanctions on schools that repeatedly fail to meet achievement benchmarks for all student subgroups. Major federal funding initiatives from the U.S. Department of Education like the *Race to the Top* and *Investing in Innovation* programs prioritize equitable outcomes as requirements for attaining funding. Professional education organizations such as the National Council for Teachers of Mathematics and the National Science Teachers Association have developed policy statements calling for increased efforts for closing achievement gaps (National Council of Teachers of Mathematics [NCTM], 2014; National Science Teachers Association [NSTA], 2000). Institutions and individuals in the research community are also raising the volume on their calls for equity in education for all students (See, for example, Barnhardt & Kawagely, 2005; Cajete, 2005, 1999; Council of Chief State School Officers [CCCSO] 2004; Lee, 2005; Lee & Buxton, 2010; Lipka, Parker Webster, & Yanez, 2005; Mackety & Linder-VanBerschot, 2008). Likewise state, tribal, and local efforts are growing in the equitable outcomes arena. American Indian tribes, for example, are increasingly developing formal education policies and sponsoring programs that aim to close the long standing achievement gap for American Indian students (e.g., Tribal Education Departments National Assembly [TEDNA], 2011; Ute Tribe, 2004).

Culturally Congruent Instruction

The construct of culture is complex and often debated. The exact definition of CCI varies in the literature, but generally it can be defined as instruction that is compatible with and builds upon students' cultures such that it validates students' cultural identities, empowers students, builds on their prior knowledge and traditional ways of knowing, and supports their achievement. Other terms relevant to and sometimes used interchangeably with CCI in the

literature include culturally responsive teaching (CRT), culturally based education (CBE), and culturally relevant teaching (CRT). Scholars have been debating and writing about CCI for decades, refining its definition, delineating its characteristics, and building a case for its importance in supporting student success (See, for example, Au & Jordan, 1981; Cardell, Cross, & Lutz, 1978; Demmert & Towner, 2003; Gay, 2002, 2010; Lee, 2005; Mohatt & Erickson, 1981; Parsons, 2008; Phuntsog, 1999, Skinner, 1999; Siwatu, 2007.).

The manifestations of CCI in the classroom are, by nature, specific to the prioritized cultural context. A people's shared history and everyday experiences, which are the basis for CCI, vary with the cultural context. The norms, accepted social interactions, power dynamics, and traditional teaching methods that are essential to CCI also vary with the cultural context. Boykin & Bailey (2000), for example, indicate that instructional practices compatible with the home cultures of African American students incorporate elements of movement, verve, and communality. Instructional practices compatible with the home cultures of many American Indian students, on the other hand, incorporate elements of private reflection and practice, practical application of knowledge, observational learning, multiple mentors from the extended family and community, spirituality, holistic learning, and communalism (Cajete, 2005, 1999; Deloria & Wildcat, 2001; Hilberg & Tharp 2002). These cultural differences require that the culturally congruent teacher possesses a deep cultural knowledge base and tailors her instruction to align with the cultural context in which she is teaching.

A small but growing body of studies is emerging in the research literature that provides evidence of the importance of culturally congruent instruction in supporting diverse students' academic achievement. A subset of these studies examines the efficacy of CCI in raising Indigenous students' science and mathematics achievement, disciplines in which the

achievement differentials between White and non-White students are particularly pronounced. Fourteen different quasi-experimental studies by Lipka and his team at the University of Fairbanks, for example, showed increased mathematics achievement in Yupik treatment students who were taught using a curriculum that incorporates traditional Yupik mathematical knowledge and teaching methods (Lipka, Parker Webster, & Yanez, 2005). Significant increases in achievement have also been correlated in studies of the use of culturally congruent instruction in mathematics and science with American Indian students (Cardell, Cross, & Lutz, 1978; Gilbert, 2005; Hilberg, Tharp, & Degeest, 2000; Matthews & Smith, 1994,). These studies provide preliminary evidence that suggests that CCI supports increased academic science and mathematics achievement in Indigenous students.

Culturally Congruent Instruction, Research, and Equitable Educational Outcomes

The calls from various sectors of society for equity in educational outcomes commonly include support for the increased use of CCI as one promising means for attaining that goal. Advocacy for more research on CCI is likewise increasing (Calabrese - Barton & Lee, 2006; Committee on Equal Opportunities in Science and Engineering [CEOSE], 2009; Executive Order 13336 American Indian and Alaska Native Education, 2004; Moses-Snipes & Snipes, 2005; Lee, 2005; NCTM Achievement Gap Task Force, 2004; Penfield & Lee, 2010; Tyler, et al., 2008). At this point, the corpus of knowledge on the nature and effects of CCI is young and undeveloped.

Significantly, the efforts to increase CCI research coincide with a greater push from federal agencies for the use of research based educational practice and increased rigor and utility in educational research (e.g., Institute of Education Sciences [IES], 2008). Studies on the efficacy and fidelity of CCI are important focuses of this research but tested methods and valid instruments appropriate for use in such studies that meet the more stringent guidelines for

rigorous research are lacking (Boykin, Tyler, Watkins-Lewis, & Kizzie, 2006; Lee, Luyckx, Buxton, & Shaver, 2007; Luyxk & Lee, 2007; Moses-Snipes & Snipes, 2005). Development of these types of instruments and protocols is complicated by the inherent specificity of CCI with each unique cultural context, a specificity which necessitates the tailoring of research methods and tools for the specific context in which they are to be used. Generating a pool of trained personnel who have the cultural and educational knowledge base for using instruments like classroom observation protocols with fidelity is also a challenge for conducting rigorous CCI research. The dissertation study described in this paper sought to address these issues in part by developing a teacher self report survey for use in assessing teachers' culturally congruent science instruction in teaching K-8 American Indian students.

Purpose of the Study

The purpose of this study was to develop an instrument for assessing culturally congruent instruction, specifically a survey in which teachers self report their frequency of use of culturally congruent instructional practices in teaching science with American Indian K-8 students in Montana. Although calls for the use of and research on CCI are increasing, there is a dearth of instruments for assessing CCI, in part due to the complexity of the construct and the required specificity of its operationalization with the prioritized cultural context. Developing a valid and reliable instrument for assessing CCI in the specific tribal contexts involved in this study is best done with the full participation of all relevant stakeholders, resulting in an inclusive, deliberate and iterative process. Such a process is also described in this paper, providing an illustrative example for others who may wish to undertake a similar endeavor.

Research Questions

This dissertation study describes the process of developing an instrument to assess teachers' CCI in teaching K-8 science with American Indian students. The research questions it addresses are:

A) What is a culturally congruent process for developing a valid instrument for assessing the use of CCI in teaching science with Montana American Indian students?

B) What is the technical quality of such an instrument?

Delimitations

1. Study participants were limited to K-8 teachers teaching in schools on and bordering Montana American Indian reservations, tribal consultants from the Kootenai, Salish, Crow, Northern Cheyenne, and Lakota tribes, and science and science education faculty members from three Montana institutes of higher education.
2. Instrument items addressed teachers' frequency of use of CCI content and practices, including student access to culturally congruent resources in science class.
3. The instrument developed does not fully represent one of the three major elements of the theory underlying CCI in this study, culturally congruent environment, due to the abstract nature of some the element's traits which present challenges to their operationalization. Resource limitations restricted the time and effort that could be expended to complete this aspect of the study.

Limitations

1. The context specificity of the instrument limits its generalizability to similar cultural contexts.
2. Treatment teacher recruitment was based in part on administrator recommendation and in

part on teacher self selection, so the sample was not random, thus increasing the potential for bias.

3. Because the study attempted to use a matched comparison group and the number of teachers teaching in similar settings and with characteristics similar to the treatment teachers is small in the rural and sparsely populated state of Montana, the comparison group teachers were recruited from a limited pool of available teachers so they were not a truly random sample.
4. The sample size used in the study's statistical analyses is moderate, thereby limiting the power of the analyses.
5. Because the survey is comprised of a set of written items and is completed in private by each teacher, it is subject to each individual teacher's interpretation of each item. Thus there is some inherent and unavoidable uncertainty in the study as to whether teachers interpreted the items as intended. This effect may have been confounded in this study by the fact that some of the teacher subjects were Second Language Learners of English or Limited English Proficient, having been raised speaking their Native languages and/or in households with Native speakers. The survey was written in the English language.

Significance of the Study

Factors important in establishing the significance of this study are previewed in the previous paragraphs and are recapped here: 1) With few exceptions, ethnically diverse students in the United States are academically underachieving compared to their White peers; 2) there is an increasing demand in the U.S. for equitable educational outcomes and for research on effective instructional strategies that will assist in attaining those outcomes; 3) there are cultural incongruities between the home cultures of many American K-12 students (45% of whom are non-White), K-12 teachers (81.9% of whom are White people of European descent) and the

culture of the typical American K-12 classroom; 4) there is an emerging body of research that suggests that CCI supports increased achievement in ethnically diverse students, perhaps by bridging the cultural divide between students, teachers and schools; 5) the evidence from the aforementioned research suggests that CCI should be pursued, both through further research and instructional practice, as a possible means of improving equitable educational outcomes; 6) CCI is an important construct that is complex and difficult to define; by its very nature, CCI's operationalization varies with the context in which it is located; 7) there are currently few valid instruments available for use in assessing CCI and its efficacy, in part due the construct's complexity and specificity to context; and 8) the processes and protocols for collaboratively developing assessment instruments are also specific to the cultural context in which the work is occurring. This study, by utilizing a culturally congruent process to collaboratively design and validate an instrument for assessing CCI in a specific tribal context, directly or indirectly addresses all of these factors. It directly addresses item 7 by making available a validated instrument that can be used in research and professional development focusing on CCI. The study also directly addresses item 8, through the development and use of a culturally congruent process for designing and validating that instrument. The description of the development process that is part of this study, while unique to the tribal contexts in which it was used, provides a model that may be useful to others endeavoring to engage in the development of instruments for use in ethnically diverse contexts.

The significance of this work ultimately lies in its ability to move CCI research and practice forward. The instrument itself and the process used to develop it have the potential to assist researchers in delineating effective instructional practices that support Indian students' science achievement, which will ideally translate into the increased use of CCI in schools. These

advances hold promise for improving the equity of educational outcomes for ethnic minority students. Improved educational outcomes are correlated with other positive outcomes that benefit individuals as well as society overall, for example, lower unemployment rates and greater earning power, translating into improved standards of living, less dependence on social welfare programs, lower incarceration rates, improved health, a stronger tax base, a stronger economy overall, greater creativity, more powerful research and development, and increased civic participation (Baum, Ma, & Payea, 2013; Dee, 2004; Milligan, Moretti, & Oreopoulus, 2003).

The significance of the CCI instrument (CCIS) development process cannot be overstated and deserves additional comment. The processes used to develop instruments such as the CCIS are integral to the generation of a valid instrument that is valuable to research and to all of the research partners (Connor, 2004; Kirkhart, 2005; LaFrance, 2004; Nelson-Barber, LaFrance, Trumbull, & Aburto, 2005). The participatory processes that were employed in developing the CCIS, for example, were deliberately designed to be culturally congruent both out of respect for the participants involved and to help ensure the validity of the instrument. With these ideas in mind, the uniqueness of every culture dictates that instrument development processes are customized specifically for the prioritized context. This includes important elements such as the observation of cultural norms for the prioritized cultures involved in the research and the equitable participation of relevant stakeholders, particularly those identified by the communities as respected representatives. The stakeholders for the development and use of the CCIS, for example, included both American Indian and non-Indians from several groups ranging from college and university educators, to K-8 educators, to professional developers in science education, to members of the tribal communities who were not formally employed as educators. Efforts were made to ensure the cultural congruence of every possible aspect of the process from

start to finish, from the identification of participants for the development process, to the formats used for gathering and validating information, to the methods used for member checking each item for appropriate content and language. Without such attention to cultural congruence in the development processes, an instrument is open to additional threats to its validity, the ethics of the development process are subject to question, and the likelihood of future research with the partners is hindered. Thorough details of the development process used in this study can be found in Chapter Four of this paper and serve as one example of an effective model of instrument development that others may take lessons from and/or customize for use in their own settings to insure the validity of their research and the perpetuation of positive research partnerships.

Definitions of Key Terms

The following terms will be a key part of this paper. Their definitions, for the purposes of this study, are described below.

American Indian – The exact meaning of this term is widely debated but in this paper it means "a person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment" (U.S. Census Bureau, 2012, p. B-8). This study involved American Indian people from five Montana tribal cultures - the Kootenai, Salish, Pend d'Oreille, Northern Cheyenne and Crow.

Concurrent Validity – Concurrent validity is the ability of the operationalization of a construct to produce data that estimate a current status or outcome that it theoretically should be able to predict (Web Center for Social Research Methods, 2006). Evidence of concurrent validity is established in this study when the data produced by the study instrument are compared to those

of an already existing measure to determine their relationship. It is a form of criterion related validity.

Construct Validity - In the modern unitary view of validity, construct validity is the overarching framework for all validity and all evidence of validity is gathered to support the inference made about the construct (Messick, 1990). According to the Web Center for Social Research web site (2006) construct validity is the "degree to which inferences can legitimately be made from the operationalizations in your study to the theoretical constructs on which those operationalizations were based" ("Construct Validity", para. 1). In this paper, it is the degree to which the instrument developed in this study enables accurate inferences about teachers' culturally congruent instruction, as operationalized by the instruments' items. Evidence of construct validity is two pronged, stemming from both theoretical and empirical sources. A high degree of construct validity increases one's confidence in the results suggested by the data obtained from administering the instrument.

Convergent Validity – Convergent validity is a subcategory of construct validity that measures the degree to which two instruments that purport to measure the same construct actually do measure the same construct (Web Center for Social Research Methods, 2006). Evidence of convergent validity must be accompanied by evidence of discriminant validity in order to constitute evidence of construct validity.

Criterion Related Validity – Criterion related validity is the degree to which the data produced by an instrument can predict or are statistically related to an outcome or criterion (United States Office of Personnel Management, 2014). It is a measure of how well the operationalization of a construct performs in practice to generate data that can be used to accurately estimate a quality or outcome. It is often said to have two forms, concurrent and predictive.

Culturally Congruent Instruction - Lee and Buxton (2010) describe culturally congruent instruction as occurring when "teachers interact and communicate with students in ways that are familiar to students in their homes and communities, as well as use cultural artifacts, examples, analogies, and community resources." (p. 65). In this study it is defined as instruction (including content, pedagogy and classroom environment) that is compatible with and builds upon students' cultures such that it validates students' cultural identities, empowers students, builds on their prior knowledge and traditional ways of knowing, and supports their achievement.

Culture - Guitierrez and Rogoff (2003) define culture as a dynamic repertoire of beliefs and practices developed through participation in a cultural community whose members span generations and share traditions and understandings that are based in the group's experiences. Culture is a dynamic construct that is continuously being shaped by a people's history and ongoing interactions with other people and their environment.

Discriminant validity – Discriminant validity is defined as a subcategory of construct validity that measures the degree to which two instruments each measure a different trait (Web Center for Social Research Methods, 2006). To be meaningful, evidence of discriminant validity must be accompanied by evidence of convergent validity.

Factor Analysis – Factor analysis is a set of statistical techniques that is used to tease out the relative influence of various factors on an outcome. Factor analysis seeks to reduce a larger set of variables that influence an outcome into a smaller set of factors by determining the relationships of variables and grouping them into factors (DeVellis, 2003).

Indigenous – Indigenous refers to organisms existing or living naturally in an area or region (Merriam – Webster, 2014). In this paper, indigenous refers to the people who are the original known human inhabitants of an area.

Instructional congruence – Instructional congruence is defined as culturally congruent instruction within a specific discipline. Lee (2005) characterizes instructional congruence as instruction that is appropriate for specific disciplines (like science) through the merging of "discipline specific" and "diversity oriented" pedagogies (p. 858).

Internal Consistency – Internal consistency is defined as a form of reliability that measures the degree to which different items on an assessment instrument correlate, or produce similar scores. It is an indication of the probability that the items in question are measuring the same construct (Web Center for Social Research Methods, 2006).

Multicultural validity – Kirkhart (2005) states that multicultural validity "refers to the correctness or authenticity of understandings across multiple, intersecting, cultural contexts" (p. 22).

Native American – As with the term American Indian, the meaning of the term Native American is widely debated. Generally, it is defined as any of the first group of people who inhabited the Americas (Merriam – Webster, 2014). In this paper, the term means indigenous people living within the United States and so it includes American Indian people. As opposed to American Indian, the term Native American, as used in this paper, includes other indigenous people such as Alaska Native and Native Hawaiian people.

Process validity - The extent that the methods employed in conducting research are adequate, sound, and appropriate for the study's context such that they enable ongoing learning for all of the research partners (Anderson & Herr, 1994)

Reliability – Krathwohl (1998) defined reliability as the "consistency of results produced by a measure" (p. 436), or the tendency for an instrument to produce similar results on repeated administrations under similar conditions. Reliability can be estimated by examining the data

produced by the instrument for internal consistency, temporal stability, and equivalence using parallel forms of the instrument. Finding evidence of reliability is not enough to establish validity although it is a necessary element of validity.

Relational accountability – The importance of relationship with and accountability toward a community. Wilson notes that research methodology with relational accountability "needs to be based in a community context (be relational) and has to demonstrate respect, reciprocity and responsibility (be accountable as it is put into action)" (Wilson, 2008, p. 99).

Self-efficacy- Self-efficacy is a person's belief about their ability to attain a goal or to influence an outcome (Science Education Resource Center, 2014).

Validity – Validity is the extent to which the data generated by an instrument enable accurate inferences about the construct the instrument intended to measure (Web Center for Social Research Methods, 2006).

Outline of the Paper

Chapter One provided a brief overview of the focus, purpose, and significance of this study to develop an instrument to assess CCI. The subsequent chapters of this paper describe and discuss the study in fuller detail. Chapter Two consists of a more detailed examination of the construct of CCI and issues related to it and a review of the relevant literature. Chapter Three provides a broad overview of the methods employed in the study in investigating each of the research questions. Chapter Four addresses the results of the study for Research Question #1, describing in detail the participatory processes undertaken to develop and validate the survey. Chapter Five describes the results of the quantitative methods used to analyze teachers' survey data to generate evidence of the instrument's validity. Finally, Chapter Six examines and discusses the findings from the study and their implications, discusses the delimitations and

limitations of the study, and makes recommendations for further research and development of the instrument generated in this study.

CHAPTER TWO – REVIEW OF RELATED LITERATURE

Chapter Two discusses in eight sections issues related to this study and a review of relevant literature. It begins with an examination of the academic achievement gap that has long existed for ethnic minority students in the K-12 schools in the United States, particularly in mathematics and science. The second section describes current advocacy efforts by a variety of educational stakeholders for equitable educational outcomes for ethnically diverse students. Historic and contemporary calls for the increased use of CCI in American Indian education are the focus of the third section. In the fourth section, the constructs of culture and culturally congruent instruction are examined. This is followed in the fifth section by a description of research studies that suggest that CCI supports improved science and mathematics achievement in Indigenous students are described. The sixth section of this chapter discusses the calls emanating from education stakeholders in the U.S. for an increase in rigorous and meaningful research on CCI. Challenges for assessing CCI are the subject of the seventh section. Finally, Chapter Two concludes with a discussion of the protocols and procedures for working collaboratively with tribal communities, particularly in education and research.

Achievement Gaps in Science and Mathematics for Ethnically Diverse Students

According to the 2010 national census, the overall population of the United States is growing more ethnically diverse, with 43% of the nation's population identifying themselves as non-White (United States Census Bureau, 2010). The diversity of the U.S. K-12 student population is likewise increasing, with 45% of U.S. students identified as being of ethnic minority descent. Projections by the National Center for Education Statistics predict that non-White students will make up the majority of students enrolled in K-12 by the 2014 (NCES, 2013a). Meanwhile, the overwhelming majority of U.S K-12 teachers (81.9%) are White people

of European descent (Institute of Education Sciences [IES], 2008). Western culture, based on northern European values, norms, and worldviews, continues to be the predominant cultural influence in much of U.S. society and its institutions, including education (Hollins, 2008; Singh, 2011).

Although in some cases the disparities are slowly narrowing, a wide array of measures of academic achievement continues to indicate that most groups of ethnic minority students attending America's schools are significantly underachieving compared to their White peers. On the National Assessment of Educational Progress for mathematics taken by nine year-old students in 2008, for example, U.S. Latinos/as scored 16 points lower and African Americans scored 26 points lower than White students on a 500 point scale (Rampey, Dion, & Donahue, 2009). In most cases, differentials in test scores by ethnicity increase as students progress through elementary and secondary school.

Other measures of academic success affirm this pattern of underachievement for most ethnic minority student groups. For example, White students drop out of precollege education at a rate of 6% of their total number, African Americans drop out at a rate of 11% and Latino/a students drop out at a rate of 22% (Rampey et al., 2009). Forty one percent of all dropouts are Latino/a, even though they only make up 17% of U.S school age youth. College degree completion rates show similar patterns, with 66% of all degrees awarded between 2005 and 2007 to non-Hispanic Whites, on par with their total population percentage. Nine percent of degrees in that same time period were awarded to African Americans and 7% were awarded to Latinos/as, ethnic groups that made up 12% and 15% of the U.S. population respectively (Rampey et al., 2009).

In Montana, scores on measures of achievement are likewise out of balance for American

Indian and White students, the state's two largest ethnic groups. On the 2007 NAEP fourth grade test of reading, for example, only 17% of Montana's American Indian students scored in the proficient or higher categories compared to 42% of White students. Similarly, on the 2007 NAEP eighth grade mathematics test, only 15% of American Indian students scored in the proficient or higher categories compared to 41% of White students (The Education Trust, 2009). Although Montana students overall scored second highest among the fifty states on the 2005 eighth grade NAEP science test, only 14% of Montana's American Indian students scored in the proficient or higher category on the test, compared to 45% of White students (U.S. Department of Education, 2008). Students overall score higher on Montana's relatively new criterion referenced tests (CRT), but the gaps in scores between American Indian and White students persist. Only 63% of the state's American Indian students scored in the proficient or higher categories on the 2010 eighth grade reading state CRT, compared to 87% of Montana's White students. Similar disparities occurred in the scores on the 2010 eighth grade mathematics state CRT, with only 40% of Montana's American Indian students scoring in the proficient or higher categories, compared to 71% of White students. In science the same pattern emerged, with 29% of Montana's eighth grade American Indian students and 62% of White students scoring proficient or above on the science section of the Montana eighth grade CRT (Montana Office of Public Instruction, 2010). On-time high school graduation rates were 58% for Montana's American Indian students in 2006, compared to a rate of 84% for White students in the state for the same year. In 2006, 11% of American Indian and 28% of White adults in Montana over the age of twenty-five had attained a four-year college degree (The Education Trust, 2009).

The Push for Equitable Education Outcomes

The achievement differentials between ethnic subgroups described above have not gone unnoticed by U.S. public, private, and government sectors. National education initiatives and

federal legislation and policies are increasingly prioritizing the closing of achievement gaps. Last renewed in 2002 and at this time up for renewal in Congress, the federal Elementary and Secondary Education Act (more commonly known as No Child Left Behind or NCLB) which holds every state accountable for the academic success of *all* students, is currently the most pervasive and influential example of federal legislation with equitable outcomes as a primary goal (No Child Left Behind, 2002). NCLB aims to improve equity in educational achievement outcomes by ensuring that every school holds high standards and provides challenging curriculum for all students. Under NCLB, states are required to define high standards for K-12 student achievement, to collect annual data that measure student achievement (in this case, standardized test scores), and to publicly report student achievement test results disaggregated by subgroups such as ethnicity, socioeconomic status and gender. Failure to consistently meet the law's student achievement benchmarks in mathematics and reading results in prescribed and required measures that modify the school in an attempt to improve student test scores. Also largely as a result of NCLB mandates, teacher preparation programs at U.S. colleges and universities are being retooled with the purpose of preparing teachers who can better support all students' learning. This law has shone a bright light on the achievement gap and has profoundly and inexorably changed the culture of American education, decidedly prioritizing equitable outcomes.

Current federal funding initiatives for education are also closely tied to the achievement of equitable educational outcomes. Major funding from the U.S. Department of Education (USDOE) known as "Race to the Top" (RTTT), which is being distributed to states through a competitive application process that began in 2009, has as a final priority the closing of achievement gaps. In the RTTT application's explication of RTTT Priority #1: Absolute Priority

- Comprehensive School Reform, it states that applications should demonstrate commitment to use RTTT funds to "increase student achievement, decrease the achievement gaps across student subgroups, and increase the rates at which students graduate from high school prepared for college and careers" (p. 51). Likewise, RTTT's Priority #2: Competitive Preference Priority - Emphasis on Science, Technology, Engineering, and Mathematics, emphasizes that applications for RTTT funding should include a high quality plan to "prepare more students for advanced study and careers in the sciences, technology, engineering, and mathematics, including by addressing the needs of underrepresented groups and of women and girls in the areas of science, technology, engineering, and mathematics" (p. 51). Nearly every state has applied for RTTT funding, so large is the carrot. As of Summer 2014, only eighteen RTTT awards had been made; each of the successful applicants included ample proof of past success in improving the equity of their educational outcomes.

A second major funding source currently being distributed through the USDOE known as "Investing in Innovation" (I3) has also identified closing achievement gaps as a priority. Further, I3 lists as criteria for eligibility that applicants "have significantly closed the achievement gaps between groups of students and have demonstrated success in significantly increasing student academic achievement for all groups of students" (United States Department of Education, 2009, p. 1). RTTT and I3 are two examples of major federal funding sources that prioritize equitable educational outcomes and whose stringent proposal guidelines require successful applicants to commit to educational systems designed to improve educational outcomes for *all* students.

National professional STEM education organizations are also focusing on equitable educational outcomes. The National Council for Teachers of Mathematics Achievement Gap Task Force (AGTF) 2004 report highlights NCTM's focus on reducing the mathematics

achievement gap through attention to equity issues in four categories: leadership and infrastructure, research, policy and political action, and professional development. Further evidence of NCTM's commitment to equitable outcomes appears in their 2008 position paper titled "Equity in Mathematics Education" which states

> A culture of equity depends on the joint efforts of all participants in the community of students, educators, families, and policymakers:
>
> - All members of the community respect one another and value each member's contribution.
> - The school community acknowledges and embraces all experiences, beliefs, and ways of knowing mathematics.
> - All necessary resources for optimal learning and personal growth of students and teachers are allocated.
> - High expectations, *culturally relevant practices* [italics added], attitudes that are free of bias, and unprejudiced beliefs expand and maximize the potential for learning.
> - All students have access to and engage in challenging, rigorous, and meaningful mathematical experiences. (NCTM, 2008, p. 1)

Similarly, the National Science Teacher Association, in a policy statement on multicultural education adopted by their Board of Directors in July 2000, documented the organization's commitment to equity and CCI by declaring that

- Schools are to provide science education programs that nurture all children academically, physically, and in development of a positive self-concept;
- Children from all cultures are to have equitable access to quality science education experiences that enhance success and provide the knowledge and opportunities

required for them to become successful participants in our democratic society;

- Curricular content must incorporate the contributions of many cultures to our knowledge of science;
- Science teachers are knowledgeable about and use culturally-related ways of learning and instructional practices;
- Science teachers have the responsibility to involve culturally-diverse children in science, technology and engineering career opportunities; and
- Instructional strategies selected for use with all children must recognize and respect differences students bring based on their cultures (NSTA, 2000, para. 3).

Clearly, as the United States student population becomes increasingly diverse and the country continues to value an educated citizenry, diversity and equity issues are growing in importance in United States education, and the cultural congruency of instruction is gaining ground as a potential strategy for achieving equitable education outcomes.

The History of Advocacy for CCI in American Indian Education

CCI and the United States Government. As discussed previously, American Indian students are underachieving compared to their White peers as measured by a wide array of assessments of academic achievement. This evidence of an achievement gap has long existed and, in many cases, is substantial. Documentation of American Indian/White achievement gaps date back at least as far as the Meriam Report of 1928, a study commissioned by the U.S government and conducted by the Brookings Institute on the economic and social conditions of American Indians, including their education. The study's report described the inequities in educational access and outcomes for American Indians at that time and suggested specific strategies for improving them. While the language of the report is decidedly racist in its

discussion of American Indian people, it is often cited as one of the first major reports that advocated the use of culturally congruent content to improve education for American Indian students. Unfortunately, education in the United States in the 1920s was still largely seen by most White people as a means to assimilate tribal peoples and culturally congruent practices did not show up significantly in schools educating American Indian students until several decades later.

Since the time of the Meriam Report, the federal government has on many occasions shown its support for culturally specific programs and practices for American Indian education with the passage of a series of federal laws and policies. The Federal Indian Education Act of 1972 (at that time known as Title IX) was an important piece of this legislation that marked the point at which the U.S. government officially pledged to support educational programs and practices specifically designed to support American Indian learners. Now part of NCLB as Title VII, the act states that

> The Federal Government will continue to work with local educational agencies, Indian tribes and organizations, postsecondary institutions, and other entities toward the goal of ensuring that programs that serve Indian children are of the highest quality and provide for not only the basic elementary and secondary educational needs, but also *the unique educational and culturally related academic needs of these children*. [italics added] (p. 483)

Just what is meant by "unique educational and culturally related academic needs" is left undefined in the act, but NCLB further states that the government will provide federal assistance to educational agencies in meeting these needs and in funding research, evaluation, and training on these topics. Tribes, researchers and schools serving Native students have made use of these

grant funds over the years to improve the CCI in their schools for example, by developing culturally relevant curriculum, and by providing training in cultural competence for their school employees.

Other notable federal legislation and policy affecting the education of American Indian people include the Indian Self-Determination and Education Assistance Act of 1975 which advocated the development of tribally controlled schools and educational programs, and the Native American Languages Act of 1990 which advocated the use of Native language in instruction to improve educational access and achievement for American Indian students. More recently, President Bill Clinton in 1998 signed Executive Order 13096 on American Indian and Alaska Native Education (American Indian & Alaska Native Education, 1999). It emphasized the federal government's commitment to working with schools to develop and evaluate the effectiveness of CCI practices in supporting American Indian student achievement and to the dissemination of such practices so as to assist tribes in meeting the educational needs of their people and in increasing American Indian student achievement. In a similar vein in 2004, George W. Bush signed Executive Order 13336, revoking the previous order but reiterating the government's commitment to assist Native students in meeting the academic standards of NCLB "in a manner that is consistent with tribal traditions, languages, and cultures" (p.1), and to study and disseminate instructional practices that support Native student achievement (American Indian and Alaska Native Education, 2004).

Advocacy for CCI by American Indian People. American Indian people themselves have long called for, both formally and informally, the use of CCI in schools as a means to improve the equity of Native students' educational experiences and outcomes. In their 2007 annual report the National Indian Education Association (NIEA) noted that

> NIEA's top priority is to strengthen the education of American Indians, Alaska Natives, and Native Hawaiians through effective and meaningful *education programs and approaches that reflect Native cultures, traditions, and languages* [italics added], including promoting these programs as part of a strategy to reform high schools so that they prepare Native students for graduation and college. NIEA is committed to strengthening Indian education through provisions that provide for *meaningful tribal involvement in setting the educational priorities for Indian students and the inclusion of Native language and cultural instruction… NIEA will focus on the promotion of instructional practices designed to meet the needs of diverse learners, specifically, cultural based education for American Indian, Alaska Native, and Native Hawaiians through national and state policy.*[italics added] (p. 13)

NIEA is very active at the grassroots and government level with efforts to influence federal legislation supporting the use of CCI in Native students' education. In 2005, NIEA sponsored a series of eleven hearings, held across the United States, to discuss the impacts of NCLB on Native education. Over 120 witnesses testified, the majority of whom were representatives of their respective tribes, and numerous letters and e-mails were also submitted, affirming tribes' support for the use of CCI. The resulting report titled the "Preliminary Report on No Child Left Behind in Indian Country" (National Indian Education Association & Center for Indian Education, 2005) stated that

> Many witnesses identified what could generally be labeled the unintended consequences of the statute (NCLB) that has resulted in major disruptions to the education systems that may fundamentally alter the education potential of schools while significantly and coincidentally narrowing the broad public purposes of schools. This later concern is most

> directly related to the impacts of the statute upon culturally based education including the use of culturally appropriate pedagogy and curriculum that is connected to the social, cultural, and linguistic heritage of the children, the role of Tribal governments and Native communities and parents in determining the education purposes of schools and the role of teachers, parents and community members in the education lives of Native students. (p. 6)

Also as a result of these hearings, NIEA wrote a report on their policy recommendations for NCLB and has submitted to Congress a number of recommended amendments to the act, some of which have been included in the new version of the bill, currently awaiting renewal. Areas of emphasis in particular include:

- Improving Title VII to address the unique cultural and educational needs of Native children
- Strengthening NCLB to provide support for instruction in Native American languages
- Improving cooperation among tribes, states, and the federal government
- Improving support for teachers of Native students and
- Funding for NCLB, especially Title VII. (NIEA, 2007, p. 15)

Again on the national level, a series of six meetings was held across the country in 2010 between American Indian and Alaska Native leaders and personnel from the United States Department of Education to discuss the state of Native students' education. Among other concerns, Indigenous leaders emphasized the cultural mismatch of school curriculum and instruction, standards, and assessments for Native students that has resulted in students' lowered self-esteem, loss of cultural knowledge, and lagging achievement. Many singled out the detrimental effects of NCLB in encouraging a generic educational system that does not consider

the uniqueness of Native students and called for federal support in the development and delivery of CCI designed to meet the unique educational needs of Native students (U.S. Department of Education, 2010).

On the tribal level, many individual American Indian tribes, in an exercise of their sovereignty, have officially defined and adopted their own educational policies. CCI is commonly a central aspect of tribal education policies. The Yankton Sioux Tribal Code includes the Education and School Code that states:

> Since education is, in part, the transmission of culture and values, education within schools and other educational institutions chartered or operated by the Ihanktowan Oyate shall include the teaching of the N/Dakota and Ihanktowan culture and values…These declarations are in accord with the policies of the Congress of the United States, which recognize a primary means by which a child learns is through the use of such child's native language and cultural heritage, and instructional use and development of a child's non-English native language promotes student self-esteem, subject matter achievement and English language proficiency. (Yankton Sioux Tribe, 1995**,** Sec. 15-2-3, Findings and Declarations, paras. B3 & B5**)**

The Yankton Sioux Tribal Code also declares that education for Yankton people should include the involvement of parents, tribal elders, and "eminent leaders" to the "maximum extent practicable," a strategy that has been identified by scholars as an essential element of CCI for American Indian students (Yankton, Sioux, 1995, Sec. 15-5-3 N/Dakota Language and Cultural Courses, paras**.** B & C).

Likewise, the Ute Tribe, in their Ute Tribe Education Department Goals and Actions, emphasize the use of CCI to improve the educational experience and academic achievement of

Ute people. Actions recommended include teacher training in cultural sensitivity, the inclusion of cultural sensitivity criteria in teacher performance evaluations, the codevelopment and implementation by teachers and tribal community members of curriculum that is place based and includes Ute culture and language, and the development of collaborative relationships between educators, parents, other members of the tribal community, and tribal departments so as to include them meaningfully in the education of their children (Ute Tribe, 2004). The Ute and Yankton Sioux tribes are but two of the many tribes who have formally advocated for the use of CCI in schools educating their people, a testimony to the significance of CCI for American Indian people and to the promise it holds for them in improving educational outcomes for their children.

The Construct of Culturally Congruent Instruction

Culture is a complex construct that is not easily defined. Carter (2000) describes culture as "learned patterns of thought and behavior that are passed from one generation to another and are experienced as distinct to a particular group (p. 865). Demmert and Towner (2003) state that culture can be "viewed as the beliefs, behaviors, and characteristics of a particular social, ethnic, or racial group, and includes application of both traditional and contemporary mores and understandings as influenced by individuals and groups" (p. 5). Guitierrez and Rogoff (2003) take a historical perspective in defining culture as a *dynamic* repertoire of beliefs and practices developed through *participation* (as opposed to just membership) in a cultural community whose members span generations and share traditions and understandings that are based in the group's experiences. They emphasize that cultural community membership and shared practices undergo constant transformation making culture a dynamic, rather than a static, construct. Lee (2010) writes that culture "generally refers to the values and worldviews shared by the members of a

social group" and notably points out that culture "serves as a framework for how we interpret and interact with other individuals and with the broader world around us" (p. 12). Regardless of the exact definition, it is generally agreed that culture is dynamic and is continuously being shaped and reshaped by a people's history and ongoing interactions with other people and their environment.

The term "culturally congruent" first appeared in the education literature in the 1980s. Au & Jordan (1981) used the term in an article describing a reading program designed specifically for Native Hawaiian students that is based on the Hawaiian cultural traditions of *talk story* and *storytelling*. The cultural congruence of the program lies in its use of these two traditions, which are normal modes of communication for Native Hawaiians, and are characterized by the mutual participation of all present, in this case teachers and students, in the co-narration of stories. The authors note that cultural congruence in classroom teaching varies with the cultural context but can include "the behaviors of the teachers, the social organization of the class, the types of participation structures, and the physical arrangement of the classroom" (p.152), among other things.

In the same volume, Mohatt and Erickson (1981) invoke the term cultural congruence when discussing their study of the social interactions of two teachers, one Indian and one non-Indian, with students in their classrooms in an Odawa Lakota school. The Indian teacher's social interactions more closely mirrored those of the Odawa students' home lives (e.g., less authoritarian in nature, less likely to put students on the spot, and slower paced and more personal discussions) and so were judged to be more culturally congruent.

Pewewardy and Hammer (2003) describe culturally congruent instruction as that which builds a bridge between the student's home culture and that of the school to support students'

learning and achievement. Lee and Buxton (2010) describe culturally congruent instruction as occurring when "teachers interact and communicate with students in ways that are familiar to students in their homes and communities, as well as use cultural artifacts, examples, analogies, and community resources" (p. 65). Some researchers also explicitly include the potential for social action and transformation as an integral part of CCI, realized through its ability to empower students as agents of change who can effect more equitable power relationships in society (cf. Gay 2010; McGee - Banks & Banks, 1995).

Related terms that are often used interchangeably with CCI in the education literature include culturally responsive education (CRE), culturally based education (CBE), and culturally relevant teaching (CRT). For example, Gay (2010) defines culturally responsive teaching as "using the cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning more relevant to and effective for them. It teaches to and through the strengths of these students" (p. 29). While each scholar tends to favor one term over another in her/his own writing and research and some have written about the nuances of each term, there does not appear to be general agreement on their precise definitions and overlap in their use still occurs in the literature.

Lee and her co-researchers have postulated a subcategory of CCI in a construct they call "instructional congruence," which they describe as a *discipline focused* form of CCI. Lee (2005) characterizes instructional congruence as instruction that is appropriate for specific disciplines (like science) through the merging of "discipline specific" and "diversity oriented" pedagogies (p. 858). Lee (2003) describes instructional congruence as

> the process of mediating academic disciplines, such as science, with students' language and culture to make the academic content accessible and meaningful for

> students….Instructional congruence focuses on articulating academic disciplines with students' linguistic and cultural experience to develop congruence between the two domains….Instructional congruence emphasizes the role of instruction (or educational interventions) as teachers explore the relationship between academic disciplines and students' linguistic and cultural knowledge and devise ways to link the two. (p. 474)

Just as there are distinct differences and yet considerable overlap in how scholars refer to and define CCI, common themes with slight variations also exist in what scholars identify as the essential behaviors and dispositions that operationalize CCI. Illustrative examples of these are found in Table 1, which lists actualizing characteristics of CCI as identified in the literature by four different sets of authors.

Examination of the four synopses found in the table reveals that each author(s) lists a set of traits characteristic of CCI, derived either from his or her original work or from a review of the work of others, that are uniquely worded and nuanced but that also overlap substantially in their content with that of others in the table, and, in fact, with the work of CCI scholars in general. For example, the use of culturally responsive pedagogy and culturally relevant content are commonly identified in the literature as essential elements of CCI, as they are by all four sets of authors in the table. Another common emphasis is that of the culturally congruent instructional environment. Two of the authors listed in the table explicitly identify the establishment of culturally responsive learning environments as important to culturally competent instruction. Meanwhile Demmert and Towner (2003) implicitly include classroom environment through their identification of the observance of cultural mores of behavior and traditional interactions between adults and Indigenous students as essential to culturally competent instruction. Further, all four either implicitly or explicitly emphasize the

Table 1

Elements of CCI as Identified by Four Sets of Authors

Author(s)/Yr	Elements of Culturally Competent Instruction Identified
Phuntsog, 1999	Five critical elements of culturally responsive practice, derived from the author's review of the literature: • Stresses respect for diversity to engage the motivation of all learners • Creates a safe, inclusive, and respectful learning environment • Integrates responsive teaching practices into all disciplines • Transforms curriculum to promote social justice and equity in society • Is culturally literate
Gay, 2010	Five traits of culturally responsive teaching, identified by the author: • Acknowledges the legitimacy of the cultural heritages of different ethnic groups, both as legacies that affect students' dispositions, attitudes, and approaches to learning and as worthy content to be taught in the formal curriculum • Builds bridges of meaningfulness between home and school experiences as well as between academic abstractions and lived sociocultural realities • Uses a wide variety of instructional strategies that are connected to different learning styles • Teaches students to know and praise their own and each others' cultural heritages • Incorporates multicultural information, resources, and materials in all the subjects and skills routinely taught in schools (p. 29)
Demmert & Towner, 2003	Six elements of culturally based education for Indigenous students, synthesized from the authors' review of the literature: • Use of Native language • Pedagogy that uses traditional cultural characteristics and adult child interactions • Pedagogy that emphasizes both traditional and contemporary ways of knowing • Curriculum based on traditional culture and contemporary contexts and that recognizes the significance of spirituality • Significant community involvement in the planning and operation of education • Use of community mores in classroom interactions
Siwatu, 2005	Four traits identified by the author as generally accepted as characteristics of culturally responsive teaching: • Uses students' cultural knowledge, experiences, prior knowledge, and learning preferences to facilitate the teaching and learning process • Incorporates students' cultural orientations to design culturally competent classroom environments • Provides students with multiple opportunities to demonstrate what they have learned using a variety of assessment techniques • Provides students with the knowledge and skills needed to function in mainstream culture while helping them maintain their cultural identity

acknowledgement and validation of cultural diversity as a trait of culturally competent instruction. It could be argued that this element is necessarily ubiquitous throughout all of the other elements, i.e., the observation of cultural uniqueness as manifested through the use of culturally congruent content, pedagogy and environment at least implicitly and frequently explicitly validates cultural diversity.

Figure 1 presents a visual representation of the general construct of CCI for ethnic minority students that incorporates the critical elements of culturally congruent content, environment, and pedagogy commonly attributed to CCI in the literature. The graphic portrays the importance of all three of these major elements in fostering CCI and does not prioritize one over another. The double headed dashed arrows in the graphic represent the interactive and nonexclusive nature of the actualizing elements of CCI; many of the sub elements of CCI could arguably be placed in more than one of the three categories and often one cannot be used effectively without others. For example, the use of culturally congruent pedagogy through the inclusion of Kootenai elders as mentors and teachers is ideally accompanied by the observation of the cultural protocol of gifting the elder for their work. Likewise, the use of cultural content in the form of traditional Coyote Stories of the Inland Salish requires the observation of cultural protocols that dictate they only be told in winter.

Delineating the common elements of CCI for ethnic minority groups as shown in Table 1 and Figure 1 assists the reader in broadly defining CCI. This type of general characterization of CCI, however, lacks precision and provides only vague notions of the types of practices characteristic of CCI for a specific ethnic group and context. Indeed, if one reconsiders the array of definitions of culture from several scholars provided earlier in this paper, one can see that the common thread running through them is the emphasis on the uniqueness of each culture, a

uniqueness that is shared by the members of one culture but not necessarily by members of other cultures. It follows then that such uniqueness requires that the manifestations of CCI are also unique and specific to each cultural context. The literature provides examples of CCI practices

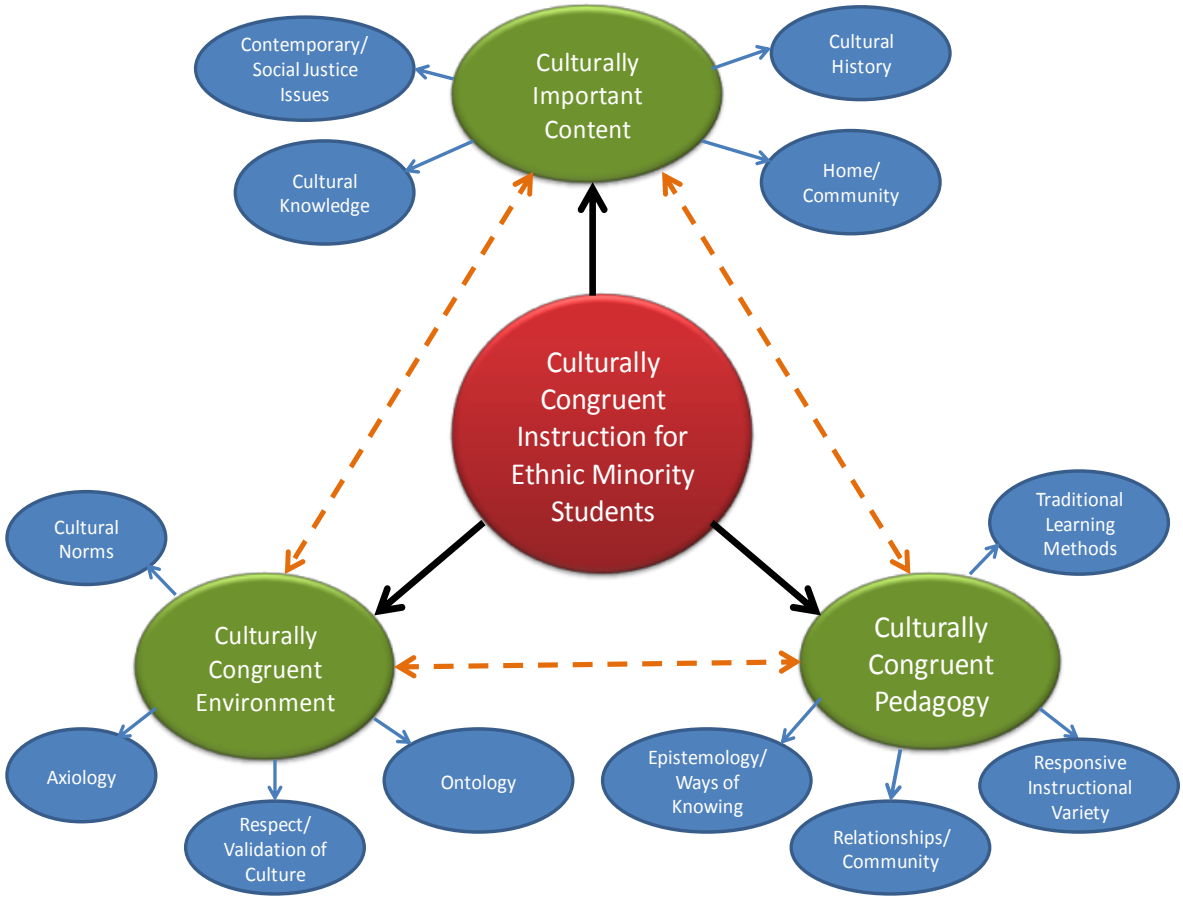

Figure 1. Model of the construct of culturally congruent instruction for ethnic minority students.

specific to various ethnic groups that illustrate this point. For example, instruction cited by scholars as compatible with the home cultures of many African American students incorporates elements of movement, verve, and communality (Boykin & Bailey, 2000; Boykin, Coleman, Lilja, & Tyler, 2004; Hurley, Boykin, & Allen, 2005). In contrast, instruction that has been identified in the literature as compatible with the cultures of many American Indian

students incorporates elements of student reflection and private practice, observational learning, multiple mentors from the extended family and community, spirituality, holistic learning and communalism (Cajete, 2005; Deloria & Wildcat, 2001; Hilberg & Tharp, 2002). Also in contrast, instructional practices that scholars describe as compatible with the home cultures of White European American students incorporates individualism, competition, linear logic, and risk taking (Cajete, 1999; Deloria & Wildcat, 2001; Gay, 2010; Hollins, 2008). Thus, while scholars commonly identify overarching elements that characterize CCI, it can be seen from the aforementioned examples that the actualizing elements of CCI for a specific cultural group of people are often unique and specific to that group.

Research on CCI in Science and Mathematics Education with Indigenous People

For decades, tribal entities, educational scholars specializing in diversity and equity, the federal government, and national education organizations have advocated the use of CCI to improve educational outcomes for ethnic minority students, including Native American students. Meanwhile, empirical evidence supporting the efficacy of CCI for improving Native American students' achievement, particularly in mathematics and science education, remains somewhat limited. The number of relevant CCI studies involving Native American students is small and many of the studies that have been reported did not employ rigorous research methodologies, such as the use of treatment and comparison groups or the random assignment of subjects, thereby weakening the scientific credibility of the evidence they provide. Demmert and Towner (2003), in an extensive literature review, found few studies that used rigorous methodology and even fewer that provided evidence of the efficacy of CCI in improving student achievement. This section describes studies that, relative to other studies of CCI and Indigenous students, utilized research designs with at least moderately high rigor and that provide evidence regarding the

efficacy of CCI in affecting Native American students' mathematics and science achievement.

Studies of CC mathematics instruction with Native students. Between spring 2001 and spring 2005, Lipka, Parker Webster, and Yanez conducted fourteen quasi-experimental trials of culturally congruent mathematics instruction in which treatment students were taught using the research group's original curriculum known as *Math in a Cultural Context* (MCC). Designed in collaboration with Yupik elders and mathematics reform oriented K-20 educators, MCC employs culturally congruent content and pedagogy in a guided problem solving curriculum designed to support students' semi-autonomy in regulating their own learning. The trials involved treatment and control groups of both Yupik and non-Yupik students in rural Alaska schools who completed pre- and post-instruction tests as a means to assess possible achievement impacts of the treatment. Over 3000 students in grades ranging from 2nd through 6th comprised the total sample for the fourteen trials. In thirteen of the fourteen trials conducted, pre- and post-assessment scores showed statistically significant greater gains for the treatment group students compared to the control group students, with effect sizes varying from moderate to strong (Lipka, Parker Webster, & Yanez, 2005). Based on this body of evidence, the researchers concluded that utilizing the culturally competent MCC curriculum was efficacious in improving the achievement of elementary school students in mathematics.

Results of another study involving the MCC curriculum conducted by Sternberg, Lipka, Newman, Wildfeuer, & Gigorenko (2006) also suggest that CCI improves Indigenous students' mathematics achievement. This study involved 156 Yupik and non-Yupik 6th grade students in seven communities and three districts in rural and urban Alaska settings. During the study period, treatment and control group students were taught the same mathematics concepts for the same length of time and over the same period of time. Treatment students were taught using the

MCC curriculum while control students were taught using a more traditional, textbook-based approach. Posttest scores for the treatment group students showed statistically significant greater gains compared to control group students, providing additional evidence for the efficacy of CCI in improving students' mathematics achievement.

Again focusing on mathematics, Hilberg, Tharp, & Degeest (2000) conducted a small scale study with twenty-four 8th grade American Indian students who were attending a middle school located on a reservation in the southwestern United States. This quasi-experimental study used random assignment and a non-equivalent control group design. Treatment group students (N=14) were taught using instructional methods that aligned with the Center for Research on Equity and Diversity in Education (CREDE) Standards for Effective Pedagogy and Learning. Specifically, the intervention pedagogy included teacher modeling with small groups of students who engaged in productive dialogue and collaborated in the creation of meaningful products. Control group students (N=10) were taught using instructional methods more typical of mainstream American schools. Analysis of student test scores on proximal achievement tests administered immediately post instruction did not show statistically significant differences in achievement between treatment and control students. However, the treatment group students attained significantly higher scores than control group students on both a mathematics content knowledge retention test and on a survey of attitudes towards mathematics completed three weeks post instruction, thus providing evidence for the efficacy of CCI in improving American Indian students' mathematics retention and attitudes toward mathematics.

Another small-scale study reported by Cardell, Cross, and Lutz (1978) investigated the interactions between peer instruction and mathematics achievement in students in two 6th grade classrooms on the Mescalero Apache Reservation in Arizona. The intervention, which consisted

of peer teaching in small groups by a peer learning leader at mathematics stations, was chosen by the study's authors because of its presumed cultural congruence with the tight knit nature of the Mesaclero Apache community and culture. The treatment and control classrooms were matched for mathematics ability and attitudes toward peer learning. Both groups studied mathematics in two-hour blocks on Monday, Wednesday, and Friday, and the study period spanned eight continuous weeks. Pre- and post-assessment scores showed statistically significant greater gains for the treatment group students compared to the control group students. The authors concluded that the culturally congruent peer learning technique supported increased achievement in the treatment students.

Studies of CC science instruction with Native students. In a study with K-8 American Indian students and their teachers, Grimberg and Gummer (2013) found positive correlations between improvements in teachers' cultural competence and increases in students' science achievement. A total of 62 teachers (27 treatment and 35 comparison) who taught on or near American Indian reservations in Montana and their students participated in the study. Data types analyzed were the teachers' pre/post responses on the Survey of Enacted Curriculum (SEC), a teacher self report survey, and students' pre/post scores on project developed science content tests. SEC results indicated that after two years of participation in a science education professional development project treatment teachers showed statistically significant gains in 1) their confidence in their ability to teach diverse students, and 2) their use of instructional practices that supported students in making connections between science content and real life issues, a culturally congruent practice. Multiple regression analyses found that gains in treatment teachers' scores on these two items accounted for 37% of the variance in their students' pretest to posttest score gains.

A small quasi-experimental study investigating the effects of CCI on kindergarten students' earth science achievement was recently reported by Sievert (2012). Two kindergarten teachers teaching in the same reservation school and their students participated in the study. Both teachers taught their students a rock cycle unit, focusing on similar Western science content taught on a similar schedule, and for a similar duration. The teacher in the treatment classroom included culturally congruent content and methods relevant to her students' tribal cultures, while the other teacher utilized methods and content more typical of a mainstream American classroom. On identical pre- and post-unit assessments, *t*-test analyses showed that the treatment students' average test score gains were statistically significantly higher than those of the comparison students.

Gilbert (2005) reported a quasi-experimental study on the interactions of science achievement and CCI involving ninety-five 5th grade Navajo students from seven classrooms in five schools on the Navajo Reservation in the southwestern United States. Students were assigned to treatment and control groups using a convenience sampling technique. Both groups studied similar science concepts over the same twelve-week period using the Full Option System Science (FOSS) curriculum. All teachers involved in the study received intensive training on the use of the FOSS curriculum in the summer preceding the study. Treatment teachers received an additional two weeks of training on the Navajo Supplemental Science Curriculum (NSSC), which focused on the integration of Indigenous culture, including language, into the science curriculum. Inclusion of elements of Navajo cultural content from the NSSC into treatment teachers' science instruction constituted the treatment intervention. Results showed that the treatment students scored significantly higher on achievement tests and attitude surveys given one week post instruction. Gilbert concluded that the CCI was efficacious in supporting

increased student achievement in and improved attitudes toward science for Navajo students.

Matthews and Smith (1994) investigated the effects of CCI on science and language arts achievement in a research study involving 4th through 8th grade American Indian students. This quasi-experimental study utilized a stratified sample of Bureau of Indian Affairs teachers who were randomly assigned to treatment and control groups. For the ten-week study period, both teacher groups taught the same set of concepts, but treatment teachers integrated culturally congruent content into their curriculum while control teachers employed the more traditional mainstream curriculum. Treatment students scored significantly higher pre-post gains on both attitude surveys and achievement tests compared to control group students. It should be noted that in the research article describing the study, the authors caution readers to consider a number of uncontrolled variables and alternative explanations that could have influenced the results and their interpretations, though just what these are is not clearly explained by the authors.

Though limited, this small body of research begins to paint a picture of the potential that CCI holds for supporting increased science and mathematics achievement in Native American students. Additional studies have been conducted and provide similar evidence for the value of CCI in supporting science and mathematics learning in other groups of ethnically diverse students, providing further testimony on the value of CCI in improving equitable educational outcomes for all students.

Calls for Increasing Research on CCI

The small but promising body of research on the efficacy of CCI in supporting Native American students' achievement described in the preceding section and the increasingly prominent political and social agendas for equitable educational outcomes for all students provide impetus for greater study of CCI. In fact, many scholars in the field of education have

increased their advocacy for CCI related research in recent years (e.g., Lee & Buxton, 2010; Penfield & Lee, 2010; Tyler et al., 2008; Lipka, Sharp, Adams, & Sharp, 2007). Agencies of the federal government and national professional organizations are also emphasizing the need for more research on CCI in recent years (cf. Committee on Equal Opportunities in Science and Engineering [CEOSE], 2009; NCTM Achievement Gap Task Force, 2004). Convened by the National Science Foundation, the CEOSE noted in their 2008 Biennial Report to Congress that

> K-12 teachers need to be taught how American Indian and other minority students learn, in order to better communicate classroom content…American brand schools tend to "initialize" American Indian students, erasing the significance of their culture, language, and perspective—which destroys self-identification and self-worth. This initialization, in turn, negatively impacts the learning process. (p. 32)

The report details the 2009 CEOSE recommendations for an increase in evaluation and research on American Indian education issues including the identification of "elements that are effective in producing successful Native American education programs" (p. 33).

The executive branch of the federal government has also increased its advocacy of CCI research in recent years. In 2004, for example, President George W. Bush signed Executive Order 13336 initiating the *American Indian and Alaska Native Education Act*. This act mandated the formation of a working group of federal agencies whose charge is to support American Indian tribes and Alaska Native groups in improving equitable education outcomes in alignment with NCLB. Significant aspects of the act include the generation of a multiyear study and report of the state of American Indian and Alaska Native education and the strengthening of the capabilities of tribal entities to conduct education research. The study and report, now known as the *National Indian Education Study* (NIES), has been widely released every two years since

2005, providing detailed analysis of AI/AN student achievement. In December 2011, Bush's successor, President Barack Obama, likewise signed Executive Order 13592 - *Improving American Indian and Alaska Native Educational Opportunities and Strengthening Tribal Colleges and Universities* – which reauthorized the NIES and reaffirmed the federal government's commitment to developing and studying research based interventions designed to improve Native students' academic outcomes.

NCTM has also come out strongly in favor of increasing CCI related research. Recommendation #2 on research from the aforementioned AGTF report reads

> NCTM should take a prominent position in support of research related to closing the achievement gap and ensuring that it is addressed at NCTM meetings and conferences at all levels. This research should go beyond suggesting causal relations between underachievement and racial/socioeconomic identity to investigate the social, political, and cultural issues that contribute to causing and closing the achievement gap. (p. 8)

The AGTF report also provides recommendations for a broad research agenda whose ultimate aim is to improve equitable mathematics education outcomes through the systematic study of:

- Race, ethnicity, social class, and language issues pertinent to closing the mathematics achievement gap.
- Characteristics of school curricula that empower students from underrepresented groups to learn.
- Cultural factors that influence mathematics teaching and learning, including analyses of the function of teachers' *worldview* in the process of teaching and learning.
- Characteristics of effective teacher preparation, teacher induction, and mentoring

> programs, including alternative teacher certification programs, in regard to its effect on the mathematics learning of students from underrepresented groups. (p. 9)

The examples described here are but a few of many that provide testimony to the growing recognition of the potential of CCI to improve learning and, consequently, of CCI's increasing significance as a research priority to enable greater understanding of its use to improve equitable education outcomes.

The Challenges of Assessing CCI

Complexity of the construct. The construct of CCI is a complex one. As explicated earlier in this paper, scholars are not in full agreement on the definition of the construct of culture, though there are some elements that are commonly cited in defining it. Culture itself is intangible, but its manifestations are frequently detectable, for example in language, art, and social norms. Culture and its manifestations are also multifaceted, including less tangible elements like attitudes and perspectives, and more tangible elements like dress and food. Measuring intangibles, for example in conducting research, is difficult. Furthering the complexity of the construct is the uniqueness of every culture; for researchers this means that measurement methods and processes must be customized for each cultural context in which they are used in order to provide valid information.

Context specific instruments and methods. Common concerns encountered in instrument development like item relevance, generalizability, and grain size, are often even more challenging when developing instruments and methods for assessing CCI. The context dependent nature of CCI requires that the assessment instruments and methods employed be customized for each context to align closely with the specifics of the prioritized culture and the elements of CCI relevant to that culture. Many of the instruments and methods for assessing CCI currently

available in the research literature were developed by each study's author(s) specifically for their study's context and so cannot be assumed to be suitable for use in other contexts. This customization may mean that instrument items are not relevant (or are even taboo) in different cultural contexts, or conversely, that essential items for other contexts are absent in the instrument.

Generalizability. Generalizability, a characteristic commonly viewed as an asset in instrument development since it extends an instrument's usefulness across contexts, is often difficult to achieve and even undesirable in CCI assessments. Instruments written so as to be generalizable across cultural contexts may be limited in their ability to provide useful information about specific elements of CCI relevant to an individual culture and/or study due to irrelevance of items, omission of items, or inadequate grain size of the data they elicit (LaFrance, 2004). Mainstream instruments designed to be used broadly across science education classrooms that focus on instructional methods such as student centered inquiry and argumentation, for example, may be inappropriate for use in Indigenous cultures in which observational learning with recognized experts (with little accompanying debate) is the traditional preferred method of instruction (Lee & Buxton, 2010; Solanos-Flores & Nelson-Barber, 2001).

Culturally specific assessment methods. The specific methods used in assessing CCI may also have limitations. Strategies appropriate and useful in one culture may not be compatible with the norms of another culture. An example of this would be an interviewer trying to conduct interviews with tribal elders with whom they have no prior relationship. (Nelson-Barber, LaFrance, Trumbull, & Aburto, 2005). Some methods may not be logistically feasible, for example, using online surveys with subjects who may have limited Internet access or limited experience in using the Internet. Incongruencies in language between the assessor and assessee

can seriously limit the use of research methods that are heavily language dependent, such as focus groups and interviews (del Rosario Basterra, Trumbull, & Solano-Flores, 2011; Solano-Flores & Nelson-Barber, 2001). Language differences can threaten the validity of assessments, for example, by hindering the assessee's accurate interpretation of survey items and/or their provision of meaningful responses to survey, focus group, or interview questions. Another example of the limitations of CCI assessments involves the use of observation protocols, for example in conducting observations of classroom instruction. Valid and reliable use of these types of instruments is dependent on the availability of highly trained observers who possess both discipline specific knowledge and deep cultural knowledge in order to make informed observations of often subtle behaviors in a given cultural context. Kirkhart (2005) notes methodology as one of the five justifications of multicultural validity, a term that refers to the authenticity of understandings across intersecting cultural contexts (p. 22). Choosing and structuring appropriate culturally congruent assessment methods is important to the validity of research because it increases the likelihood of generating data that enable accurate inferences about the construct in question.

Bridging cultures. The challenges described in the preceding paragraphs necessitate that a delicate balance be struck in the development of instruments for use in studying CCI. Developing instruments and processes that are appropriate for use in a prioritized cultural context, that gather information important to a particular study and its participants, and that provide information valuable in furthering our understanding of teaching and learning requires what could be called "bridging" knowledge and behaviors on the part of the research stakeholders including (a) deep knowledge of the norms and protocols of the prioritized cultural context as well as of the educational research community; (b) strong relationships and

collaborative partnerships between all of the stakeholders involved in the study; and (c) a thorough understanding of the priorities and needs of each of the stakeholders. The current study is an illustrative example of this type of bridging in action, as detailed in the Methodology section of this paper.

This study was part of a larger research effort designed to evaluate the efficacy of CCI in supporting science achievement in students from several Montana tribal cultures. Such an effort requires both a means to assess teachers' use of CCI and their students' science achievement. These two data sets can then be analyzed to determine if they are correlated, thereby providing evidence for the efficacy (or lack thereof) of CCI in supporting student learning. Considering the many potential types of CCI assessment challenges discussed earlier in this section, for this study it was deemed essential to generate a CCI assessment instrument and development methods that were designed collaboratively with representatives from all of the project's stakeholders to be specifically compatible with the contexts in which they were to be used. While the instrument and methods may be compatible with other American Indian contexts to some extent, their use outside of the Montana tribal cultures for which they were designed will likely require some customization for the specific culture with which they are being used. Even given the potential limitations to generalizability discussed, it is believed that the instrument and methods used in the present study will have some value in serving as a model for instrument and methods development that are useful to others attempting this type of work.

Engaging in Research with American Indian People.

In addition to the issues discussed in the previous section regarding the challenges of developing instruments and methods for assessing CCI, researchers working with American Indian people will likely encounter other issues particular to the cultural community with whom

they collaborate. These issues may stem from many, often overlapping, sources including stakeholders' political statuses, worldviews, epistemologies, norms of behavior, and values. For example, the worldviews held by many Indigenous cultures are frequently incompatible with the perspectives of Western research paradigms (Aikenhead & Michell, 2011; LaFrance et al., 2012). One's philosophy can have profound and far-reaching effects on the entire research process, from generating the research questions, to designing the study, to gathering and interpreting information and reporting findings. Likewise, epistemological differences may mean that the information that the tribal community considers important for the study does not align with the information that researchers from outside the tribe consider valuable to furthering the research.

Issues of tribal sovereignty and intellectual property rights must also be considered when conducting research with tribal communities. Non-tribal and tribal research partners are prudent in proactively working together to resolve questions of consent for participation in the research, ownership of knowledge and products generated by the research efforts, and permission to use the knowledge generated, for example. These types of questions have ethical as well as potentially legal ramifications due to the sovereign status held by many American Indian tribes.

Other research related issues that can arise in partnering with tribal communities are more logistical in nature. Accessibility to tribal elders, for example, may be hindered due to their living in remote locations, language differences between the researchers and the elders, and differences in social norms for interaction and communication. Disparities in communication styles and/or language can be particularly problematic, confounding the research process by leading to misinterpretations of ideas, the use of inappropriate methods, and/or inaccurate data analyses, thus reducing the validity of the research and potentially offending research partners (Hall & Ward Hood, 2005; Quigley, 2001).

Researchers should be cognizant and respectful of the research context, to honor the people and the reciprocal relationships in which they are engaged, to improve the quality and validity of the research, and to ensure beneficial outcomes for the people with whom they are collaborating (Nelson-Barber et al., 2005). A selection of issues relating to Indigenous research that was particularly significant in conducting this study is discussed in the following paragraphs. This is by no means an exhaustive treatment of this subject, which would require a full dissertation in itself, but rather a set of illustrative examples for the readers' knowledge.

History of hegemony. An overarching contextual factor influencing research throughout Indian Country is the historically hegemonic relationship that exists between European Americans and American Indian people. In the eyes of many early European immigrants to this continent, Native people were often seen as less civilized, less intelligent, less advanced people, whose assimilation would improve their lives and also enable the takeover of their resources by non-Indians. History provides many examples in which research and evaluation, including education research and evaluation, have been used to subjugate American Indian people by providing "evidence" to justify their assimilation and even their cultural genocide (Chawla-Sahota, 2010; LaVeaux & Christopher, 2009). Even with their sovereignty guaranteed by treaties with the federal government, Indigenous peoples of the United States are, out of necessity, continuously working to prevent the erosion of their sovereignty as nations and their rights to self-determination. This hegemonic legacy has contributed to an attitude of mistrust in some American Indian people toward working with researchers, especially when the researchers are cultural outsiders. In order to conduct research with American Indian people, researchers must make concerted efforts to establish credibility as just and equitable people worthy of trust, to develop truly collaborative relationships with tribal community members, and to work

collaboratively with stakeholders to design and conduct research that is culturally congruent and that addresses issues important to tribal communities (LaFrance, Nichols, & Kirkhart, 2012). In the present study, the strong relationships between several of the key stakeholders living in each community were valuable to engaging in valid research in their home settings. Because the research design included stakeholders from multiple communities across Montana, relationship building between stakeholders across settings was an important priority for the project partners and significant attention was given to attaining that objective throughout the design and implementation of the project activities. Specific examples of these types of efforts are described in detail in Chapter Four of this paper.

Cultural norms and values. Another issue of importance for those engaging in research in Indian country is the observance of the cultural norms and values for the specific research context (LaFrance & Nichols, 2004; Quigley, 2001). These should be known and practiced when interacting with Native peoples out of respect for the community, to foster the development of trusting relationships and credibility with stakeholders, and to improve the gathering and interpretation of valid, relevant and useful information. This is a challenging charge, since there are hundreds of American Indian tribes in the United States and every tribal culture is unique. The researcher would be mistaken in assuming that knowledge of one culture can be applied across cultural settings. Working with the Apsåalooke, or Crow people of Montana in this study, for example, requires extensive knowledge of their clan system, of people's relationships within their clan, and of the norms for interacting with specific members of a clan (G. Whiteman, personal communication, May 25, 2008). Ignorance of these norms may result in offensive interactions and little or no exchange of reliable information. The culturally competent researcher must assume responsibility for becoming knowledgeable about their specific research

contexts and must realize prior to beginning the research the level and type of effort it will take to develop their cultural competence. The development of cultural competence in the project stakeholders was a central focus of the professional development project on which this study focused, realized in large part through the sharing of cultural knowledge and experiences as an integral part of nearly every project activity. This enabled the research team to become more familiar with the local cultural norms, values, and worldviews, thus ensuring greater cultural congruency and validity for the research.

Keeping in mind the tribal specificity of cultural norms and values, the literature provides some guidance on what are considered to be common norms, values, and protocols for American Indian people in general. Sources vary somewhat, but generally a list of these items include the following:

- Community centeredness
- Cooperation, collaboration
- Respect for people
- Non-interference/Respect for autonomy
- Family/Relationships
- Concise expression/a\Active listening
- Fluidity of time
- Time for reflection
- Sense of humor
- Harmony with Nature/Bond to place
- Centrality of spirituality/Monism
- Respect for ceremony

- Honesty
- Bravery
- Sharing/Generosity
- Humor
- Open mindedness/Accepting of multiple truths
- Holistic perspective
- Value for practical knowledge
- Observational learners
- High context learners
- Humility
- Fine arts perspective (stories, symbols, metaphors) (Aikenhead & Michell, 2011; Gilliland, 1999; LaFrance et al., 2012)

The culturally congruent researcher, as well as the culturally congruent teacher, will want to keep these in mind when working with Native people, while also remembering that tribal cultures and individual people within tribal cultures vary in their norms and values, so assumptions about culture are not always safe and care should be taken to become aware of the nuances of culture in each context.

Relationships, community and participatory methods. Interpersonal relationships, a strong sense of community and the recognition that knowledge is bound to experience and should be used to benefit community are fundamental values in many American Indian cultures (LaFrance et al., 2012). The legacy of exploitation of Indigenous peoples coupled with the cultural values prioritizing relationships and community make the use of participatory research methods, in which all stakeholders have a voice in the conversation and decision making, a

research paradigm of choice for working with tribal communities (LaFrance, 2004; LaFrance & Nichols, 2004; Quigley, 2001). Participatory methods center on the co-construction of knowledge by all stakeholders and dictate collaborative examination and consensus in delineating the who, what, where, when, why and how of the research. These methods encourage the identification of research topics that are embraced by all stakeholders as valuable to their communities and compatible with their cultures. Inclusive discussions to identify and negotiate the types of information that stakeholders consider to be knowledge of value to the study as well as how that information is best collected and interpreted are other central features of the participatory paradigm. The non-hierarchical, equitable approach of participatory research methods stands in stark contrast with much of the history of research conducted with Indigenous people in which they have been merely subjects to be studied and in which findings have frequently been used in ways that were destructive to tribal cultures. The current study utilized participatory techniques extensively in the design and validation of the CCIS. More details on the specific methods used are available in the Methods section of this paper.

Qualitative methods. Scholars have also noted that qualitative methods may be more appropriate in conducting research with tribal people (cf. Denzin & Lincoln, 2008; LaFrance, 2004). The detailed, contextualized story that can emerge using qualitative methods such as case studies, interviews, and focus groups, for example, may be more valuable and informative for tribal people whose cultures embrace oral history and storytelling. Particularly given the uniqueness of each tribal culture, the richness of the information gained using such methods may be more valuable to the research in helping paint with fidelity an ample picture of the research context and outcomes. While the generalizability of findings is normally considered advantageous since it enables their broader relevance and application, the limitations on

generalizability sometimes encountered with qualitative findings may be less problematic in Indian country where it is recognized that tribal cultures are unique and that knowledge generated is important within the specific context (LaFrance et al., 2012). Indeed, generalizing across cultures and settings, while maintaining culturally validity, is difficult and often undesirable in research and evaluation with Indigenous peoples.

Process validity. Attention to context is vital in designing and conducting research with diverse communities from both a technical (to improve validity) and ethical (to honor the cultures of the research participants) standpoint. Process validity refers to the extent that the methods employed in conducting research are adequate, sound, and appropriate for the study's context such that they enable ongoing learning for all of the research partners (Anderson & Herr, 1999; Anderson, Herr, & Nihlen, 1994; Herr & Anderson, 2005). It stipulates attention to the relationships, cooperation, and co-learning cultivated between research partners (Cullen, 2008). Writing on the topic of action research, Anderson et al. (1994) state that methodological adaptations that researchers employ to accommodate their specific study context and that serve to foster and capture the flow of action in that context contribute to process validity. Specific examples of research strategies identified by Anderson and Herr (1999) that support process validity include the development of relationships with research participants, the equitable inclusion of multiple perspectives, methods, and data types to allow triangulation of data and avoid bias, the collaborative identification of what counts as evidence in a specific context, and the dissemination of knowledge generated in the study. Process validity prioritizes a reflexive cycling back by the research partners to reexamine the assumptions underlying the framing of the problem. By attending to the proper use of research process, process validity contributes to outcome validity (Herr & Anderson, 2005) and arguably contributes to construct validity.

Doing things the right way. In more recent and related work, LaFrance, Nichols, and Kirkhart (in press) discuss the importance of doing things *the right way* in working with Indigenous communities, which in turn ensures the trustworthiness (or validity) of a study's findings. Emphasis is placed on developing relationships, observing cultural protocols and using culturally compatible methods, and giving back to the community by sharing and utilizing the information that was gathered through collaborative study. Similar to the notion of process validity, these authors note that validity grows out of attending to doing things the right way.

Relational accountability. Related to these ideas of process validity and doing things the right way are the concepts of relational accountability and the three r's of Indigenous research (respect, reciprocity, and responsibility) described by Wilson (2008). According to Wilson, to have relational accountability means that the research methodology "needs to be based in a community context (be relational) and has to demonstrate respect, reciprocity, and responsibility (be accountable as it is put into action)" (Wilson, 2008, p. 99). Methods like personal narrative, participatory action research, talking circles, and storytelling are respectful to Indigenous people because they are relational and so fit within an Indigenous paradigm for doing things the right way. Doing things the right way, conducting research with relational accountability to the specific research context, supports the validity of the research outcomes.

Community based validity. Likewise, Kovach (2009) explains that in an Indigenous research paradigm, validation emanates from the community participating in the research. Research must be conducted in the right way for *each* community, in respectful adherence to an Indigenous research paradigm. Research methods that are congruent with Indigenous epistemologies and decolonization methodologies include storytelling, oral history, unstructured interviews, open-ended conversations, sharing circles, and similar methods compatible with local

tribal protocols that give representation and voice to research participants. To honor community accountability and as a check for validity, Kovach describes her efforts to regularly share her findings with Indigenous graduate students, researchers, and other community members. It is again emphasized that researchers must be aware of and observe the cultural protocols and epistemologies for their specific research contexts and realize the importance of relationship to Indigenous people and Indigenous research paradigms.

Validity and context. In a related but more broadly argued position, Moss and her colleagues (Moss, 1998; 2005; Moss, Phillips, Erickson, Lather, & Schneider, 2009) call attention to the significance of context to validity and to the influences of sociocultural history and context in conducting and interpreting research. Existing definitions of validity that privilege one perspective are unfavorably criticized. The authors instead suggest that validity is inseparably influenced by a broad range of perspectives that are shaped by culture and context, and make the case for a reconsideration of validity to include additional perspectives (Moss et al., 2009). Efforts by agencies to rigidly define the parameters for rigorous education research whereby the quality of research findings is substantiated are also criticized. It is argued that such narrow boundaries for research again disenfranchise other perspectives and research paradigms. Instead, discourse that considers the legitimacy of diverse perspectives and the many forms of research that exist is advocated. Such discourse will enable researchers to learn from each other and avoid the generalization of ideas in an attempt to encompass multiple perspectives but, in doing so, misrepresents them.

This literature review shared some examples of cultural issues related to research that should be respectfully addressed when collaborating with American Indian people. Acknowledgement of the importance of culture and context in designing and conducting research

is increasing in many types of social science research. Engaging in effective research with American Indian people, as with any group of people, requires that researchers embrace the importance of developing their cultural competence, put forth the effort it may require to attain cultural competence for the specific group of people with whom they wish to collaborate, and apply their knowledge in order to conduct their research in an equitable, ethical, valid, and symbiotic manner.

In describing the literature relevant to this dissertation study, Chapter Two raised many issues relevant to conducting research in tribal communities. Chapter Three begins to describe the research design and methodologies used in this dissertation study, which incorporated many of the recommendations discussed in Chapter Two. Further details on the process that emerged from the study and a discussion regarding the interplay of the study's context and its influence on the methodologies and process are provided in Chapters Four and Six,

CHAPTER THREE – RESEARCH DESIGN AND METHODOLOGY

This study was conducted as part of a National Science Foundation funded Math and Science Partnership (MSP) that partnered three institutes of higher education (IHEs) with five tribal communities, dozens of K-12 schools, and over one hundred teachers teaching on or near American Indian reservations in Montana. American Indians students, who comprise 12% of the student population in Montana, experience persistent underachievement compared to their White peers across academic subjects, as previously described in this paper. In response to this, the MSP was designed to increase K-8 American Indian students' science achievement by: (a) deepening K-8 teachers' science content knowledge; (b) improving teachers' knowledge of and proficiency with using CCI; (c) developing partnerships between IHE science faculty, tribal community experts, and K-8 teachers; and (d) increasing teachers' leadership roles in strengthening science education in their schools. The project leadership team consisted of IHE science and science education faculty, professional developers, and cultural experts from the five tribal cultures involved in the study. Project activities included ten-day summer institutes, academic year courses, and three- to five-day culture camps in which teachers worked side by side with cultural experts in natural settings. Teacher membership in the project lasted for three years. Two cohorts of teachers participated in the project over a five year period.

As a part of the MSP evaluation and research efforts, representatives from all partner groups worked collaboratively to design and validate the CCIS, a 41 item instrument that operationalizes culturally congruent instruction in terms of content, pedagogy, and instructional environment for K - 8 science education for the five tribal cultures in the partnership. Cultural protocols were carefully considered and practiced throughout the development process. Both qualitative and quantitative research methods were used to design the CCIS, to begin to

characterize its nature, and to gather evidence of its validity. A description of the methods employed in this study to address each research question is provided in this chapter.

Research Question #1: What is a culturally congruent process for developing a valid instrument for assessing the use of CCI in teaching science with Montana American Indian students?

The process referred to in Research Question #1 includes all of the steps taken and strategies used in developing the CCI assessment instrument, from defining and operationalizing the construct, to the iterative review and refinement of drafts of the instrument, to the validation of the data it generates. This aspect of this two part dissertation study focuses on the evolution of the development process and the products that were generated en route. The processes involved the significant use of culturally congruent strategies and participatory methods, engaging all relevant stakeholders in the design of the instrument.

The development of the Revised CCIS followed a series of standard steps for instrument development. In line with recommendations made by Hinkin (1998), for example, the development process began with the establishment of a theoretical framework for CCI. The framework was generated using the common techniques of in depth literature review and multiple and substantive conversations with stakeholders to help define the construct. Both methods were used to determine the domain of content and to generate items, the next steps in the instrument development process. Prototype instruments were produced using the findings from these methods. Items were included in the instruments only if all stakeholders agreed that they were accurate and appropriate for the specific cultural contexts for which the instrument was intended. Prototype instruments were reviewed iteratively by multiple and diverse stakeholders throughout the process to help ensure content validity. Additional details regarding

the generation of instrument items are found in Chapter Four.

The second prototype instrument was administered to 128 teachers in the study and the data were subjected to analyses of internal structure. Results of these analyses were used to revise the instrument through item reduction, the replacement of the response scale with finer grained scales, and a reorganization of the instrument to improve its clarity and user friendliness. Confirmatory factor analyses were conducted on data generated through the administration of the revised instrument and various types of evidence of validity were gathered. These are discussed in Chapter Five.

While the development process did in many ways mirror a series of steps typical for instrument development, what set the development processes for the Revised CCIS apart from standard instrument development processes were the culturally congruent accommodations made to ensure the equitable and significant contributions of all stakeholders, which helped promote the content validity of the instrument. The development process for the instrument will be described in depth in Chapter Four.

Research Question #2 – What is the technical quality of an instrument for assessing the use of CCI in teaching science with Montana American Indian students?

Quantitative methods were employed in this study to generate information about the nature of the CCIS, including evidence of its validity. This aspect of the research involved analyses of data collected through the administration of the CCIS and several other instruments. The additional instruments' data analyses contribute to evidence of validity for the CCIS. The methods for data collection and analyses are detailed in the following paragraphs.

Sample selection and characteristics. The study employed a quasi-experimental design, involving two cohorts of non-randomized treatment teachers and matched comparison teachers,

all of whom were teaching in the project's K-8 partner schools. Treatment group teachers were participants in the MSP professional development project described earlier in this chapter. The project's intervention activities spanned five years; each cohort was formally active in the project for three years with one year of overlapping membership in Project Year Three, during which both cohorts were active.

The project's design and delivery were led by the three IHE partner sites located in western and south central Montana. Each site then worked closely with partner schools and tribal communities in their respective surrounding area. The overarching goal of the MSP was to increase American Indian K-8 students' science achievement, therefore the project's partner schools were located on or near American Indian reservations in Montana where American Indian student enrollment was high. The specific reservations involved – the Flathead, Crow and Northern Cheyenne – are the homes of five distinct tribal cultures. There were also a number of partner schools, largely associated with one of the IHE partners on the western side of the state, which bordered reservations. While the faculty and staff from the three IHE partners worked closely in the design and delivery of the intervention (identifying common learning objectives, sharing PD activities, and designing common instruments for evaluating project impacts, for example) it was also recognized that this mélange of partner schools, tribal communities and IHEs enveloped a diverse set of contexts that required some customization of project activities by site to meet the distinct needs of their partners. This was particularly true in regards to the cultural congruence aspects of the project since the five tribes involved – the Northern Cheyenne, Crow, Pend d' Oreille, Kootenai, and Salish – each possess different cultural traditions, norms, histories, and land bases, all of which were integral to the project's success.

Treatment teachers ultimately applied for project membership voluntarily. Project staff

from each of the three IHE sites began the teacher recruitment process by approaching local school administrators to discuss the MSP project and to ask for recommendations of teachers in their district who they believed would be good candidates for the MSP, i.e., for grooming as science teacher leaders per the project's objectives. As a result of these discussions, about half of the treatment group applicants were initially identified by their administrators and subsequently approached to apply for project membership. The remaining treatment teacher applicants heard about the project largely through word of mouth from their peers and opted to apply on their own. Teacher applications were submitted to project leadership committees at each IHE partner site; committee members at each site then collaboratively vetted them and chose the treatment teachers for their site's cohorts based on a predetermined set of common criteria. The criteria for membership selection had been defined proactively by the project's leadership team during the development of the MSP grant proposal and were based on teacher traits found in research studies to correlate with high teacher leadership potential, such as number of years teaching, respect of their peers, and previous leadership roles (York-Barr & Duke, 2004).

Similarly, comparison group teachers were recruited by the project staff at each IHE site working in conjunction with school administrators to identify teachers that "matched" with the treatment group teachers. They were chosen based on their similarities for relevant characteristics such as total number of years teaching, number of years teaching science, grade levels taught, their ethnicity and gender, their student demographics, the location of their school (rural, urban, on reservation, or off reservation, e.g.) and the number of college science courses each had completed.

Similar to the national statistics cited in Chapter One, the overwhelming majority of the treatment and comparison group teachers (84%) were White (See Table 2). Eighty-six percent

were female. Table 3 shows the locations of the partner school by categories. Fifty-four percent of teachers were teaching in schools located on American Indian reservations. The others taught in schools bordering reservations, or rural or urban schools within forty miles of a reservation. Table 4 provides student demographics details for partner schools. American Indian student enrollments in the partner schools ranged from 1 to 100%. One in five teachers taught in schools with American Indian student enrollments exceeding 80% and half were teaching in schools in which American Indian students constituted at least 41% of their student enrollment. Seventy percent of teachers had taught science for at least six years (See Table 5).

Table 2

Self Identified Teacher Ethnicity

Ethnicity	Number of Study Teachers	Percentage of Study Teachers
Asian American	3	2
Mixed American Indian/Alaska Native/White	2	2
American Indian/Alaskan Native	16	12
White	107	84

Table 3

Location of Treatment and Comparison Schools

Location of School	Number of Study Teachers	Percentage of Study Teachers
On an American Indian reservation	69	54
In a border town, serving American Indian students	11	9
Rural - off reservation	17	13
Urban/Suburban	31	24

Table 4

Study School Demographics

Percent American Indian Student Enrollment	Number of Teachers Working in this Category	Percentage of Teachers Working in this Category
1 to 20	49	38
21 to 40	15	12
41 to 60	8	6
61 to 80	31	24
81 to 100	25	20

Table 5

Total Years Teaching Science

Total Years Teaching Science	Number of Study Teachers	Percentage of Study Teachers
Less than 1 year	6	4
1 to 2 years	11	9
3 to 5 years	22	17
6 to 8 years	25	20
9 to 11 years	14	11
12 to 15 years	12	9
More than 15 years	38	30

Data collection. The second prototype CCIS was administered via hardcopy in spring of 2007 at preliminary meetings of the Cohort 1 treatment teachers at their respective sites, prior to their participation in PD intervention activities. Findings from this administration were used to improve the instrument, resulting in the Revised CCIS.

During this study the Revised CCIS was administered a series of five times to treatment and comparison group teachers, beginning in spring 2008 and then annually each year through spring 2011. Cohort 1 treatment teachers completed the CCIS four times. Cohort 2 treatment teachers joined the project in spring of 2009 and so completed the Revised CCIS annually through 2012, also for a total of four times. For the first three administrations, the surveys were distributed in hardcopy form and scoring and data compilation was conducted by hand; the last

two administrations were online, utilizing the web based application SurveyMonkey. It was anticipated that the administration of the survey online would assure timely compilation of the data. By the date of the fourth administration, all of the treatment teachers had had extensive experience in the use of online learning and resources due to their membership in the project; no difficulties in the use of the online survey were reported. Treatment teachers worked on the survey as a group during a time set aside during the professional development activities, increasing the likelihood of consistent administration. Comparison group teachers were encouraged to complete the survey during the proctored session in which they completed the science content test, though a small percentage completed it at a different time. Data from the fourth administration were downloaded and submitted to evaluators as an Excel file from the web site. Maximum sample size was 128 and included both treatment and comparison group teachers; sample sizes for each analysis conducted varied depending on the type of test performed.

Data analyses employed. To generate information about the nature of the CCIS, a series of tests was conducted on the data collected from the administration of several instruments. These tests are described in the following paragraphs using a two pronged approach – those that provided evidence of validity and those that specifically addressed reliability. Because reliability is an aspect of validity, some tests served to provide evidence of both.

Tests of reliability. Reliability is defined as the degree to which an instrument generates consistent results; it is an aspect of validity. The greater evidence a researcher has of an instrument's reliability, the more they can be confident in the instrument's ability to produce the same results when administered under repeated, similar circumstances. There are several forms of reliability and a number of ways to generate evidence of reliability. In this study, the CCIS

was tested for temporal stability and internal consistency. The tests used are described below.

Test/retest for temporal stability. Test/retest analyses were conducted on data obtained from successive administrations of the CCIS as a means of gathering evidence of the temporal stability of the instrument and the data generated from its use. Temporal stability is an aspect of instrument reliability; if analysis of the data from two administrations of the instrument occurring within a short time frame and under consistent conditions show significant positive correlations, this provides evidence of the temporal stability of the instrument's internal structure.

To gather data for a test/retest analysis, 68 teachers (35 treatment teachers from Cohort One, 26 comparison teachers who would soon change roles in the project and become Cohort Two treatment teachers, and 7 comparison only teachers), completed two administrations of the CCIS, one occurring in early May 2009 and the other six weeks later, in mid June 2009. These administration dates coincided with the end of Cohort One's second year of membership in the PD project, and with the beginning of the first year of Cohort Two's membership, prior to their participation in the project's intervention activities. No intervention activities occurred in the interim period between the CCIS test and retest administrations. It was predicted a priori that there would be significant positive correlations between the first and second administration scores for each group. Analyses were conducted to calculate Pearson's Correlation coefficients for the overall CCIS test/retest scores for the whole group and for each of the three teacher subgroups. Because the survey underwent revision from the second prototype of the CCIS used in the spring test to the Revised CCIS used in the retest administration, analyses were conducted only on the 29 items common to both survey versions and a common four point scale was used. The data resulting from the administration of the Revised CCIS, which utilizes a six point scale,

were converted mathematically to their four point equivalents using the formula (N-1) x 3/5 + 1 where N is a six point scale datum point.

Factor analyses. Exploratory and confirmatory factor analyses (EFA and CFA respectively) were conducted on the 2009 - 2012 data sets to investigate potential interrelationships existing between the survey items, uncover evidence of latent variables, and to gauge the instrument's internal consistency and structural stability. An instrument's internal consistency and stability contribute to its tendency to provide reliable data. Uncovering the relationships between an instrument's items also helps to delineate the factors, or scales, that comprise an instrument. Hypothetically, each factor is comprised of a subset of survey items that can be identified through factor analysis, in this case using iterative tests that calculate correlations for each item with every other item. Those items that tend to "load" with higher correlations on each other together comprise a factor. The subset of items that make up an individual factor are believed to tap a specific aspect of the larger construct that the instrument as a whole is designed to address. In this study, extraction of each factor was accomplished using Principal Components Analysis and Varimax rotation. Using these methods in an EFA, items with high factor loadings were identified to form the initial additive factors believed to be underlying the overall instrument. Measures of internal consistency are a product of factor analysis, including Cronbach's alphas for the scales and inter-item and overall total item correlations. In this study, conducting factor analyses over time using three successive years of data provided two follow up opportunities for CFA, potentially confirming the factors identified in the EFA. Obtaining similar results over time provides evidence of the instrument's internal consistency stability.

Inter - scale correlations. Additional paired tests of correlation were conducted to gather

evidence of the instrument's internal structural stability, exploring the correlations between the survey's four scale scores and with the overall scores for each year's data from 2009-2011. It was predicted a priori that each year the four scales would exhibit positive correlations with each other and with the overall scale score.

Tests of validity. Validity is the extent to which the data generated by an instrument enables accurate inferences about the construct it is intended to measure (Web Center for Social Research Methods, 2006). According to unitary theory, construct validity is the overarching framework for all other types of validity and all evidence of validity is evidence of construct validity (Messick, 1990).

Content validity. Content validity is defined as the relationship between the content of an instrument and the construct it is designed to measure. In this study, evidence of content validity was accrued through the extensive qualitative processes employed in developing the instrument such that the construct was well represented in every aspect of the instrument including its items, format and scale. The deliberation and care taken in involving multiple stakeholders in an iterative and culturally congruent development process enables confidence in the content validity of the inferences drawn from the instrument.

Tests for concurrent validity. To gauge project impacts on teachers and as a means of gathering evidence of the instrument's concurrent validity, ANOVAs and *t* tests for differences in pre- and post-CCIS overall scores and scores for each factor were conducted. These were conducted for each data set from 2009-2011 for Cohorts One and Two treatment and comparison teachers. Four a priori predictions were made relevant to these tests: 1) The overall baseline scores for the treatment groups and comparison groups would show no statistically significant difference in their means; 2) comparison group scores would show no statistically significant

change over the three years of data collection; 3) after joining the project, treatment teachers' mean scores would show statistically significant positive differences compared to comparison group teachers and those differences would increase over time; and 4) Cohort One Treatment teachers' mean scores would initially be statistically significantly higher than Cohort Two Treatment teachers' mean scores due to their additional time engaged in project activities. Differences would decrease over time as Cohort Two treatment teachers' increased their time engaged in project activities.

Tests for convergent validity. Two additional surveys identified in the research literature as instruments designed to assess CCI were also administered to treatment and comparison group teachers one time concurrently with the Revised CCIS in spring of 2010. These data were then analyzed for correlations with the CCIS overall scores as a means of gathering evidence of the Revised CCIS's convergent validity, a type of criterion related validity. Convergent validity is based on the idea that two or more constructs (or measures of said constructs) that should be related to each other are in fact related, as shown by correlations between the data they produce. The Teacher Multicultural Attitude Survey or TMAS (Ponterotto, Baluch, Greig & Rivera, 1998) and the Culturally Responsive Teaching Self-Efficacy Scales or CRTSES (Siwatu, 2006) were chosen from the review of instruments described in Chapter Four. They were selected for their relevance to the current study, in terms of the specific constructs they were designed to evaluate, and for the quality of the evidence of validity provided by their authors. The constructs addressed in the TMAS and CRTSES, multicultural attitudes and teacher self-efficacy respectively, are thought to be important traits held by culturally competent teachers. The theoretical frameworks used in designing each instrument, described by their respective authors in the papers cited above, provide evidence of this. Both instruments were shown in previous studies to have high

internal reliability. The authors of the TMAS also reported a test-retest Pearson's correlation of .80. In the current study, tests were first run to affirm the internal reliability of each instrument prior to conducting correlations analyses for soliciting evidence of their convergent validity with the Revised CCIS. It was predicted a priori that there would be statistically significant positive correlations found between the scores of the three instruments.

Additionally, the correlations between the Revised CCIS scores with the TMAS and CRTSES scores for 2010, since the three assessments were completed concurrently, provide a second form of evidence of concurrent validity. It was hypothesized a priori that each of the instrument scores would positively correlate with the others' scores, given that the TMAS and CRTSES were designed to tap constructs believed to be positively associated with CCIS.

Tests for Discriminant Validity. Tests for discriminant validity were also conducted on data sets to gather further evidence of validity for the Revised CCIS. Discriminant, or divergent, validity, is a type of criterion related validity that examines whether two constructs that are thought to be unrelated in fact provide evidence that they are unrelated, i.e., correlation tests conducted on the scores from two instruments administered to the same group are expected to provide evidence of low or no significant correlation between the scores on the two instruments. Scores from a science content test taken by all treatment and comparison teachers in spring 2010 were chosen for the discriminant validity test. Paired correlation tests were run comparing individual teacher's content test scores with their Revised CCIS overall and factor scores from the 2009-2011 data sets. It was predicted a priori that no significant correlations, positive or negative, would be found between teachers' scores for the two instruments.

This chapter briefly described the methods that were used in this study to develop and collect evidence of the validity of the Revised CCIS. Chapter Four describes the methods used in

the instrument development processes in more detail as part of the results found in investigating Research Question #1.

CHAPTER FOUR – REVIEW OF THE RESULTS FOR RESEARCH QUESTION #1

This chapter describes the results of the study relevant to Research Question #1 *What is a culturally congruent process for developing a valid instrument for assessing the use of CCI in teaching science with Montana American Indian students?* The nature of Research Question #1 requires that this results chapter describes the framework of the study that emerged, the development process that evolved, and the evidence that the process was culturally congruent. The chapter includes elements of methodology in describing the development process, and describes the theoretical underpinnings that influenced the design of the process by laying out the details of context that affected it.

The chapter begins by describing the preliminary framework on which the instrument was built. Next, it explains in detail the qualitative processes used to define and operationalize CCI for the five prioritized tribal cultures. Then it describes specific traits of the American Indian cultures in the study that were addressed in the development process and that contribute to its cultural congruity. Finally, it provides evidence in the form of the project partners' behaviors and feedback that testify to the quality of the development process.

Acknowledging the Specificity of Tribal Cultures

While the elements of CCI are often found described in the literature for a cultural group overall, it must also be recognized in developing a CCI assessment instrument that a wide range of variation can and commonly does occur across subgroups and between individuals within a cultural group. The study described in this paper involves five distinct American Indian cultures, all indigenous to North America. In acknowledgment of the uniqueness of each tribal culture, the work described herein is constrained to and specifically prioritizes the elements of CCI identified as common to these five tribal cultures – the Bitterroot Salish, the Kalispel (or Pend d'Oreille),

and the K'tuxna band of Kootenai, all of northwestern Montana, and the Northern Cheyenne and Crow people of eastern Montana.

Within these five American Indian cultures there exist substantive cultural variations – between the Salish and Kootenai peoples for example, who have shared the same reservation for over 150 years but who have distinct cultural differences on many levels. An example of a cultural difference relevant to CCI is the seasonal tradition of Coyote Story telling observed by the Salish people residing on the Flathead Reservation. Coyote Stories are only told by the Salish people in the winter months, in keeping with a generations-old tradition. Stories are "put away" in a special ceremony once winter ends, not to be told again until the following winter. The people of the Kootenai band of the Flathead Reservation, on the other hand, do not have seasonal restrictions on the telling of Coyote stories. This difference in cultures should be accommodated by teachers in their practice and by researchers when assessing culturally congruent instruction.

The development of the instrument in this dissertation study was keenly focused on capturing cultural specificity in the instrument's items through the use of participatory methods to generate them. The specific elements of CCI were first broadly identified through a review of literature that focused on American Indian education. Then, in recognition of the specificity of each of the five American Indian tribal cultures involved in the project, input from members of the five tribes was gathered and used to custom design the instrument items for the five communities. Each draft of the instrument was reviewed by tribal experts, as well as other members of the development team. Those items that members of all five tribes agreed were essential to CCI comprise the instrument.

The Three Element Framework

For the purposes of this study, a preliminary framework for CCI was proposed by the author in a project meeting that included IHE faculty, professional developers and tribal experts. The model delineates CCI into three major interacting elements of content, pedagogy, and learning environment, as depicted in Figure 1. These three elements are commonly used in the literature in discussions of CCI (cf. Lee & Luykx, 2006). The team agreed that the three elements provided an efficient and comprehensive encapsulation of the main focuses of teaching practice. By content, it is meant the culturally congruent topics that are addressed in the curriculum, which in this study's contexts includes tribal oral history, Indigenous science knowledge, and contemporary and historical issues related to science (both tribal and those of the larger society of which students are members). By pedagogy, it is meant the specific types of culturally congruent instructional methods used by teachers and students in teaching and learning, particularly those that build on students' traditional ways of knowing and that are congruent with their cultural norms. By learning environment, it is meant those things that contribute to a culturally congruent classroom atmosphere including the resources available to students and teachers and the interactions and power dynamics that are reflections of the classroom norms and contribute to the classroom ambience.

The three-pronged framework consisting of content, pedagogy, and environment was used as a preliminary framework for the development of the instrument that was the subject of this study. These three elements provided adequate inclusiveness and a parsimony that is a desirable attribute of a technically correct instrument – i.e., too many or two few categories and items can have deleterious effects on the validity of the instrument and its ease of use. It was found that the three-pronged approach was also effective in the collaborative work done with the

project partners who together generated and honed the survey's items and format; nearly all of the items that emerged from the collaborative discussions fell into one or more of the main three survey categories. Any unassigned items were put in a fourth miscellaneous category because no common theme emerged from them. The instrument development process is now described in more detail.

Defining and Operationalizing CCI for Five Montana Tribal Cultures

Evaluating culturally congruent instruction can be a complex undertaking on many levels, as discussed in earlier sections of this paper. The design of the instrument in the current study was likewise complex, for example, in deciding what aspects of CCI should be assessed, in operationalizing CCI for the prioritized cultural contexts, and in identifying by what means it should be assessed. After a review of existing instruments in the research literature by the author of this study and in-depth conversations with project stakeholders and assessment specialists, it was decided by the project leadership team that one form of evidence of teachers' CCI, the one that is the subject of this paper, would be collected through the administration of a survey in which individual teachers would self report the frequency with which they employed specific culturally congruent practices in their science instruction. These practices would address the three elements identified by the project leadership as key to CCI – content, pedagogy, and instructional environment. While the self report survey is one way to evaluate CCI, the project leadership team recognized that this method has limitations in that it does not provide qualitative information about the nature of CCI occurring in teachers' classrooms, only the frequency of specific practices. To compensate for this limitation, other types of data about teacher instructional practice were also collected by the project partners, including curriculum artifacts, classroom observations, and teacher reflections on their work. These other types of data are not

considered in this dissertation study because they are beyond the scope of the research questions that are the focus of this study.

Review of existing CCI instruments. Initially, a survey of existing instruments designed to assess culturally congruent instruction, attitudes, and beliefs was conducted by the author of this dissertation study to determine if an appropriate instrument was already available that could be used in studying CCI with Montana teachers. Table 6 lists the reviewed instruments and their salient characteristics. As can be seen by scanning the table, several of the instruments contained a significant number of irrelevant items and others had no psychometric data or weak results from their analyses of validity and reliability, characteristics that rendered them unsuitable for use in the present study. Based on this review, it was determined by the author of this study that it was important to develop a new instrument that would meet the specific needs of the research to be conducted. Notably, two instruments in the review, the Culturally Responsive Teacher Self Efficacy Scale (CRTSES) and the Teacher Multicultural Attitude Survey (TMAS) stood out as potentially valuable for use in this study due to their evidence of high reliability (Cronbach's Alpha = .96 and .86 respectively) and the relevance of their items. These instruments were in fact used to evaluate the validity of the instrument under development in the current study. Details on their use and analyses of the data collected can be found later in Chapter Five.

Table 6

Review of existing instruments designed to evaluate teachers' culturally congruent instruction, attitudes, and/or beliefs

Instrument Name	Author(s)	Topics Covered	Reliability	Validity	Comments/Suitability	Dimensions
Cultural and Educational Issues Survey	Pettus & Allain, 1999	Culture, ethnicity, race, religion, gender and sexual orientation	Internal consistency CA = .92; no test/retest; item correlation < .2	Content validity by 4 judges; construct validity for gender and discipline	Items on Version B on sexuality irrelevant	None reported
Personal Beliefs about Diversity	Pohan & Aguilar, 2001	Culture, ethnicity, race, religion, gender, sexual orientation, SES, language, exceptionality	Internal consistency CA = .71 to .81; no test/retest; item correlation < .3	Content validity by panel; construct validity for gender, coursework and experience	Some items irrelevant; 15 items; 5 point scale	None reported
Professional Beliefs about Diversity	Pohan & Aguilar, 2001	Culture, ethnicity, race, religion, gender, sexual orientation, SES, language, exceptionality	Internal consistency CA = .78 to .90; no test/retest; item correlation < .3	Content validity by panel; construct validity for gender, coursework and experience	Some items irrelevant. 25 items 5 point scale	None reported

Culturally Responsive Teaching Self Efficacy Scale	Siwatu, 2007	Based on Bandura and Culturally Responsive Teaching Competencies (Siwatu, 2007).	Internal consistency CA = .96; no test/retest	See factor analysis for scales	40 items, all relevant. Measures teacher's confidence in using cc teaching practices. Scale is 0 to 100. Correlates at r = .7 with CRTOE below at p = .001	7 factors with Eigen values > 1 account for 67% of variance. Scree test showed 2 to 3 factors but uninterpretable. Authors used one factor accounting for 40% of variance. Factor loadings .39 to .79
Culturally Responsive Teaching Outcome Expectancy	Siwatu, 2007	Based on Bandura and Culturally Responsive Teaching Competencies (Siwatu, 2007, 2009).	Internal consistency CA = .95; no test/retest	See factor analysis for scales	26 items, all relevant. Measures belief in positive outcomes from CC teaching practices. Scale is 0 to 100. Correlates with CRTSE above, r = .7 at p =.001	4 factors w Eigen values > 1 account for 60% of variance. Scree test showed 2 to 3 factors but uninterpretable. Authors used one factor accounting for 45% of variance. Factor loadings .55 to .75
Cultural Diversity Awareness Inventory	Henry, 1986; 1990; revised 1995; Larke 1990; Milner et al., 2003; Brown, 2004a, 2004b; others	Based on lit. review, e.g., Banks (1997), Grant and Gomez (1996), and Sleeter (1995)	Internal consistency CA = .90; test/retest = .66	Content validity by expert panel review	Widely used; items all relevant; 28 items; 5 point scale; some reversed items	4 to 5 designated scales but no factor analysis: diversity awareness, classroom environment, family/school interaction, cross cultural communication, alternative assessment

Culturally Responsive Teaching Questionnaire	Phuntsog, 2001	Based on 5 critical elements from lit review - issues and characteristics of cc teaching	None reported	Content validity by two judges	Relevant items; no pyschometric analysis; items wordy - possible concern for ESL teachers; 4 point scale	None reported
Teacher Beliefs in Urban Schools with African American Students	Love and Kruger, 2005	Based on Ladsen-Billings 1994 study of culturally relevant beliefs and practices	Internal consistency CA < .75	Only factor analysis indicated by authors	Some items irrelevant. Results mixed: Some showed correlations between teachers' cc beliefs with student standardized test scores in math, reading and Language Arts; others showed none	6 designated scales, no FA: knowledge, students' ethnicity/culture, social relations, teaching profession, teaching practice, student needs/strengths
Teacher Multicultural Attitude Survey	Ponterotto et. al., 1998; Cichelli & Su-Je Cho 2007	Multicultural awareness and sensitivity	Internal consistency CA = .86; test/retest = .8	10 grad student panel plus 2 teacher focus groups for content validity. Construct validity through convergent correlations with 3 other instruments. Criterion validity shown for PD, but not gender or race.	Items relevant. 20 revised items; 13 positive and 7 negative. 5 point scale	After two studies and repeated factor analysis, one factor used, accounting for 32.5% of variance. Factor loadings .19 to .78

Multicultural Efficacy Scale	Guyton & Wesche, 2005	Based on lit. review. Items address multicultural experience, multicultural efficacy and multicultural attitude.	Internal consistency CA = .89 overall; no test/retest	Confirmatory factor analysis. Review of prototype by over 12 nationwide experts	Many items highly relevant; significant number irrelevant. 20 efficacy, 7 attitude, 7 experience items. 4 point scale, various response categories. One 5 item multiple choice	Factor analysis showed 3 scales: Efficacy, attitude, and experience

CA= Cronbach's alpha

CCIS first prototype. The first prototype of the Culturally Congruent Instruction Survey or CCIS, as it came to be known, had been developed during a project preceding the MSP that was also directed by this study's author and also focused on teacher professional development in science education but in Flathead Reservation schools only. The process of developing the first CCIS prototype began with an extensive literature review, conducted by the author of this paper. A large body of literature relevant to American Indian culture and culturally competent teaching with American Indian students was pored over and issues, practices, and other ideas relating to CCI were identified and recorded. Research studies and other scholarly writings by experts in American Indian culture and education were reviewed. The original list of specific ideas recorded in 2005 is found in Appendix A. The ideas were then categorized and vetted for use in generating items for the draft survey. Vetting decisions were based on (a) the apparent relative importance of each idea as portrayed by their frequency and emphasis in the literature; (b) the relevance of ideas to the prioritized tribal context; and (c) the relevance of ideas to K-8 science education. The author's personal experience from her graduate studies in American Indian education, her seven years of teaching in a tribal high school, and consultations with tribal members who worked at the tribal school also factored into the choosing of items. The comprehensiveness of the items, i.e., whether they were well distributed across the many elements thought to constitute CCI for this context, was the final criterion for choosing ideas for inclusion in the draft instrument. Ideas were then changed into individual survey items as statements of instructional practice, compiled into categories, and formatted into the original prototype version of the CCIS. A four point Likert-type scale indicating frequency of use was also applied to each item. This 27-item first prototype instrument was then used in assessing impacts of the earlier PD project on teachers' use of CCI, but no analyses of validity, reliability,

or instrument structure were conducted. The first prototype instrument can be found in Appendix B.

Garnering tribal elders' support. The MSP project that is the subject of the current study covered a broader geographic area than the earlier PD project, to include five tribal cultures, three IHEs, and teachers from three reservations across the state. Before the second phase of instrument development work began, numerous visits were made to respected elders and educators in each of the partner tribal communities. Visits to the elders were made out of respect for the tribal communities involved and in observance of their protocols for looking to the elders as leaders, guides, and gatekeepers in issues related to the tribal community and culture. The purpose of the visits was to discuss the nature of the intended work, to request approval for the work from the elders, and to invite their collaboration in the development work.

Visits with elders involved considerable time and deliberate effort since the five tribal community partners lie on opposite sides of a large state. Each has unique protocols and histories, all of which should be considered in approaching elders. On the Flathead Reservation, for example, visits had to be paid to both of the two elders committees, the Salish/Pend d'Oreille Culture Committee (SPO-CC) and the Kootenai Elders Committee (KEC), which lie at opposite ends of the reservation and meet on different days. Arrangements to be placed on the agenda must be made well in advance of the monthly meeting for each by calling the Salish/Pend d' Orielle Longhouse and the Kootenai Community Center. Although given a time slot, presenters can expect either delays or earlier than scheduled appearances on the day's schedule, depending on how well business is flowing that day, so a presenter is well served to arrive an hour before they are scheduled to present. Eating and/or helping to serve lunch is not required of presenters, but is a well accepted tribal protocol that contributes to relationship building with the elder

committees, thus a visit may extend across many hours. The SPO-CC frequently request advance written documentation that supplies details about the issue a presenter is bringing to them. The SPO elders and Kootenai elders tend to engage in different levels and types of interaction and a presenter should be well prepared for a range of outcomes from very little interaction to very pointed and sometimes intense questioning if their agenda item is perceived as controversial. Select elders are recognized as spokespeople for the elders committees in many cases and may respond at length on behalf of the committee about agenda issues raised. Conversely, they may choose not to speak at all, thereby limiting interactions.

Much of the groundwork had already been laid by the author for the three tribal communities on the western side of the state during the work she completed during the earlier PD project. A tribal community member who was working as a program coordinator with the project led this phase of the work in the two tribal communities on the eastern side of the state. On both sides of the state, the relevant project personnel deliberately identified and invited elders who were recognized as respected knowledge keepers and teachers in their respective communities to join the development team. Approval for the project was given by the elders of each of the five tribal communities involved in the study.

Revising the first prototype CCIS. The main author of the first draft of the instrument (also author of this paper) is a non-Indian woman who had worked at that point for sixteen years as an educator in the Flathead Reservation tribal secondary school and tribal college.. During that time she had formed close relationships with tribal members with whom she had worked extensively in these educational settings. When the MSP project's leadership team decided that the CCIS should be used as part of the project evaluation, she collaborated with four of these local tribal consultants to revise the items that comprised the first prototype instrument to

improve the items' clarity, accuracy and ease of use. This work began in winter 2007 and extended into spring 2008.

Meetings for revising the instrument were informal and often were conducted one-on-one but occasionally occurred in groups of three, and were commonly conducted after sharing a meal, in line with local tribal traditions. Two of the tribal collaborators were women who were members of the Salish tribe. One was a 65 year-old elder who had worked in many different jobs with the tribes, including as a paraprofessional in the tribal secondary school and as a tribal cultural specialist. The other was a 50 year-old tribal educator who had held many prestigious positions in tribal education and was well known in the state for her expertise and advocacy in Indian education. The other two collaborators were members of the Kootenai band, a man and a woman. The Kootenai woman was a 63 year-old elder who had also held many different positions as a tribal employee and was currently acting as a designated cultural representative for the Kootenai people, often in formal and informal educational settings. The Kootenai man was a Kootenai language specialist who worked for the Kootenai Elders Committee and held advanced degrees in Native studies and education.

Each person contributed suggestions for improving the instrument's content and layout, which were then incorporated into the second prototype instrument. One person suggested improvements in the wording of items to make them more applicable to other tribes that were involved in the project, so an item previously worded as "Salish or Kootenai words or phrases posted" became "Posted words or phrases in local Native languages." Another person suggested deletion of the item "Student use of instructional technology," noting that she did not consider it a culturally congruent practice. In all, three items were deleted (items #14, #19, and #24), nine were revised, fifteen remained intact, and eleven were added to the second prototype of the

instrument. The items that were added to the second prototype were items #25 through #28 and items #30 through #35, along with an open-ended item, #13. Five of the added items originated with suggestions made by tribal experts including #28 - Art based instructional methods, #32 – Teaching Indigenous science along with Western science, and #31 – Place-based science. Six of the new items were found in the literature review as recommended instructional methods for American Indian students and were approved by the tribal experts. These included item # 34 – Observational learning strategies, item # 25 – Mentoring by adults other then the classroom teacher, and #26 – Opportunities for private practice precede public demonstration of proficiency.

The resultant survey was a 35-item instrument (34 forced-choice and one open-ended item) that employed the same four point Likert-type scale indicating frequency of use for each item that was used on the first prototype. The number of items included on the second prototype was increased from the 27 that were on the first prototype. This second prototype of the CCIS was piloted to collect baseline (pre-treatment) data with the MSP project's first cohort of treatment group teachers. The data were also analyzed to gather evidence of the technical quality of the instrument. These analyses and their results are described in Chapter Five. The second prototype instrument can be viewed in Appendix C.

Revising the second prototype CCIS. The next step taken in the instrument development process occurred nearly a year later in February 2009 at a two-day meeting with representatives from every party of stakeholders involved in the PD project participating – elders from the five tribal cultures involved in the project, project leadership from each of the IHEs, project professional developers and classroom mentors (former K-8 teachers), practicing K-8 teachers from participating schools, an external evaluator, and science and science education

faculty along with graduate students in science education from the three partner IHEs. The sole evaluator involved was a member of the Turtle Mountain Chippewa Tribe. Otherwise, each partner group participating included both Indian and non-Indian representatives, creating a well balanced ethnic mix of professional educators and non educators. Many of the participants knew each other, having worked together previously. These measures were taken deliberately to ensure a balanced and friendly group and to help create a safe environment for all that would enable candid conversation.

Meeting environment. The meeting lasted two days and was held in a hotel conference room located about midway between the eastern and western reservations involved in the study. This was considered "neutral territory," as opposed to holding the meeting at a university or school, in which it was anticipated that some of the tribal participants may not have been comfortable due to the history of negative experiences with schools that many had encountered in their lives.

The meeting was scheduled over two days not only because of the distances people had to travel, but also in an attempt to provide time for the group to become comfortable with the meeting venue and to bond as a group. Most of the group stayed overnight at the meeting hotel and everyone ate meals together, a culturally competent practice that was also deliberately observed. Participants were seated around tables arranged in a U shape, with the evaluator in the middle of the U, a typical arrangement in tribal settings that promotes a more equitable and open atmosphere for conversation, as opposed to sitting in rows which limits interaction with others.

These types of details for the meeting's format were deliberately observed as a means to foster a safe environment and enable relationship building, a cultural value that permeates the lives of the Native people involved in this study. These measures also were taken to provide

extensive opportunities for participants to reflect deliberately on CCI and related topics, thereby supporting them in sharing their experiences and deep thoughts on these sometimes controversial and emotion-laden topics.

Meeting enactment. The external evaluator who facilitated the meeting conversation was carefully chosen for her extensive evaluation experience in Indian Country and her known expertise in facilitating emergent conversations using participatory evaluation processes. The meeting began with a prayer, as is traditional for the tribal people who participated. The evaluator then initiated an open-ended conversation with all participants about the meaning and significance of culture and CCI. Specifically, the questions posed were *How do we define culture in terms of cultural competency for teaching, especially in science? What are the different elements of culturally competent teaching?* After an extended discussion about the first question held over several hours in which people freely expressed their views and told stories about their family members' and their own experiences with school, the facilitator slowly and deliberately moved the group toward discussing the nature of CCI for the specific tribal communities involved, and an examination of how CCI would look in K-8 classrooms, particularly for science instruction.

Open-ended discussion. During the conversations there was no interview protocol employed or adherence to formal rules for contributing to the conversation, although the evaluator did consult with project representatives in advance of the meeting to discuss the meeting's objectives and again during breaks in the meeting as a touch point to determine what other types of information were desirable. The format of the meeting was similar to that of a *Talking Circle* in which any participant was welcomed but not obligated to speak. Protocols for Talking Circles can vary with specific tribal cultures, but generally they are semi structured,

naturally flowing, informal conversations focused on a central topic but often addressing many related topics. Participants are free to contribute when they feel that they have something important to say, and are allowed to speak as long as they desire without interruption. Other members of the circle listen respectfully and may respond to any speaker after their conclusion, or they may choose to move on to related topics. This type of format is a thorough and egalitarian one commonly used in tribal settings, and can be very time consuming compared to more structured meetings that follow an agenda and are pushed along so that all items are discussed. The extra time sometimes associated with a Talking Circle is time well spent in attaining objectives like those of this phase of the instrument development; the open-ended nature of the Talking Circle enables ideas to emerge naturally from the conversation, often with more depth than could be achieved in a more structured format. For the development of the CCIS, this format was particularly effective given the diversity of the stakeholders and the objectives of defining and operationalizing the construct of CCI for the specific cultures and contexts involved. Several of the university faculty involved in the project had limited familiarity with the construct and with the tribal cultures participating in the project. Two of them related to the author after the meeting that its open format was very valuable in deepening their personal understanding of cultural issues and in providing an opportunity for them to build relationships with tribal partners in the project.

Equitable opportunities for stakeholder input. It was anticipated that the most significant outcome of this meeting of stakeholders would be the important input about CCI and science education provided by the members of the tribal cultures with whom the instrument would be used. Although it is estimated that about 95% of the contributions to the discussion were made by the tribal partners, the meeting also provided opportunities for input from the other project

partners, many of whom were non-Indian K-20 science educators and science faculty. It was anticipated that the feedback received from the science educators, both Indian and non-Indian, would be especially useful in ensuring that the CCIS was designed to comprehensively address relevant aspects of K-8 science instruction and that the survey would be teacher friendly, for example, by utilizing jargon familiar to professional educators.

Reacting to the CCIS. The meeting's discussion was recorded in two ways: the evaluator wrote brief notes about emergent big ideas on a large chart for all participants to see while the project director recorded the conversation in greater detail on a laptop computer. The two sets of notes from the conversation of the first day of the meeting were analyzed by the evaluator at the end of the day to identify overarching themes that had emerged on Day One that could be probed further to elicit more in-depth information during Day Two's discussions. On the second day of the meeting the evaluator approached the group with four teaching scenarios that emerged from the previous day's conversation as characteristic of the professional development project and whose examination she felt would provide additional valuable information. Those four scenarios were:

1. American Indian teachers teaching mostly American Indian students of their same culture in reservation schools
2. American Indian teachers from a different tribal culture teaching mostly American Indian students in reservation schools
3. Non-Indian teachers teaching mostly American Indian students in reservation schools
4. Non-Indian teachers teaching mostly non-Indian students in off-reservation schools

These four scenarios provided fodder for the second day's conversation as the whole group engaged again in a Talking Circle-type format and discussed the differences in CCI for

each scenario. Midway through the second day of meetings the evaluator shared the second prototype CCIS with the meeting participants and solicited their feedback on the instrument, again through open discussion. The resultant ideas from the two-day discussion were used by the author of this study to modify the instrument items into a third generation prototype referred to as the Revised CCIS.

Follow up focus groups. In the next few months following the two-day meeting described above, the evaluator conducted smaller focus groups with subsets of treatment teachers from the two project sites located on the western side of Montana. The teacher participants in the focus groups were selected by the project leadership at each site based on their advanced level of cultural competence compared to their peers, as judged by the project leadership. One meeting was held on a reservation with teachers who all taught in reservation schools with high percentages of American Indian enrollment. The other meeting was held off reservation with teachers who all taught in off-reservation schools with low percentages of American Indian student enrollment.

The focus group facilitator, the same person who conducted the two-day meeting, utilized a semi-structured interview protocol in the focus groups whose questions were co-designed by the author of this paper and the evaluator to address important issues that emerged from the original two-day meeting. The first question posed to start the conversation was "What are the attributes of a culturally competent teacher?" with a request for specific examples used as a follow up prompt. The next question was "What does a culturally congruent classroom look like?", again with a follow prompt requesting specific examples. The conversation was then moved to a discussion of the challenges of culturally competent teaching and ways to effectively address them. These first three questions were very similar to those asked in the two-day

meeting with tribal experts. A large flip chart was used to record and display teachers' responses to the facilitator's questions. A list of CCI elements was brainstormed by the teacher participants during each of the focus groups.

The facilitator then turned the conversation to an examination of how the PD project had changed individual teacher's understanding and use of culturally congruent practices in their classrooms. About midway through the session, the facilitator shared with the teachers the list of CCI attributes brainstormed by the tribal experts at the two-day meeting and asked them to compare it to the list that they had just generated in response to similar questions. The next question asked teachers to rate the influence of each of the project activity types on developing their cultural competence. The focus group concluded with a discussion about the essential elements that should be included in a PD model for developing culturally congruent teachers. The ideas generated in the focus groups mirrored those that surfaced in the two-day meeting with tribal experts, except in the omission of an emphasis on poverty that was raised by the tribal experts. The full set of focus group questions is available in Appendix D.

Expert review for final item revisions. The next step in the development of the CCIS was to engage in member checking of the instrument. The author of this study met face-to-face with several members of the original two-day meeting group to go over the instrument items one by one. Reviewers were asked to check the CCIS items for face validity, inclusiveness, bias, clarity of the language, user friendliness, and accuracy in portraying the ideas that emerged from the instrument development meetings. Additional feedback was gathered via electronic mail communications with project personnel, both those who had participated in the three meetings (the two-day meeting plus the two focus groups) and those who had not. Feedback was also collected via electronic mail from two local tribal professionals external to the project who are

recognized experts in American Indian education. The feedback from the face-to-face meetings and electronic mail communications consisted of positive comments and two suggestions for revisions: the rewording of items to specify Montana Indian tribes, rather than local American Indian tribes, and the removal of redundant items.

One tribal expert, a female Salish elder, also made multiple suggestions for additional items. She had worked for years teaching in both formal and informal settings, including in the tribal alternative secondary school with the author, and stated that she drew her ideas from those experiences. New items that were generated based on her suggestions are:

- Two items addressing the availability of cultural games and toys, web sites and software (#30 & #31)
- An item about time flexibility (#22)
- Two items about the cultural compatibility of the classroom environment and management (#37)
- An item about the use of symbols and analogies (#15)
- An item about student self regulation of learning (#10)
- An item about student's application of knowledge to solve problems relevant to their communities (#18)

These new items were approved by other tribal experts on the development team. The final revised survey resulted in 37 forced-choice items plus four open-ended items divided into four categories. This differed from the second prototype instrument which was comprised of 34 forced-choice items plus one open-ended item divided into three categories. Forced-choice items were written to be unidimensional. Open-ended items consisted of a text box that allowed respondents to add comments or items that were not included on the instrument, increased from

one to four. The final revised survey items were also more clearly delineated into their categories with prominent category headings, and a short paragraph included beneath each heading with explicit directions about how to respond to the category items. An example illustrating how to use the scale for responding to an item was also provided at the beginning of the instrument.

Revising the response scale. In conversations with the statistician who conducted the analyses of the data from the second prototype instrument it was determined that, while significant findings were found, the granularity of the data could be improved to provide finer detail about teachers' frequency of use of CCI. It was decided that on the revised instrument a finer grained scale should be used as a means to try and improve the quality of the data. This study's author combed the literature on survey scale design and consulted with three assessment design experts external to the project, using the information gathered to design two new six point scales that were used on the Revised CCIS, replacing the four point scale found on earlier versions of the instrument. The Revised CCIS can be found in Appendix E of this paper.

The domain of content for CCI. A synthesis of the tribal expert input regarding CCI and what it looks like in practice gathered through the various strategies employed in the development process is displayed in Figure 2. This figure lays out the domain of content for the construct of CCI for the five tribal cultures involved in this dissertation study. Some of the items originated in the literature review and were presented to tribal experts during meetings for their scrutiny. Others originated from the tribal experts' conversations. Regardless of their origin, all were affirmed as culturally congruent by the tribal experts on the development team and so are included in the domain. This domain, displayed graphically in Figure 2, presents another layer of specificity in the examination of CCI, a finer grained representation of the aspects of CCI for the five tribal cultures who participated in this dissertation study.

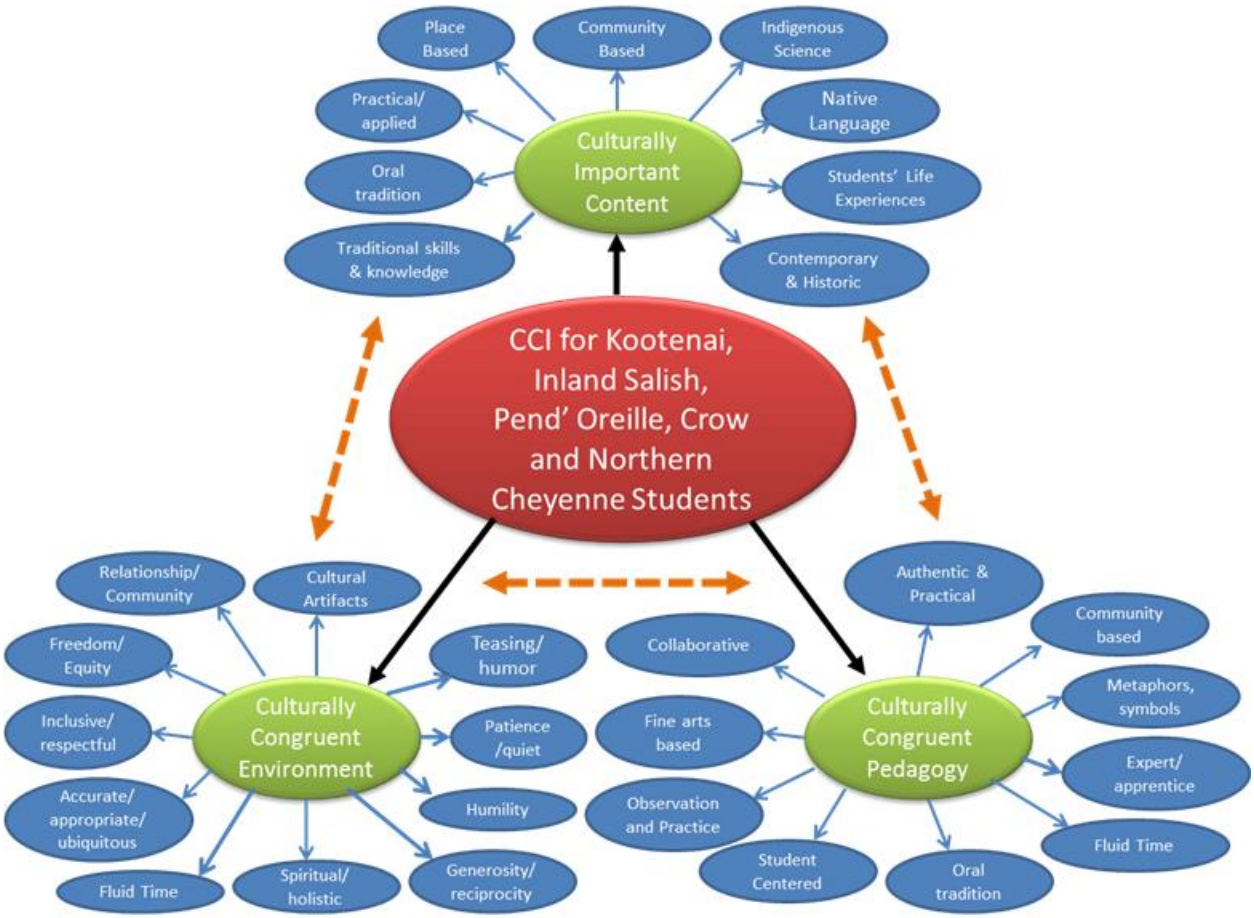

Figure 2. Model of CCI for Kootenai, Inland Salish, Pend' Oreille, Crow and Northern Cheyenne students. A list of traits was compiled through literature review and discussions with the development team. Traits in the graphic are those agreed to by members of the development team as essential to CCI for students from these tribes.

Grounding the development process. The processes used in the development and validation of the CCIS were in large part based in the cultural norms and protocols of American Indian people, particularly for the five tribal cultures for which the instrument was designed. The work was grounded in the research literature and in the knowledge and experiences of the project partners who collaboratively developed the instrument. An extensive literature review of American Indian cultural norms and education had been conducted by the author of this study during the development of the prototype instrument for an earlier PD project. The literature review was expanded during her master's and doctoral degree studies. Knowledge gleaned in the

literature review was used to generate a list of common characteristics and themes related to American Indian cultures and education that was used in generating items for the earliest version of the survey.

The next steps in developing the instrument were grounded in the real life experiences and knowledge of American Indian adults and education professionals. This stage of the work employed participatory, inclusive, and culturally congruent methods to engage development team members in conversations about culture and education that related to the instrument development. The experiences of the project leadership team significantly influenced the development process. Some of the project leaders had lived and worked in their respective communities for many years and held deep understanding of aspects of the community that were relevant to the development work such as world views, epistemologies, histories, norms, and values. Their knowledge was valuable in arranging development meetings that observed cultural norms, in identifying and approaching knowledgeable community members who could serve on the development team, and in contributing to conversations in development meetings. The instrument development team included tribal member partners, K-8 teachers, and IHE faculty, therefore the cultural considerations addressed in the instrument development process were many and varied.

The often overlapping themes that guided the design of the development processes and their manifestations in the processes are discussed in the following paragraphs. Examining the manner in which these themes were addressed in the study assists in characterizing the development process in response to Research Question #1.

Community benefit. Placing the good of the community, rather than that of the individual, at the center of life is common in American Indian cultures (Cajete, 2001; LaFrance

& Nichols, 2010). Work is valued for its practical application in improving the lives of community members (Aikenhead, 2011). In acknowledgement of this value, the instrument development work consciously and consistently focused on issues and outcomes that partners agreed were important because they held potential benefit for their community in improving science education and student achievement. This was actualized from the start of the work with the project leadership's visits with elders and other tribal members to solicit their input and approval of the project, and continued throughout each stage of the work. Individuals who were viewed or, in some cases, formally designated by their community as community spokespeople were deliberately identified and invited to participate in the development work. Individuals tapped for the work consistently stated that they were committing to join the team in recognition of the potential value of the instrument for improving teacher practice and student achievement. No one turned down the invitation to participate in the process. Conversations in development meetings reflected the prioritization of community, focusing on the consideration of community history, values, norms, needs, perspectives, and particularly issues surrounding American Indian education, with individuals commonly sharing personal stories that illustrated these ideas. Tribal experts placed emphasis in their conversations on the good of the community; stories told illustrated events that, though personal, presented a message about community. An example of this occurred during a small group meeting between the author and two tribal experts from the Flathead Reservation. One tribal expert relayed stories of her own experiences in school, those of her relatives, and those of other people she knew in the community. She ended the story by underscoring the need to improve education to support more American Indian students to graduate from college so they could return to the reservations and help their tribes.

Trusting relationships. Relationships are of paramount importance in tribal communities

and they significantly influence the outcomes of collaborative work, for example, by affecting the candor of conversations and people's willingness to share personal or sensitive information (Aikenhead, 2011; Gilliland, 1999). Every effort was made to foster relationships between the development team partners, in large part by carefully observing cultural norms. Close attention was paid to organizing meetings that fostered a relaxed, egalitarian, and culturally congruent environment, for example, by arranging seating in circular shapes to enable equitable communication, opening and closing meetings with a prayer, encouraging flexibility in the meeting schedules, and allowing generous time to thoroughly discuss ideas, sharing meals and lodging, locating the meetings at central locations so no partner would have an undue travel burden, sharing personal stories and oral histories, and encouraging and honoring every person's contributions to the work. Given the hegemonic hold that non-Indians have historically imposed on American Indians and the legacy of the use of education as a tool of assimilation, the project team was particularly sensitive to the need to work at developing relationships of trust. Because about half of the leadership group had very little experience working with tribal communities, the idea of relationship building was prioritized as a regular topic of conversation in project leadership meetings and the means to develop relationships were overtly discussed and employed early and often in the project partnership. Also, in an effort to make things more comfortable for the tribal members involved, the evaluator who facilitated the multiple-day development meeting was herself American Indian, with extensive experience working in Indian Country and knowledge of the cultural norms of many tribal cultures.

Collaboration and egalitarianism. American Indian communities are generally collaborative and egalitarian in nature, rather than hierarchical and competitive (Aikenhead, 2011; Gilliland, 1999). These norms were reflected in many ways throughout the instrument

development processes. Work was conducted collaboratively in nearly every stage. Conscious efforts were made to discourage any perception of hierarchy among the K-8 teachers, tribal consultants, and college faculty who participated in the instrument development. Decisions regarding the design of the development processes as well as the design of the instrument itself were made through collaborative discussion that included the voices of all stakeholders. Decisions were rendered through consensus, with every person's vote given equal weight. The inclusion of specific survey items in the instrument from the second prototype onward, for example, only occurred if all members considered them accurate and valuable. The wording of items was crafted meticulously, soliciting input from every member of the development team either in full group, small group, or one-on-one meetings with the study's author. During group discussions of American Indian culture and education, non-Indian members of the group spent the bulk of the time listening, sharing ideas sparingly and judiciously, thus enabling tribal partners to share their expertise more fully. These norms of collaboration and egalitarianism are reciprocally reflected in other tribal norms that were observed in the study's design, such as the value of relationships and the importance of community.

Fluidity of time. American Indians commonly have a fluid conception of time and tend to be less rigid about schedules, instead emphasizing process over deadlines (LaFrance & Nichols, 2010). Tribal elders have on several occasions told the author of this study, for example, that things happen not according to a schedule arbitrarily set by humans but when they are meant to occur. Many elements of mainstream American society hold the opposite perspective, for example, emphasizing the importance of punctuality and following schedules to the letter, multitasking to make efficient use of time, and cutting corners in order to make deadlines. Realizing the Native perspective on time, development work for this project observed fluid time

frames, engaged in thorough and deliberate conversations, and allowed ample time for relationship building. This did not lead to superfluous conversations, wasted effort, empty time, or unproductive meetings. Instead, the project partners enjoyed the development of trusting relationships, invested partners committed to the work, and in-depth conversations regarding personal experiences, oral histories, and cultural values that informed the instrument design in ways that would not have been likely otherwise.

Generosity and hospitality. American Indian people generally are a low-income population, often living in rural, even remote areas, commonly with limited infrastructure and few well-paying jobs. Even so, hospitality and generosity are traditional values in many American Indian cultures (Gilliland, 1999). Guests are welcomed and made comfortable. Sharing food and resources and providing lodging are common practices. Giving gifts is a traditional means of honoring people's contributions to the community. The values of generosity and hospitality overlap with the importance of developing trusting relationships and valuing community. The current study observed these values in consistently providing food, lodging, transportation, and safe, well-equipped, comfortable work environments for the development team partners. Partners lived on opposite sides of the state, about eight hours' drive apart. In many cases for the tribal partners, the closest major medical facility or shopping mall was several hours' drive from their homes. Multiple day meetings in urban centers made the long distance travel required more valuable and palatable, enabling access to comfortable hotels with in-house meeting facilities, increasing the time the group could spend together, providing opportunities for sharing several meals per day, and enabling access to shopping and other resources not available on many reservations.

Communication norms. The norms of communication utilized in the project were

carefully chosen for their cultural compatibility. In today's world, many people, professionally and socially, communicate via electronic media, such as electronic mail, blogs, wikis, and text messages. In the small towns and tribal communities involved in this study these forms of communication are present also, but in many of the study's scenarios, the leadership relied instead on oral conversations for project related work, realizing that they were the preferred mode of communication in those contexts. An example is seen in the means used to invite people to join the development team. The project director and a site coordinator, both of whom had deep knowledge and experience in the tribal communities involved in the study, took the lead in identifying and inviting respected community members to work as part of the team developing and validating the instrument. In alignment with each community's cultural norms, most invitations to participate were extended face-to-face and in some cases via phone call, rather than by means of electronic mail or text. This enabled the project personnel to visit on a personal level with invitees about their families, mutual friends, events in the community, and so forth, thereby fostering and honoring their relationships with the potential participants. They were also able to discuss the project thoroughly with each invitee, answering any questions they had, and soliciting ideas for meeting times and locations and other potential points of contact that should be invited to participate in the process. Meetings with tribal elders and meetings involving discussion of potentially personal or emotion-laden topics were always conducted face-to-face. In other aspects of the project, such as meetings between multiple faculty members to discuss business issues of the project, electronic mail and teleconferencing were the communication modes more frequently used because partners were located hundreds of miles apart and accustomed to communicating electronically.

Open-ended discussion. It is common among American Indian people, particularly

traditionally raised American Indians, to listen and observe quietly, intently and respectfully, and to use words judiciously. In many American Indian cultures, words are considered to have primordial power and so should be used carefully and ideas considered thoroughly (Northwest Indian Applied Research Institute, no date). Patient, deep deliberation of ideas is also a common occurrence. The pace of conversation with elders, for example, may be measured and immediate feedback limited. A holistic perspective, considering ideas in the larger context, is common when talking with American Indian people about serious or complex topics.

The open-ended nature of the discussions used throughout the development processes reflected observation of the communication norms of the five tribal cultures involved in the study. In encouraging conversation at the multiple-day development meeting, for example, specific ideas relevant to CCI were not suggested to partners as topics of conversation until the latter half of Day Two when the second prototype instrument was introduced for review by partners. On Day One, partners generated ideas organically through deep open-ended discussions that began with the posing of broader questions like *How do we define culture?*, *What have your experiences with education been like?*, and *What do you think culturally congruent instruction looks like?* Discussions were only occasionally lightly facilitated, allowing partners to freely discuss ideas they considered relevant and important and to contribute when they felt it was valuable. Members were encouraged to voluntarily share their ideas, with no parameters placed on the topics discussed or the duration of the conversation. This is reminiscent of the cultural norms of the Talking Circle, a form of discussion that is common in tribal communities, and whose design honors each person's contributions.

Storytelling and oral history. Oral history and storytelling are traditional forms of communication observed by many American Indian cultures (LaFrance et al., 2012). Stories

serve as teaching tools and provide entertainment, and are a regular part of everyday life in many American Indian communities. It is through storytelling that the history and values of Indian people are handed down from generation to generation. Storytelling was a frequent occurrence in project meetings with the development team. It was used to convey people's experiences in schools, both good and bad, for example. Stories were shared that illustrated the cultural overlap and incongruencies that exist between non-Indian and American Indian people. Stories were also used in describing traditional teaching methods and learning objectives. Storytelling was one of the most common and important means of communication used in the development processes. The fluid use of time, the development of trusting relationships, the open-ended nature of discussions, and the provision of safe meeting environments enabled and encouraged storytelling.

Cultural specificity and commonality. There are over 500 federally recognized tribes in the United States, and many other unrecognized tribes, each with their own unique culture. Even so, there is a tendency in American society to treat American Indians as one cultural group. Project partners, both non-Indian and Indian, were keenly aware of society's tendency to lean toward "pan-Indianism" and every effort was made to avoid this faux pas. Respected representatives from each culture participated in the development work and each shared tribally specific ideas for their respective community. The contributions to the conversation made by each were acknowledged and honored in the design of the draft survey items. At the same time, it was also conceded in the group conversations that American Indian cultures share some commonalities. In the end, survey items were chosen for inclusion in the final revised instrument only if partners agreed that they were applicable to all five tribal cultures involved in the study, in recognition of the two sides of this coin.

This section of the dissertation study described the cultural accommodations that were observed in the development process that contributed to its cultural congruence. The next section discusses evidence in the form of behaviors and feedback that provide testimony about the quality of the development process.

Evidence of the Quality of the Study's Processes. Evidence of the quality of the development and validation processes came in many forms, most of which are reflections of people's reactions and behaviors regarding the processes and the instrument. These types of evidence are discussed in the following paragraphs.

Development of trusting relationships. One form of evidence of the quality of the development and validation processes was seen in the trusting relationships that grew as the project's collaborative work progressed. Although some of them had just met and the cultural divisions were wide in many cases, participants in the multiple-day meetings exhibited a growing sense of trust in their fellow group members in their willingness to talk increasingly candidly, particularly about controversial subjects regarding American Indian education, as the meeting progressed. By the second day of the meeting, members of the group were freely mingling before and after the meeting, as well as during breaks in the meeting. Also on Day Two, American Indian members described painful experiences from their boarding school days, talking openly with a group in which nearly half the members were non-Indian people. This is significant not only because of the cultural differences that existed between group members but because historically it was the hegemonic relationship with non-Indians that led to the oppressive and emotion wrought events that they were describing. Further, American Indian people generally tend to be reserved in unfamiliar public situations and use their words judiciously; the willingness of participants to freely speak about sensitive subjects suggests that they held some

level of trust for the people present at the meeting. Without the safe atmosphere and the relationships that the process fostered through the observation of cultural traditions and norms, such openness would most likely not have occurred and the personal experiences that informed the instrument design would not have been shared publicly.

Participation and feedback. A second form of evidence of the quality of the development process is provided in the level of investment made in the project by participants. Nearly every person who was approached to work with the development team accepted the invitation to participate and followed through by attending several meetings and participating fully. No one balked when asked to review the instrument a second and sometimes a third time in follow up meetings. Their willingness to join the project and fully engage in the work is testament to participants' receptiveness to the process and to its appropriateness for the context and purpose for which it was used.

Another form of qualitative evidence of the nature of the process came in the form of participants' comments about the methods used to develop it. Participants involved in the development meetings were highly complimentary in their reactions to the process, providing unsolicited positive comments about the value of the methods used. One woman, a traditionally raised elder, noted to the author privately at the end of the two day meeting that the time spent with the group members helped her to feel comfortable in sharing her experiences and opinions about education and American Indian people. She also thanked the group for "taking on such an important issue". A second female elder reiterated appreciation to the group and added that she was happy to be part of such important work. Another woman, a traditionally raised elder and elementary teacher, thanked the group at the end of the two-day development meeting for the time and effort they were spending on this important work to improve American Indian

education. She noted that the fact that so many tribal members were part of the meeting and the conversation had been so deep indicated to her that the group was "on the right path" and serious about the work. The site coordinator who had invited many of the tribal participants to the meeting and who was also American Indian, noted privately to the author that rarely had she seen a meeting of this type in which the tribal people outnumbered the White people. She attributed the willingness of the tribal people present to contribute so openly in large part to the ethnic makeup and format of the meeting.

The CCIS was also reviewed in the later stages of development individually by five people who were not involved in the group meetings. The instrument was sent to them by electronic mail and they reviewed it on their own, checking its face validity, accuracy, format, user friendliness, technical quality, and cultural congruence, depending on their expertise. Three members of this group of reviewers were American Indian, one Salish, one Kootenai, and one Paiute; two of them held degrees in education and had worked in schools as teachers, the third was a statistician and emeritus professor at one of the partner IHEs. Each responded favorably to the instrument. One reviewer stated "I like the content and what the questions are getting at. Certainly illustrates concepts that teachers should be thinking about when delivering lessons to students…I really like the work you've done!" Another stated "Your survey looks great. There are so many strategies listed. I would love to hear how they turn out." One recommended additional people in the tribal community that could be contacted to review the instrument, indicating his belief that the endeavor was worthy enough to involve elders and other community experts in the process.

The other two reviewers of the five external to the project development team were non-Indians who held advanced knowledge in assessment design and statistical analysis. Both were

involved in extensive conversations with the study's author via electronic mail and teleconferences about the technical quality of the survey. Three members of the group of five reviewers provided ideas for minor, cosmetic changes to the instrument verbiage or format, which were incorporated by the author. As with those involved in the group work, none of the reviewers responded negatively to the instrument or refused to review it. In each case, feedback was provided within a week after the instrument was electronically mailed to the five reviewers. These positive reactions and prompt responses could be considered testaments to the quality of the development process. They underscore the importance reviewers felt toward the development of the instrument and its potential to benefit the tribal, education, and research communities.

Instrument dissemination and use. The widening use of the instrument in the research and education communities indirectly speaks to the quality of the development processes used to generate the Revised CCIS. In at least two cases, partners in the instrument development process chose to use the Revised CCIS in work outside the project or shared it with others, recommending its use in their communities, particularly in educational settings. A highly respected Crow elder and teacher participant reported that she had given the Revised CCIS to her superintendent and recommended its use in the school district to assist teachers in understanding and self assessing their CCI, for example. A graduate student who was involved in the study's work with one of the partner IHEs used the instrument in her as yet unpublished master's degree research.

Numerous Native and non Native researchers external to the project have requested information about the development work and in some cases are actively using the Revised CCIS in their research. Two research projects operated through IHEs in neighboring states are currently using the Revised CCIS in investigating interactions between culturally congruent

curriculum, teacher professional development, and American Indian student achievement, for example. Many of the early requests for information came as a result of project participants spreading the information by word of mouth in their communities.

Presentations delivered at professional conferences about the CCIS and its development and validation processes likewise generated much interest. Presentations consistently sparked enthusiastic conversations among the audience members and presenters and numerous requests for the conference papers and instrument were received from researchers from several continents. One of the more unique presentation settings occurred at a national conference for the American Indian Science and Engineering Society, where American Indian K-20 education professionals comprised the majority of the audience. After the presentation detailing the development processes and instrument was delivered, the audience members were asked to discuss with others sitting in proximity how they would customize the development processes to make them culturally appropriate if they were to try to develop a similar instrument for use in their own settings. This led to very engaged and protracted small group discussions as participants described the culturally specific ways that they would adapt the methods. In the follow up sharing of ideas with the whole group, the audience was again engaged in deep conversations and the session ran beyond its allotted time. A sign up sheet was circulated for those who were interested in receiving the presentation, paper, and instrument to provide contact information; forty-eight people attending the session requested the information.

This chapter focused on Research Question #1 by delineating the details of the development process, describing the cultural accommodations that contributed to its cultural congruence, and providing evidence of the quality of the process. The development process utilized standard practices in instrument development, such as literature review, expert feedback,

and iterative review. It set itself apart from more typical development processes in the culturally congruent accommodations that were infused throughout the process.

The next chapter addresses Research Question #2regarding the technical quality of the instrument through the examination of both the qualitative processes that contributed to content validity, and also the results of the statistical analyses of data produced through the use of the instrument.

CHAPTER FIVE – RESULTS FOR RESEARCH QUESTION #2

In investigating Research Question #2, What is the technical quality of the instrument?, this dissertation study examined the results of the analyses of several types of data and collected a body of evidence of validity for the inferences made from such analyses. This chapter discusses those results and their significance to characterizing the instrument's technical quality.

The value of an instrument like the Revised CCIS lies in its ability to provide information that is accurate, credible and trustworthy for the purpose for which it was designed, i.e., its validity. Validity is not an intrinsic characteristic of an instrument and it is not something that is measured directly. Rather it is a judgment of the appropriateness of the inferences made from the instrument as supported by theoretical rationales and the data generated through its use. Carmines and Zeller (1979) note "One validates not the instrument itself but the instrument in relation to the purpose for which it is being used" (p. 17). In this study, the validity of the Revised CCIS for assessing teachers' CCI in science classrooms serving Montana American Indians was investigated.

Evidence of Reliability and Internal Structure

Reliability is the "consistency of results produced by a measure" (Krathwohl, 1998, p. 436) or the tendency for an instrument to produce similar results on repeated administrations under similar conditions. Ascertaining the reliability of the data produced in using an instrument is an important part of characterizing the instrument and contributes to establishing evidence of validity. Finding evidence of reliability on its own does not allow one to conclude that use of the instrument enables valid inferences. In fact, an instrument may prove to have high reliability even though it does not result in valid inferences about the construct it was intended to measure. Many factors can affect reliability.

Reliability can be estimated in several ways. In the present study, evidence of reliability was sought through the calculation of test-retest scores correlations as indicators of temporal stability, and through the calculation of Cronbach's alphas and inter-item and item total correlations as indicators of internal consistency.

Pearson's Correlation Coefficient for the test-retest. For the test-retest reliability estimate, Pearson's Correlation Coefficient was calculated using the scores from the 68 treatment and comparison group teachers who completed the two administrations of the CCIS spaced six weeks apart in spring 2009. The strength of association between the two sets of scores was found to be .74; it was statistically significant at the $p = .001$ level for a two-tailed t test. This moderately high association supplied good evidence of the temporal stability of the data provided by the instrument under the study's conditions. The correlation coefficient also indicates that 55% of the variance in the scores is accounted for by the construct being measured while 45% of the variance is due to factors unrelated to the construct.

Reactivity. The test-retest method is subject to some potential influences that should be

taken into consideration when interpreting it. Carmines and Zeller (1979) describe the effect of reactivity, for example, in which subjects may change their stance on an issue or behavior toward a phenomenon as a result of measuring the phenomenon. If, for example, a teacher in this study became more aware of specific strategies relating to CCI as a result of completing the first administration of the survey and began using those strategies in their teaching, she might change her responses on the second survey due to this reactive change in their behavior. Reactivity can lead to deflated estimates of reliability. Since the second administration of the survey occurred six weeks after the first one during the summer about three weeks after the school year ended, this effect most likely did not significantly influence teachers' responses on the second survey.

Time between administrations. The time between administrations of the test and retest can also influence subjects' responses. Too short an interval between administrations can mean that memories from the first administration may influence subjects' responses on the second administration, sometimes resulting in subjects' attempts to duplicate their answers, and causing an overestimation of the Pearson's correlation coefficient. Too long of an interval between instrument administrations may result in a change in subjects' perception of the construct. In the present study, the six-week interval was judged as an appropriate interval to minimize the potential effects of time between instrument administrations. The first administration occurred between four and six weeks before the end of the school year, depending on each school district's calendar. Teachers were finishing up the school year for most of the interval between administrations, a busy time of year that involves tying up loose ends, fieldtrips, special events, grading, and the like, thereby limiting other influences or distractions that might have affected teachers' perceptions of the construct. The second administration occurred before the start of the professional development intervention, to prevent the training from influencing responses.

Consistent administration. Another potential influence on the data produced from using an instrument has to do with the setting in which it is administered. If the settings differ for the pre- and posttest, for example, if there are distractions present in one setting such as people talking that are not present in the other setting, this can influence the data that are produced in each administration. To safeguard against influences due to inconsistencies in administration, the proctors for the instrument administration were given explicit instructions regarding how, when, and where to administer the instrument and were asked to hold these conditions steady for the next administration six weeks later. The accommodations and controls employed in this study served to limit each of the influences on the test-retest method that are described in these paragraphs, enabling increased confidence in the reliability of the data produced by the instrument.

Factor Structure and Internal Consistency. Instruments may address multiple aspects of a single construct, and so may be comprised of multiples factors or scales, each representing a different aspect. Factor analysis can provide evidence of validity and reliability in characterizing the underlying factor structure of an instrument, play a data reduction role in decreasing the number of dimensions or scales from the total number of items to a smaller set of representative dimensions, and contribute to theory building about the nature of the construct. Choosing and refining the factors in developing an instrument should involve a number of considerations. Pett, Lackey, and Sullivan (2003) state that "the decision as to the number of factors to be retained should be based on an artful combination of the outcomes obtained from the statistical indicators…the factors' theoretical coherence, a desire for simplicity, and the original goals of the factor analysis project" (p. 167). A balance should be struck between parsimony and scale reliability when selecting and refining an instrument's factors.

Four factor structure. The factor analyses conducted on the 2009-2011 data sets from the Revised CCIS provided three years of results indicating that the survey contains four internally consistent subscales that correspond closely to the topic divisions designated during the design of the Revised CCIS. These scales also loosely correspond to the elements of CCIS predefined in the MSP project – curricular content, pedagogy, and classroom environment.

Tables 7 - 11 provide the results of the factor analyses by year for the four subscales and an Overall CCIS scale. Survey items 7, 23, 32 and 41 were open-ended items and so were not included with any of the scales. Items #33 and #34 did not correlate strongly with any of the survey's scales and so were also not included in any of the scales. Scale names for the four subscales are those used in the four sections of the Revised CCIS and are descriptive of the items that comprise each factor.

As can be seen in the tables, factor analyses resulted in high Cronbach's alphas, ranging from .86 to .96, indicating a high level of internal consistency within the scales. With a few exceptions that will be addressed later in this dissertation study, the inter-item correlations fall within a desirable range from .31 to .58, levels that indicate that items are likely tapping into the same construct (perhaps different aspects of the construct) but correlations are not so high as to indicate that items are redundant. Likewise, the corrected total item correlations all fall within a desirable range between .2 and .75, again indicating that the items are substantively correlated, this time with the larger construct (CCI), but are probably not redundant.

In conducting the factor analyses, the Eigen value was set to a minimum of 1.8, ensuring that the amount of variance accounted for by each factor would exceed that of two items. Total variance accounted for by all four factors was calculated at between 55% and 62% over the four years of analyses, with each individual factor accounting for between 12 and 20% of the

variance; these are considered to be acceptable levels for each scale and for the instrument overall.

The four-factor structure was confirmed over the three years of factor analyses, with a fourth year of data analysis added later. These results provide solid evidence of the stability of the instrument's internal structure.

Table 7

Factor Analysis Results for Scale 1, Curriculum Content, for Revised CCIS 2009-2011 data

Year	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
2009	2.28	.88	.89	.57	.70
2010	2.52	.91	.91	.62	.74
2011	2.58	.89	.89	.58	.71

N = 128
Items 1 – 6

Table 8

Factor Analysis Results for Scale 2, Instructional Strategies, for Revised CCIS 2009-2011 data

Year	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
2009	3.12	.89	.89	.34	.55
2010	3.30	.90	.91	.39	.59
2011	3.41	.91	.92	.40	.60

N = 128
Items 8-22

Table 9

Factor Analysis Results for Scale 3, Classroom Resources, for Revised CCIS 2009-2011 data

Year	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
2009	2.40	.88	.88	.48	.65
2010	2.58	.90	.90	.54	.69
2011	2.65	.91	.91	.56	.71

N = 128
Items 24-31

Table 10

Factor Analysis Results for Scale 4, Other Education Related Practices, for Revised CCIS 2009-2011 data

Year	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
2009	2.50	.86	.86	.43	.61
2010	2.35	.88	.88	.48	.65
2011	2.71	.88	.88	.47	.64

N = 128
Items 35-40

Inter-scale and cross year correlations. Evidence of the stability of the internal structure and reliability is also seen in three years of statistically significant positive inter-scale correlations. Two-tailed tests investigating correlations between the four subscales scores and between each subscale and the overall CCIS score were conducted for each year of the three years of data obtained from the administration of the CCIS from 2009 to 2011. Pearson correlations ranged from .42 to .91 and in each case were found to be statistically significant at the $p = .01$ level, providing additional evidence of the instrument's internal consistency, stability and reliability

across time. Table 12 shows the results for 2009 data, Table 13 exhibits results for 2010 data, and Table 14 shows the results for 2011 data.

Table 11

Factor Analysis Results for the Overall Scale for Revised CCIS 2009-2011 Data

Year	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
2009	2.69	.94	.94	.31	.54
2010	2.81	.96	.96	.41	.63
2011	2.96	.95	.96	.37	.59

$N = 128$

Further evidence of internal consistency is found in the inter correlations of CCIS data with future CCIS data. The Revised CCIS scores from 2009 correlate strongly with the 2010, 2011, and 2012 Revised CCIS overall and subscale scores. Pearson's *r* values for the overall scores ranged from .59 to .76, indicating large positive correlations; all were statistically significant at the $p = .01$ level. Subscale *r* values ranged from .43 to .70, indicating medium to large correlations; all were statistically significant at $p = .01$. These results are consistent for both treatment and comparison group data. The findings serve as evidence of internal consistency reliability in that the data produced from the Revised CCIS each year are strongly associated with every other year's data over these four years of its use with a specific population.

Table 12

Inter – Scale Correlations for 2009 Revised CCIS data

	Overall Scale	Curriculum Content	Instructional Strategies	Classroom Resources	Other Educ Practices
Curriculum Content					
Pearson Correlation	.81**	1.00	.61**	.61**	.56**
Significance, 2-tailed	.00	.00	.00	.00	.00
Instructional Strategies					
Pearson Correlation	.84**	.61**	1.00	.42**	.52**
Significance, 2-tailed	.00	.00	.00	.00	.00
Classroom Resources					
Pearson Correlation	.78**	.61**	.42**	1.00	.61**
Significance, 2-tailed	.00	.00	.00	.00	.00
Other Educ Practices					
Pearson Correlation	.82**	.56**	.52**	.61**	1.00
Significance, 2-tailed	.00	.00	.00	.00	.00

N=128

** Significant at p=.01

Table 13

Inter – Scale Correlations for 2010 Revised CCIS data

	Overall Scale	Curriculum Content	Instructional Strategies	Classroom Resources	Other Educ Practices
Curriculum Content					
Pearson Correlation	.87**	1.00	.72**	.71**	.79**
Significance, 2-tailed	.00	.00	.00	.00	.00
Instructional Strategies					
Pearson Correlation	.91**	.72**	1.00	.63**	.68**
Significance, 2-tailed	.00	.00	.00	.00	.00
Classroom Resources					
Pearson Correlation	.82**	.71**	.63**	1.00	.91**
Significance, 2-tailed	.00	.00	.00	.00	.00
Other Educ Practices					
Pearson Correlation	.89**	.79**	.68**	.91**	1.00
Significance, 2-tailed	.00	.00	.00	.00	.00

$N = 128$

** Significant at $p = .01$

Table 14

Inter – Scale Correlations for 2011 Revised CCIS data

	Overall Scale	Curriculum Content	Instructional Strategies	Classroom Resources	Other Educ Practices
Curriculum Content					
Pearson Correlation	.83**	1.00	.68**	.66**	.59**
Significance	.00	.00	.00	.00	.00
Instructional Strategies					
Pearson Correlation	.90**	.68**	1.00	.59**	.63**
Significance	.00	.00	.00	.00	.00
Classroom Resources					
Pearson Correlation	.82**	.66**	.59**	1.00	.59**
Significance	.00	.00	.00	.00	.00
Other Educ Practices					
Pearson Correlation	.83**	.59**	.63**	.59**	1.00
Significance	.00	.00	.00	.00	.00

$N = 128$
** Significant at $p = .01$

Factor alignment with three elements of CCIS. The alignment of the instrument's factors with the three elements of the CCI model used in this dissertation study provides evidence supporting the theory underlying this model. The alignment is imperfect, not surprisingly, given that the three elements (pedagogy, content, and environment) are often interdependent and the lines between them blur in instructional practice. This interdependence is indicated in Figure 1 by the dashed double-headed arrows connecting the three major elements of the models. Curriculum content and pedagogy are contemporaneous in science, for example, when students are building their understanding of the nature of science through the use of inquiry

based learning. An example taken from the Revised CCIS is seen in item #11 *Local tribal elders or other tribal community members were guest teachers,* which on the instrument is grouped under Instructional Strategies, but which consistently loaded on to the Culturally Important Content scale, perhaps because teachers integrating cultural content also invited tribal elders to teach cultural content to their students.

Composite Factors. Closer examination of the factor loadings for the 2009-2012 factor analyses affirms the overlap of elements and items, with some items loading on two or even three scales, and some items changing scales from year to year. The overall pattern of the scales remains relatively constant over the years, as exhibited in Table 15 which displays the items' primary factor associations based on their loadings for the data from 2009 through 2012. Items listed in the table were retained in a given year if they exhibited a factor loading of at least 0.4, a communality of at least 0.4, and a significant loading on one factor (at least 0.1 higher than its loading on other factors). The composite factors displayed in the bottom row of Table 15 were then determined based on items' loading values, communalities, and their pattern of loadings over the four years of factor analyses. Item #12 *Used teaching strategies that support Limited English Proficiency or Second Language Learners* is not listed in the composite factors because of its low communalities in two separate years and its primary loading on two different factors in the other two years. Item #19 *Supported mentoring of students by adults other than the classroom teacher or paraprofessionals* loaded weakly on several factors, with the exception of the 2011 data, so it was not included in the composite factors listed in Table 15. Item #21 *Used observational learning strategies* is not listed in Table 15 because it had no single factor on which it primarily loaded. Item #33 *Communicated with every student's parent or guardian to discuss their student's progress* is not listed because it exhibited weak loadings in two years, no

primary loading in two years, and low communality in three of the four years of analysis.

Table 15

Factor Item Assignments for CCIS Data 2009-2012

	Scale 1 – Cultural Content Items	Scale 2 – CC Instruction Items	Scale 3 – CC Resources Items	Scale 4 – Other CCI Practices Items
2009	1, 2, 3, 4, 5, 6, 11, 16, 20	8, 9, 10, 13, 14, 17, 22	24, 25, 27, 28, 29, 30	35, 36, 37, 38, 39, 40
2010	2, 5, 6, 11, 25, 26, 38, 39, 40	8, 9, 12, 13, 14, 15, 17, 18, 22	3, 24, 25, 28, 29, 30, 31, 35, 36, 37	1, 16, 19, 20, 21, 25, 27, 33
2011	1, 2, 3, 4, 5, 6, 16, 18, 19, 20, 21, 27	8, 9, 10, 13, 14, 15, 17, 22	24, 25, 26, 28, 29, 30, 31	34, 35, 36, 37, 38, 39, 40
2012	1, 2, 3, 4, 5, 6, 11, 38, 39	9, 13, 14, 15, 17, 18, 22	16, 24, 25, 26, 27, 28, 29, 30, 31	12, 34, 35, 36, 37
Composite Loadings	1, 2, 3, 4, 5, 6, 11, 16, 20	8, 9, 10, 13, 14, 15, 17, 18, 22	24, 25, 26, 27, 28, 29, 30, 31	34, 35, 36, 37, 38, 39, 40

Item reduction or revision? When items load weakly or on multiple factors or do not contribute to the instrument's internal consistency, standard practice is to omit them from the instrument or at the very least to revise them. Pett, Lackey, and Sullivan (2003) argue that consideration of an item's importance to the overall construct should be part of the decision-making process regarding the item's omission or revision. They suggest that some of the items in question might be important enough to retain and become part of a new subscale in future versions of the instrument. They recommend that their status should be explained in descriptions of the instrument development.

Likewise, when items load strongly on multiple factors, they are frequently omitted from an instrument. In this case, Pett, Lackey, and Sullivan (2003) again suggest examining the item's

relevance to the construct and, if deemed important enough to retain, they should be placed with the scale that seems to best suit them conceptually. Essentially, the decision making process involves consideration of both quantitative and qualitative issues and attempts should be made to strike a balance between them that is most beneficial to the validity of the instrument.

Items #33 and #34. Generally, the items in each of the four headed sections of the survey correspond to the four factors that comprise the instrument as determined by their factor loadings. The fourth section of the Revised CCIS, "Other Educational Practices," is composed of a collection of nine items, one of them open-ended. One of the eight forced-choice items found in this section of the survey, #33, showed low communality or did not load consistently and significantly on any one of the scales. Item #34 made the cutoff in two of the four years and so was included in the Composite Scale 4, but was weak in the other two years, with low loadings and communalities. In examining the content of all of the items in this section of the instrument it is apparent that the two items in question differ substantively in their content from the rest. The six items that form the scale focus on teachers examining and improving their practice for cultural congruence. For example, item #38 reads *Consulted with tribal elders, culture committees, or other tribal community members about content relevant to Montana Indian tribes,* and item #36 reads *Examined instructional content for cultural bias*. The two items with low communality are item #33 *Communicated with every student's parents or guardians to discuss their student's progress,* and item #34 *Held a private conference with each student to discuss their progress.* While each has culturally congruent overtones – the first in suggesting the involvement of family or community in their students' education and the second in directly involving the student in the regulation of their learning, both aspects of CCI found in the literature and included in Figure 2 - their focus is significantly different from the other six items

in this section of the instrument, a qualitative indication that they might not belong with the other items in Scale #4. Notably, tribal consultants in the development work emphasized the importance of items #33 and #34. This fact, coupled with the items' inclusion in the CCI literature, provide theoretical support for retaining the items in the survey.

Some evidence indicates that the specific wording of items #33 and #34 may also have influenced their behavior. The author fielded questions from teachers completing the survey on the wording of these two items each time the instrument was administered. The word "every" in item #33 and "each" in item #34 were the focus of teachers' questions, suggesting that the items' wordings are problematic. One option for handling the items would be to retain them, given the importance attached to them by the tribal consultants and their alignment with the literature review, but change their wording so that they are not absolute, i.e., requiring the inclusion of every student, which makes them challenging for teachers to attain. In item #33, contacting every student's parents or guardians for example, might be a difficult benchmark for teachers to realize, given that they have little control over the behavior of their students' families. Several treatment group teachers informed the author that though they had tried repeatedly over the years, they had never succeeded in contacting every student's parents or guardians. Commonly, conflicts with parents' work schedules hinder communication with the school. In this study's setting, the negative legacy of education for American Indian people may also be a factor limiting parents' involvement with their children's schools. Further, the lack of telephone and Internet access for some families might play a role in successfully contacting parents or guardians.

Likewise item #34, addressing conferencing with every student, may be difficult for teachers to attain in part due to some students' low attendance rates. Student mobility and

absenteeism rates are high in American elementary schools. The United States General Accountability Office in 2010 reported that 13% of American students change schools four or more times before high school and that mobility rates are higher among minority students and students of low socioeconomic status. The schools where the teachers in this study were employed are no exception; a significant fraction of their K-8 students were low SES and/or minorities with high rates of mobility. In fact, 34% average transience was reported in one of the schools involved in the project (personal communication, April 3, 2010). In light of the theoretical importance of items #33 and #34 to the CCI construct for American Indian students and the evidence that the item wording may be a influencing the item responses, it is recommended that the items be retained in the instrument, their wording revised, and further analysis of their behaviors be conducted in future administrations of the instrument.

Item #12. Item #12 *Use teaching strategies that support Limited English Proficient or Second Language Learners* also did not show a strong pattern of association in the four years of factor analyses. It is possible that, given the small number of ethnic minority students in the overall Montana K-12 student population, Montana teachers have not received much training in ESL and LEP instructional strategies and are not using them regularly in their instruction, which would potentially influence the item's behavior in this study. Prominent scholars of CCI have noted the importance of accommodating language incongruities in teaching diverse students (e.g., Luykx, Lee, Mahotiere, Lester, Hart, & Deaktor, 2007; Nelson-Barber & Trumbull, 2007). Given the significant number of LEP and ESL American Indian students in the classrooms of the teachers involved in this study, this item is highly relevant to the CCI construct. Tribal members on the instrument development team identified language incompatibility between students' home languages and the language of instruction in schools as an issue influencing their students'

science achievement. Thus, on a theoretical basis it is important to include this item on the instrument. Conceptually, the item seems most suited for Scale #2-CC *Instructional Strategies*; it did load strongly on Scale #2 in the 2010 factor analysis. Additional studies of this recommendation to retain the item on Scale #2 are suggested to determine how it affects the behavior of the factor and the instrument overall.

Item #19. Item #19 *Supported mentoring of students by adults other than the classroom teacher or paraprofessionals* loaded weakly on multiple factors for three of the four years of data analyses. It is difficult to determine exactly why this occurred, since response rate means and standard deviations were moderate, indicating that at least some teachers were using this strategy some of the time. This item addresses a traditional learning method for many American Indian cultures, at least outside of formal schooling, so it seems important to retain from a theoretical perspective. One recommendation would be to subject the item to think aloud techniques with study teachers to determine how they interpret the item and whether a revision of the item's wording would be useful.

Item #21. Item #21 *Used observational learning strategies* did not load on any one factor more than once in the four years and so was also not included in the composite factors. Theoretically, this is a very important item, since it represents a well-documented traditional learning method common to American Indian cultures; retention is recommended on that basis. Response means on this item over the four years of data were moderate, ranging from 2.99 to 3.61 on a six-point scale, indicating that on average teachers used observational learning strategies in about half of their lessons. One recommendation would be to subject the item to think aloud techniques with study teachers to determine how they interpret the item and whether a revision of the item's wording would be useful. Again, perhaps further probing regarding how

teachers interpret this item may shed light on how to best handle it.

Sample size. In conducting factor analyses, a large sample size will result in a more stable factor structure. Determining adequate sample size is complex, requiring consideration of item number and factor number. Guidelines for sample size estimation have been proposed by various authors, but there is no definitive agreement on them. Often cited is the recommendation by Tinsley and Tinsley (1987) of between five and ten subjects per instrument item, a guideline that they suggest holds up to a maximum of 300 subjects at which point the ratio of subject to item decreases. Riese, Waller, and Comrey (2000) advise that a sample size of 100 is adequate for a simple factor analysis when item communalities and factor loadings are high. Mundfrom, Shaw, and Tian (2005) use two different means to estimate sample size, the simpler form suggesting between 110 to 180 subjects for a four factor instrument. Using these latter guidelines, the CCIS with 37 retained forced-choice items comprising the four scales should be tested with between 100 to 180 subjects to provide findings of a stable factor structure that can be pointed to with confidence. The study's sample size of 128 that was used for the factor analyses falls within this range. Even so, further testing of the instrument with a larger sample size is recommended to gather more evidence of internal structure stability and validity.

Item interpretation. The interpretation of items is also a concern when developing an instrument and may be a confounding factor that influences reliability and validity. As an example relevant to this study, in small rural communities where many people are related, a casual conversation at a family gathering between a teacher and a student's mother about her child's progress in school might be a common occurrence but whether teachers acknowledge such a conversation as relevant to item #33 when completing the instrument is not known. Think alouds are a recommended technique for establishing evidence of validity and would be valuable

in scrutinizing every item on the instrument. If findings suggest that subjects are misinterpreting items, they can be rewritten to improve their clarity and potentially the instrument's validity.

Culturally congruent instruction or just good teaching? Scale 2, *Culturally Congruent Instructional Strategies*, exhibited the highest mean scale score for each year of teacher data. While all of the instructional strategies included in the items on this scale were identified in the literature and by the instrument development team as culturally congruent practices for American Indian learners, some of the items are also identified in the literature as good teaching practices in general. Examples include item #14 *Provided specific formative feedback to each student*, item #8 *Had students work in collaborative groups*, and item #9 *Used extended wait time in conversations with students*. This raises the question as to what distinguishes these items as culturally congruent instructional practices. As Au (2009) points out, in some cases it is not the strategy itself, but the way it is realized and why it is culturally appropriate in a given setting that distinguishes a strategy as CCI. Again, item #14 presents an example. Providing specific formative feedback to an American Indian student in a manner that observes his or her cultural norms would likely occur one-on-one with a mentor in a private setting, allowing the student to learn from the mentor through conversation and observation, and providing time for private practice or revision before the student demonstrates their proficiency publicly. This type of scenario mirrors traditional learning methods for many American Indian cultures and includes several traits that are depicted in Figure 2 as part of CCI. Item #8 also supplies an example of good teaching overlapping with culturally congruent practice. As discussed earlier in this paper, American Indian cultures tend to be collaborative rather than competitive in nature, so working in collaborative groups in school not only provides the advantages of group work that benefit students in general but also observes the cultural norms of many American Indian students of

working for the good of the group and maintaining humility, thus providing a more natural and relaxed learning experience for them. This overlap between good teaching practices and cultural congruence may in part explain why teachers scored highest on this factor if they are using these particular strategies regularly to support learning for all of their students.

Evidence of Validity

Construct validity. A traditional definition of construct validity describes it as the degree to which an instrument provides data that measure the attribute (or construct) it purports to measure. In the more modern unitary view of validity, construct validity is the overarching framework for all validity and all evidence of validity is gathered to support the inference made about the construct (Messick, 1990). Evidence of construct validity is two pronged, stemming from both theoretical and empirical sources. In the present study, the theoretical evidence of construct validity came from the review of the literature, the expert input gathered during the instrument development work, and the organization of the instrument itself. The empirical evidence was gathered through the analyses of data generated from the administration of the instrument to teachers during the evaluation of a professional development project. Evidentiary types address content validity, criterion related validity, discriminant validity, and convergent validity. Their significance in contributing to the body of evidence for construct validity is discussed in the following paragraphs.

Content validity. Haynes, Richard, and Kubany (1995) state that content validity is "the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose" (p. 238). The *Standards for Educational and Psychological Testing* published jointly by the American Education Research Association and American Psychological Association (1999) define content validity as the relationship

between the content of an instrument and the construct it is designed to measure. Considerations for content validity include all elements of the instrument that may affect the data it generates including the instrument's response format, instructions for using the instrument, and the actual content of the items that comprise it. Content validity is conditional based on many things, for example, on the context in which the instrument is used and the consequences of the data's interpretation. It is threatened by the underrepresentation, overrepresentation, or omission of facets of a construct, or by the inclusion of facets in the instrument irrelevant to the construct it addresses. Because content validity significantly influences the quality of the inferences that can be drawn from the data obtained from an instrument, it is an important part of construct validity.

To promote content validity, instrument development should begin by clearly defining the instrument's intended use, targeted construct, and the construct's operationalization. In the earliest stages of the study, the leadership team accomplished the first task by defining the intended use of the Revised CCIS. It was determined that the instrument was intended to be used as a tool for assessing the CCI of teachers teaching science in Montana schools serving American Indian students from the five prioritized tribal contexts.

The latter two tasks, defining the targeted construct and its operationalization, are commonly accomplished through (a) conducting a literature review and (b) soliciting expert input to ascertain the theoretical underpinnings of the construct, determine the domain of its content, and identify behaviors that operationalize the construct. These methods were used in developing the CCIS in its various forms as synopsized below.

An extensive literature review was conducted prior to the current study resulting in a preliminary definition of CCI, the theory underlying it, and a list of behaviors cited as elements of CCI relevant to classrooms serving American Indian students. This information and

knowledge gleaned from the author's experiences teaching in a tribal school were used to develop a prototype instrument. Local Indian education experts then reviewed the instrument for content validity. They were asked to scrutinize the accuracy and relevance of the items on the instrument, the language used in the items, and whether any items should be added, removed or revised. Their feedback was utilized to modify the prototype instrument.

Stakeholder meetings. In the current study, a larger and more diverse group of education and cultural experts met for two days to again define CCI and delineate its operationalization specifically for the five American Indian cultures involved in the PD project. After a full day of discussion of the meaning and significance of CCI and how it should look in K-8 instruction, the prototype instrument was introduced and scrutinized on Day Two of the meeting by the experts present and their feedback was recorded. The feedback on the instrument and the notes from the two-day discussion were incorporated into the design of the second prototype of the instrument. Multiple parties reviewed and provided feedback on this version of the CCIS including several focus groups of teachers and tribal elders, as well as individual tribal elders, IHE faculty members, other education professionals and statistics and assessment experts. All of the feedback received was used in refining the instrument into its third version, the Revised CCIS.

Community member review. Iterative expert reviews of the instrument were employed throughout the development process to ensure the comprehensiveness, relevance, and clarity of the instrument's items as a means to bolster content validity. Experts contributing to these efforts were tribal elders, tribal K-20 educators, non-Indian K-20 educators, and professional statisticians and assessment developers. The majority of the reviewers were members of the American Indian tribes for whom the instrument was developed, but a significant number of non-Indians also reviewed the instrument and provided advice that was used to refine it. Consultation

with members of the population with whom an instrument will be used to enhance its content validity is a highly recommended but often overlooked instrument development method (Haynes, Richard, & Kubany, 1995; Vogt, King, & King, 2004). Vogt et al. (2004) note that working with experts from the prioritized population is advantageous at the item identification stage to ensure item relevance and adequate representation of the construct, and also later in reviewing items to ensure that they are written in language that is understandable and familiar to the target population. Haynes et al. (1995) note that the use of open-ended conversations with members of the prioritized population can promote content validity by uncovering less obvious facets of the construct that should be included in the instrument and identifying and modifying already existing items in need of refinement. These strategies were integral in the development processes for the Revised CCIS.

Focus groups. The use of focus groups with members of a study's prioritized population is also a strategy valuable to improving content validity. Vogt et al. (2004) note that such focus groups allow researchers to gain insights into constructs from the perspective of the prioritized group, helping researchers to overcome their own cultural biases and avoid ethnocentric assumptions about other cultures. Focus groups also may foster richer conversations than one-on-one interviews as members of the group respond to other members' comments, either agreeing with and expanding them or refuting them, thus providing an additional perspective.

Focus groups can also be advantageous in that some people are more likely to talk when others from their community are present, particularly if the topic at hand is controversial. The current study found this latter point to be true, especially in talking with traditionally raised American Indian elders whose upbringing taught them to be judicious with their words and whose history of oppression in boarding schools may have made them reticent to discuss

education.

A well-structured focus group with an experienced and sensitive facilitator can ease anxious situations and build trust, cultivating a safe environment for conversation. The focus group facilitator in this study was an American Indian woman with extensive experience in conducting focus groups and a calm manner that enabled focus group members to discuss their experiences and feelings toward education. She was deliberately chosen based on her successful experiences working in Indian Country. It was also anticipated that her heritage would be beneficial in fostering a safe environment and building trust with the tribal experts. The effectiveness of the focus group method in enabling candid, relevant and substantive conversation about sensitive topics arguably improved content validity by unveiling important elements of the construct of CCI during the group conversations that might not otherwise have surfaced.

Instrument organization. To maximize its content validity and reliability, the Revised CCIS instrument was physically structured in line with recommendations for instrument development provided in the literature. It is organized into four separate sections with headings and explicit, easy to understand instructions for each. The first section of the instrument includes an example that demonstrates how to use the scale to choose a response that most closely reflects a teacher's perceived frequency of use of a specific instructional practice. Each item in the instrument is written to be unidimensional, clear, and specific. Each section ends with an open-ended prompt and empty text box to allow respondents to add any other examples of CCI practices they use that are relevant but were not included in that section of the instrument.

The four sections of the instrument are titled *Curriculum Content, Instructional Strategies, Classroom Resources*, and *Other Education Related Practices*. The item assignments

for the first two sections of the instrument each align with one of the three major CCI elements proposed in the study's CCI model, namely culturally important content and culturally congruent pedagogy. The third and fourth sections of the instrument align with the third element of the CCI model, culturally congruent environment.

Response scale. When trying to assess frequency, a Likert-type scale is one of the most common and useful scales used in education research. Colton and Covert (2007) recommend using three to seven response categories depending on the instrument topic and the intended audience. These authors also recommend anchoring responses with numbers when needed to quantify the responses for data analysis and/or as a means to clarify their meaning for respondents. The response scale used in the prototype instrument was a four-point scale. Analyses of the data from the use of the prototype exhibited limited variance and so finer grained six-point scales were developed for use in the revised versions of the instrument. The lead author of the instrument consulted the research literature regarding response scales and also worked closely with three assessment experts to devise the new scales. After developing several scale types and examining them with the three assessment experts, six interval Likert-type scales were deemed most appropriate for assessing frequency. These scales provided enough divisions to allow teachers to make fine discrimination, but not so many that they might be overwhelming and difficult to discriminate.

It was anticipated that teachers would require substantively different frequency scales for the various sections of the instrument given that the first three sections of the instrument address everyday instructional practices and issues, while the fourth section mainly addresses less frequent behaviors like consulting with elders about cultural content. It was decided that two different six-interval frequency scales should be included on the Revised CCIS. The scale used

in the first three sections includes six labels, five of which are anchored with quantified frequency intervals in the form of percentages. The first category, *Never*, is not quantified. They read *Never, Seldom - 1 to 20%* [of lessons]*, Sometimes - 21 to 40%* [of lessons]*, Often - 41 to 60%* [of lessons]*, Very Often - 61 to 80%* [of lessons], and *Almost Always - > 80%* [of lessons]. The response scale for the fourth section of the instrument also has six anchored divisions and they read as follows: *Never*, *1 to 2 times per year*, *3 to 4 times per year*, *5 to 6 times per year*, *7 to 8 times per year*, and *9 or more times per year*.

Scale Length. There are no widely accepted rules for setting scale or subscale length. Determining the optimal number of items per subscale and in an instrument overall involves the consideration of several issues. The item number should adequately sample the domain of content for the scale in order to avoid measurement errors. The number should be large enough to result in an acceptable Cronbach's alpha. Too many items can inflate alpha, resulting in high alphas even when the inter-item correlations are low. Too many items can also tax subjects completing the instrument, leading to a bias in their responses due to boredom or fatigue. Researcher recommendations for scale length based on the time it takes to complete the entire scale tend to range from 15 to 30 minutes (e.g., Worthington & Whitaker, 2006). Conversely, having too few items in an instrument can limit reliability. The number of item response choices available also influence alpha, further complicating decisions regarding scale length.

While there is also no widely accepted formula for determining subscale length, Hinkin (1998) suggests that 4 to 6 items per subscale is generally adequate. After factor analysis, scale/subscale length may be modified for a variety of reasons including (a) low communality or inter scale correlations, (b) low factor loadings, (c) multiple factor loadings, (d) the effect of item omission on Cronbach's alpha, and (e) items' conceptual consistency with the rest of the items

on the subscale. Items falling under conditions (a) and (b) are frequently eliminated from scales. Items whose omission would increase Cronbach's alpha are also often eliminated. Treatment of items with multiple factors loadings varies, depending on the strength of the loadings and the conceptual underpinnings of the items. Worthington and Whitaker (2006) and Pett, Lackey, and Sullivan (2003) argue that item retention should not be based solely on statistics, but also by considering qualitatively the conceptual value of each of the items to the subscales and to the scale overall.

The length of time required to complete the Revised CCIS ranges from 10 to 15 minutes, based on observations of teachers in this study. This falls within recommended time limits for avoiding subject bias due to fatigue or boredom. The number of items per subscale listed in Table 18 for the composite loadings in the Revised CCIS varies from a low of seven on Scale 4 (*Other CCI Related Practices*), to eight on Scale 3 (*Culturally Congruent Resources*), to a high of 9 on Scale 1 (*Cultural Content*) and Scale 2 (*Culturally Congruent Instructional Strategies*). All but five items from the Revised CCIS are recommended for intact retention. These are discussed later in this chapter.

Content validity and item alignment. A hallmark of content validity is how well the instrument's items reflect a comprehensive and balanced representation of the content and so adequately operationalize the construct the instrument is intended to assess. Judging how well the construct is represented can be challenging. Nunally (1978) states that "content validity rests mainly on appeals to reason regarding the adequacy with which important content has been sampled and on the adequacy with which the content has been cast in the form of test items" (p. 93).

It can be particularly difficult to fully and accurately reflect the domain of content for

constructs that are more abstract, like some aspects of CCI. CCI is a relatively new construct and still in the early stages of being defined through scholarly study. By its very nature its operationalization is not universal across contexts, i.e., it is context or culturally specific. There is an emerging theoretical foundation to work from in the literature, but it may not be exhaustive in its reflection of the entire domain of content for CCI in a given setting. CCI's domain of content varies among American Indian tribes, let alone among other ethnic groups. Further, because culture is dynamic, ever changing with the times, operationalizing CCI can be a moving target. This also makes it difficult to delineate all of its dimensions, to achieve full content validity.

Expert input and the literature review conducted as part of the instrument development process defined the domain of content for CCI for this study. This domain is represented graphically in Figure 2. The graphic depicts a model of CCI specific to the five tribal cultures involved in the study. It includes a representative set of the elements that project partners agreed characterize CCI for members of their communities.

Examination of the instrument's item alignment with Figure 2 provides a check on its content validity and an assessment of its accuracy and comprehensiveness in representing the content domain. Interpreting each item and strictly aligning each to a single trait on the graphic is difficult due to the previously discussed overlapping nature of the three major elements of CCI and the items that operationalize them. Allowing for such overlap and acknowledging the author's possible bias in interpreting and aligning the items, it is apparent that every trait in the pedagogy and content elements is represented by one or more items, as shown in Table 16.

This is not the case for the third element, Culturally Congruent Environment. Looking at the distribution of items across each of the three major CCI elements of content, pedagogy, and

environment, each is represented with between 12 to 20 items per element as shown in Table 16. Distribution within the pedagogy and content elements is adequate but items are not evenly spread across the traits in the Culturally Congruent Environment element. This suggests that some reexamination and revision of the theory and/or the instrument is warranted. Possibilities for this work are discussed in Chapter Six.

Quantifying CCI. The quantification of content validity, for example through the use of a Content Validity Index technique, did not occur in this study. The author of the study felt that the processes involved in quantifying items were not culturally congruent for the development context at the time the present study was occurring. Asking tribal elders to quantify the importance and relevance of the CCI operationalization items is not naturally compatible with the relativistic perspectives that are commonly associated with tribal people. Further, qualitative discussions of tribal elders' educational experiences at development meetings resulted in tearful stories and emotion-laden responses from participants. Requiring the quantification of a topic as sensitive as CCI would have been disrespectful, awkward, and potentially misrepresentative of the construct, given the emotion that would be involved in such a process. Asking others external to the development team to act as proxies to quantify the items would have created an artificial situation, and the content validity would potentially be negatively influenced. Though it could be considered a limitation at this point in time, the quantification of CCI to improve content validity is recommended as part of further research on the Revised CCIS, if a suitable context for the work can be found.

In summary, the use of numerous and diverse experts to define and operationalize the construct of CCI, their iterative review of the instrument in one-on-one interviews, individual review, and focus groups, the extensive literature review conducted by the author and her

personal experiences with CCI in tribal settings, and the deliberate structuring of the instrument in alignment with recommendations in the literature, were all used to promote the content validity of the data and inferences resulting from the instrument's use. The extent of the work, in terms of the number of development meetings held, the number of reviews the instrument underwent, the long running duration of the work, the diverse perspectives involved, and the depth of the literature review all contributed to building a substantive body of evidence of content validity. Conversely, the underrepresentation of Culturally Congruent Environment is recognized as a delimitation of the study that dampens content validity. Over time, the content validity will likely change, for example, as the instrument is used in additional settings, or the cultures of the tribes with whom the instrument is used evolve. Ongoing research to continually contribute evidence of the degree of content validity is recommended.

Convergent and discriminant validity. Convergent and discriminant validity are two aspects of construct validity that must be considered together if they are to be meaningful as evidence of validity. Convergent validity measures the degree to which two instruments that purport to measure the same construct actually do measure the same construct. Conversely, discriminant validity measures the degree to which two instruments each measure a different construct (Web Center for Social Research Methods, 2006). Evidence of both types must be present in order to provide meaningful evidence of convergent validity. Both were gathered as part of the present study and are described below.

Convergent validity. Two instruments, the Teacher Multicultural Attitude Survey, or TMAS, and the Culturally Responsive Teaching Self-Efficacy Scale, or CRTSES, were administered concurrently with the Revised CCIS in spring 2010 as a means to investigate convergent validity, as described earlier in this paper. These instruments were chosen from a set

Table 16

CCI Trait Alignment with Items

Culturally Important Content	
Trait	Item
Oral tradition	1
Traditional skills and knowledge	3
Contemporary and historic	2, 3
Students' life experiences	2
Native language	16
Indigenous science	5
Community based	18
Place based	4, 8
Practical and applied	13, 18
Culturally Congruent Pedagogy	
Trait	Item
Collaborative	8
Fine arts based	20
Observation and practice	21, 17
Authentic and practical	13, 18
Student centered	9, 10, 14, 34
Oral tradition	1, 11
Fluid time	9,22
Expert – Apprentice	11, 14, 19, 34
Metaphors and symbols	15
Community based	11, 18, 33
Culturally Congruent Environment	
Trait	Item
Relationship/community	33, 39, 38
Equity/freedom	10
Inclusive/respectful	33, 34
Accurate/appropriate/ubiquitous	35, 36, 38, 40
Fluid time	22
Spiritual/holistic	-
Generosity/reciprocity	-
Humility	-
Patience/quiet	9
Teasing/humor	-
Cultural artifacts	24, 25 26, 27, 28, 29, 30, 31

of instruments found in the literature that were designed to assess attributes believed to be associated with CCI. Initially, the data from the CRTSES and the TMAS were analyzed to affirm evidence of their internal consistency reliability. Cronbach's alphas and inter-item correlations provide evidence of the two instruments' internal consistency, as shown in Table 17.

Table 17

Reliability Analysis for CRTSES and TMAS 2010

Instrument	Grand Mean Score	Cronbach's alpha	Standardized Cronbach's alpha	Inter Item Correlation	Corrected Item Total Correlation
CRTSES	3.33	.97	.97	.47	.67
TMAS	4.05	.84	.87	.25	.46

N = 128

Ponterotto, Bluch, Greig, and Rivera (1998) designed the TMAS to assess teachers' multicultural awareness. They make the case that multicultural awareness is positively correlated with multicultural instructional practice when they state:

> A critical step in providing multicultural education involves the cultural awareness and sensitivity of teachers. In fact, if teachers are not culturally aware of their own socialization biases…then multicultural education efforts are likely to be ineffective …. The construct of "multicultural awareness" described herein refers to teachers' awareness of, comfort with, and sensitivity to issues of cultural pluralism in the classroom. Furthermore, teachers high in multicultural awareness see cultural diversity as a strength and feel the responsibility to address multicultural issues in the curriculum and in the teaching/learning process. (p. 1002)

Based on this argument, it was hypothesized a priori that the TMAS should produce data that positively correlate with data from the Revised CCIS, i.e., teachers who had high levels of

multicultural awareness would be more likely to use CCI more frequently. Ponterotto et al. (1998) also provided psychometric data indicating acceptable levels of validity, reliability, and internal consistency associated with the instrument, thus making it even more valuable for use in the present study.

The analyses did indeed provide evidence of convergent validity. Tests for correlations were conducted for the overall scores for the 2010 TMAS and the 2009-2011 Revised CCIS overall scores and scores for the four Revised CCIS subscales. Thirty-six missing scores from the 128 total for the 2011 Revised CCIS were imputed based on teachers' average scores for 2009 and 2010 . This was a conservative method of imputation and potentially underestimated the 2011 Revised CCIS scores, given that both treatment cohorts experienced their biggest increases in Revised CCIS scores at the end of their second year in the program, which would have been 2011 for the Cohort 2 teachers. Correlation results indicated that the two instruments' overall scores have a medium to medium-large positive correlation, with Pearson's *r* ranging from .35 to .47, all of which are statistically significant at $p = .001$ for a two-tailed test. Likewise the scores for the four Revised CCIS subscales for the three years have a medium to medium-large positive correlation with the overall scores for the TMAS, with Pearson's *r* values ranging from .32 to .44. The critical value of *r* for a two-tailed test at $p = .001$ for the sample size of 128 used is .287. All results exceeded this number so all were statistically significant at the $p = .001$ level. Correlations for the overall scores for the TMAS with the Revised CCIS are shown in Table 18.

The CRTSES was chosen as a second convergent validity instrument based on a different CCI related criterion. This instrument was designed to "assess teachers' self-efficacy to execute practices of culturally responsive teaching" (Siwatu, 2006, p. 4). It was hypothesized a priori that

there would be a positive correlation between scores on the CRTSES and the Revised CCIS. i.e., teachers who reported high self-efficacy toward executing CCI would be more likely to use CCI in their classrooms. The instrument also was seen as a good candidate for convergent validity evidence because it was backed up with psychometric data indicating acceptable levels of reliability, internal consistency, and validity.

Tests of convergence were conducted using the CRTSES data gathered from all teachers in spring 2010 and their Revised CCIS data from 2009-2011. Again, missing data for 36 teachers for the 2011 Revised CCIS were imputed using the average scores for 2009 and 2010. It was predicted a priori that the 2010 CRTSE overall scores would show positive correlations with the Revised CCIS overall scores and the scores for the four subscales from 2009 to 2011. Indeed, the CRTSES and the Revised CCIS overall data showed positive correlations over the three years of CCIS data, with Pearson's Correlations ranging from .44 to .49. These are medium-large correlations according to standards set by Cohen (1992) for Pearson's *r* effect size. Likewise the scores on the four Revised CCIS subscales from 2009 to 2011 showed medium to medium large positive correlations with the 2010 CRTSES overall scores, with Pearson's *r* ranging from .29 to .47. Again, the critical value of *r* for a two-tailed test at $p = .001$ for the sample size of 128 used is .287. All results exceeded this number so all were statistically significant at the $p = .001$ level. Correlations for the overall scores for the CRTSES with the Revised CCIS scores are shown in Table 18.

Guidelines in the literature for evaluating the strength of association for Pearson's *r* vary somewhat, but standards set by Cohen (1992) are often cited as a rule of thumb. His standards for social science research set an *r* of .3 as the lower limit for a medium correlation while an *r* of .5 or above is considered a large effect size. By these guidelines, all of the values of *r* found in

the convergent validity analyses in this study fall in the medium to large range. The combined evidence provided by the correlations between the Revised CCIS with both the TMAS and the CRTSES increases confidence in the findings of convergent validity.

Table 18

Correlations for 2010 CRTSE and TMAS Overall Scores with 2009-2011 Revised CCIS Overall Scores

	2010 CRTSE	2010 TMAS
2009 Revised CCIS		
Pearson Correlation	.48**	.41**
Significance	.00	.00
2010 Revised CCIS		
Pearson Correlation	.49**	.47**
Significance	.00	.00
2011 Revised CCIS		
Pearson Correlation	.44**	.35**
Significance	.00	.00

N=128

** Significant at *p*=.001

*Discriminant validity***.** To investigate the discriminant validity of the data produced by the Revised CCIS, they were analyzed for correlation with teachers' 2010 science content test scores. Treatment and comparison teachers in the study completed content tests each year as part of the PD project evaluation and research, but there were no theoretical grounds to believe that the science content test scores would exhibit a relationship with the Revised CCIS scores. It was hypothesized a priori that teachers' scores on the content test would not show a significant correlation with their Revised CCIS scores. All of the treatment and comparison group teachers' scores were included in the analyses for a sample size of 128. Analysis of the three years of CCIS overall score data with the 2010 content test scores resulted in Pearson's r ranging from

.01 to .12, thus confirming the lack of correlation between the two instruments. The subscales also showed very low correlations, with only one outlier of r = .22 on the 2011 Scale 4 - Other Educational Practices that did reach statistical significance for p = .05. Even so, this indicates a weak correlation; Cohen's standards set correlations less than .3 as weak. Table 19 shows the results for the discriminant validity tests for the overall instrument scores.

Table 19

Correlations for the 2010 Science Content Test with the 2009-2011 Revised CCIS Overall Scores

	2009	2010	2011
Pearson Correlation	.01	.10	.12
Significance	.88	.29	.19

N=128

The Revised CCIS data's medium to large positive correlations with the TMAS and CRTSES data as evidence of convergent validity coupled with the weak correlations with the science content test scores as evidence of divergent validity together contribute substantive evidence of construct validity for the Revised CCIS as an instrument that measures the intended CCI construct.

Criterion related validity. Criterion related validity is the degree to which the data produced by an instrument can predict or are statistically related to an outcome or criterion (United States Office of Personnel Management, 2014). It is a measure of how well the operationalization of a construct performs in practice to generate data that can be used to accurately estimate a quality or outcome. It is often said to have two forms, concurrent and predictive. In this study, concurrent validity was collected in two ways.

Concurrent validity. Concurrent validity is a type of criterion related validity defined as the ability of the operationalization of a construct to produce data that estimate a current status or

outcome that it theoretically should be able to predict (Web Center for Social Research Methods, 2006). In this study, evidence of concurrent validity was collected through the analyses of data produced from the concurrent administration of the Revised CCIS, the TMAS and the CRTSE in spring 2010. The results of these analyses found medium to large positive correlations between the 2010 data produced by the three instruments. An *r* value of .47 was calculated for the correlation between the 2010 Overall Revised CCIS scores with the scores for the CRTSE; an *r* value of .49 was found for the correlation between the 2010 Revised CCIS and the TMAS. Both were statistically significant at $p = .001$. The simultaneous administration of the instruments makes the results applicable as evidence of concurrent validity as well as convergent validity, as discussed in the preceding section.

Concurrent efficacy assessment. A second form of concurrent validity evidence was seen in the use of the Revised CCIS to assess the efficacy of the PD project in improving teachers' use of CCI. The instrument was administered annually to two different cohorts of treatment teachers and a matched comparison group of teachers over the life of the PD project (N=128). It was predicted a priori that treatment teachers' CCI would improve after they joined the PD project. Based on the assumption of validity of inferences resulting from the instrument's use, it was predicted that this change would be reflected in an increase in treatment teachers' mean Revised CCIS scores. At the same time, it was predicted that comparison group teachers' mean scores would not change significantly. It was further predicted that after they joined the PD project, treatment teachers' mean scores would become statistically significantly different from comparison group teachers mean scores and this difference would increase over time.

ANOVA and *t* tests were used to investigate the differences in groups' mean scores. Because the Cohort One Treatment teachers completed the second prototype CCIS for setting

their baseline scores, while the Cohort Two Treatment and all comparison teachers completed the Revised CCIS for their baselines, only the 29 items common to both versions of the instrument were used (N=128) in the ANOVA and related tests. Results of the Levene's Test of Homogeneity of Variance provided no evidence that the data were not homogeneous in their variance. As expected, results of the ANOVA and the post hoc tests, Tukey, Scheffe, Bonferroni, Sidak, Gabriel and Hochberg, all indicated that there was no statistically significant difference between the baseline scores for the treatment and comparison groups ($F = 2.192$, $p = .093$).

Results of the *t* tests showed that in nearly every case, treatment teachers' mean scores were statistically significantly higher than those of the comparison group teachers. Two exceptions to this trend were found in the 2009 data for the Curriculum Content and Classroom Resources scales; in these two cases no statistically significant differences in the means of the treatment and comparison group teachers were found. This effect may have in part been due to the contribution of the baseline scores for Cohort Two treatment teachers, which would have depressed the treatment teachers' 2009 mean scores.

As anticipated, as the PD project years passed and treatment teachers experienced more of the intervention, their mean scores gradually rose while comparison group scores remained nearly unchanged. By the end of the second year of project membership for each treatment group (2009 for Cohort One and 2011 for Cohort Two) their mean scores showed a statistically significant positive difference with the comparison group's scores. These results affirm the prediction that treatment teachers' scores would rise over their years in the project while comparison group teachers' scores would remain flat over the same time period. Because the score differences occurred concurrently with the PD intervention, they provide evidence of the concurrent validity of CCIS data-based inferences regarding the teachers' CCI use. This

outcome is not surprising, given the care taken to ensure content validity during the extensive instrument development process. The dual role that several people played in designing and delivering the intervention and participating in the instrument development meant that an alignment between the two was likely, increasing the likelihood that specific items on the instrument would be addressed in the PD and the likelihood that they would be reflected in teachers' instructional practice.

This chapter presented and discussed the results of the statistical analyses of data generated by the 2009 – 2012 administrations of the Revised CCIS. The amalgamation of the evidence of reliability, factor structure, and internal consistency with that of the content, convergent, discriminant, and concurrent validity presented in this chapter create a solid body of evidence that legitimizes inferences drawn from the use of the Revised CCIS in the prioritized tribal contexts. The next and final chapter of this dissertation study will include a discussion of the limitations and delimitations of the study, recommended future research focuses, and the significance of the study's findings and products to furthering CCI research.

CHAPTER SIX – DISCUSSION AND CONCLUSIONS

This study examined the development and validation of the CCIS as an instrument designed to assess teachers' culturally congruent instructional practice. The research questions focus on two separate but intertwined topics: the nature of the instrument development process and the nature of the instrument itself. The cultural incompatibilities that exist between American Indian cultures and the Western science paradigm give examination of these two questions significance to those interested in research and education with Native people.

This chapter begins with an examination of issues related to validity and an Indigenous research paradigm and their significance in the outcomes of this study. It continues with a discussion of the limitations and delimitations of the study, moves to recommendations for future research, and ends with concluding remarks about the study and the significance of its findings and products.

Validity and an Indigenous Research Paradigm

Many of the important outcomes of this dissertation study center on ideas related to the interplay of validity and an Indigenous research paradigm, ideas that are topics of current debate in the Indigenous research community and the research community at large. The significance of these ideas and their interdependence with the outcomes of this dissertation study warrant additional examination.

Relevance to this study. The positions held by Moss et al., (2005), LaFrance, Kirkhart, & Nichols (in press), Wilson (2008), Kovach (2009), and Anderson et al (1994) regarding the related ideas of process validity, doing things the right way, validation emanating from the community/context, and relational accountability, were integral to this dissertation study. Each of these ideas addresses the influences of context on research, an issue that was central to the

outcomes for both Research Question 1 and 2.

The findings for the first research question in this dissertation study, investigating what a culturally congruent instrument development process looks like, are centered in these ideas. The process developed and used in this study, which proved to be successful in the diverse tribal and K-20 education contexts involved, paid close attention to doing things the right way for the specific communities involved, thus supporting process validity, and contributing therefore to construct validity. The strong relationships with and collaborative expertise of the research partners were key in designing and implementing a respectful and responsible culturally congruent development process that observed cultural protocols and enabled the representation of multiple perspectives in the work. The multiple methods for collecting data were culturally congruent, as explicated in Chapter 4, and contributed to process validity in accommodating the cultural contexts in which the work was done. The triangulating methods and data types used to represent the study's diverse research partners also contributed to process validity. Efforts to address the reciprocity aspect of using the knowledge gained from the study is underway at this time, with the sharing of the instrument and development process with tribal partners, K-20 education professionals, and education researchers and evaluators through presentations and publications of the work.

In investigating the second research question regarding the technical quality of the instrument and gathering evidence of validity, the validity emanating from the community's involvement in the process complements the evidence of validity collected through statistical means. The operationalization of culturally congruent instruction and the iterative review and modification of the CCIS by a wide range of research partners through multiple methods (group meetings, focus groups, interviews, electronic mail feedback) enabled community based

validation and contributed to process and construct validity.

The constructs of process validity, relational accountability, validity emanating from the community, and doing things in the right way were interdependent, interwoven and integral to this dissertation study. The conceptualization and reconceptualization of these ideas is currently occurring through the work of a number of researchers, both Indigenous and non-Indigenous. They are gaining prominence as important and evolving issues as Indigenous communities increase their roles in conducting research meaningful to their communities. The outcomes of this study in developing a process for instrument development and collecting community based evidence of the instrument's quality bear witness to the importance of these ideas to research in Indigenous communities.

Limitations and Delimitations

The following paragraphs discuss the delimitations and limitations that were identified as relevant to this study.

Generalizability. Krathwohl (1998) states that "evidence of validity is always gathered in a particular context, and one of our concerns must be the generality of that context" (p. 432). The Revised CCIS was developed for use with five specific American Indian cultures in Montana. This was noted as one of the delimitations to the study, and is potentially a factor that presents a limitation to the valid use of the instrument in other cultural settings, i.e., its generalizability. There are over five hundred tribal cultures still flourishing in the United States, each one unique but frequently sharing some characteristics, especially with neighboring tribes. The partners involved in the Revised CCIS development agree that the instrument should be adaptable for use with educators working with students from other American Indian cultures. Indeed, the Revised CCIS is being used in other tribal settings at this time, as an assessment and

evaluation instrument (e.g., Nam, Roehrig, Kern, & Reynolds, 2012; Reynolds & Kern, 2012). Researchers wishing to use the instrument in other tribal cultures are advised to carefully review it and adapt it for use in their specific context. There is evidence that people external to the project and representing cultures other than those involved in this study also consider the instrument adaptable for use in their settings. In the section of this chapter that addressed Research Question #1, there is a description of a professional presentation in which audience members from different tribal groups scrutinized the instrument and discussed viable ways that it could be adapted for use in their cultural settings. So while the instrument may not be fully generalizable across cultural settings, the examples presented here provide evidence that it is likely adaptable for use in other tribal settings. Partners on the development team expressed that they believed that both the instrument and the methods used to develop it are likely to be valuable as models for researchers interested in developing instruments for use in other contexts.

Teacher selection bias and sample size. The PD project in which this study was situated required that the subject pool be delimited to teachers from the specific cultural contexts that were prioritized for the project. Teachers who comprised the study's treatment and comparison group teachers were not randomly selected. Randomization is a desirable condition for improving the rigor of the type of research conducted in this study; the inability to randomly generate the teacher groups is noted as one of the study's limitations. The inability to randomize was largely due to the small group of available teachers who fit the treatment teacher profile – a grades 3 to 8 teacher, who had taught at least 5 years but was at least 7 years from retirement, who held respect among their peers as a competent teacher with leadership potential, who was teaching in a school with significant American Indian student enrollment, and in a school on or near the Flathead, Northern Cheyenne or Crow reservations. In the relatively small communities

found on these reservations, the schools are also small and so the teacher pools are limited; in many cases, for example, there may only be one or two second grade teachers in a school. Finding ninety treatment teachers who fit the project's teacher profile in these small pools was very challenging. The same challenge was faced in generating a matched comparison group of teachers to participate in the study.

School administrators were approached by project leadership, given the project criteria for teacher participation, and asked to apply those criteria in recommending teachers for participation in the project. In all cases, treatment teachers voluntarily applied to join the project. Approximately half of them applied after their administrators recommended them as candidates. When the time arrived for working with the second cohort of treatment teachers, the pool of potential candidates was even more limited by the exemption of the Cohort 1 teachers, so the challenge to meet the project teacher criteria was escalated, and as a result, the criteria were somewhat relaxed – e.g., some newer teachers and those teaching lower elementary students were selected to join Cohort Two. It was a constant balancing act to maintain a useful sample size while also selecting teachers who fit the research study's theoretical profile.

The sample sizes used in the data analyses generally met or exceeded the lower limits recommended in the literature for attaining acceptable power levels. The multiple years of data analyses that mostly reaffirmed findings from year to year or that allowed the examination of trends over several years may be viewed as partially compensating for the limitations of the non random selection of subjects and moderate sample sizes. On the other hand, these limitations are important ones that should be considered in drawing conclusions about the evidence of validity generated in the study. Recommendations for future research would include efforts to overcome the limitations on random selection and sample size to allow increased power of the statistical

analyses and increased confidence in research findings.

Item interpretation and language incompatibility. The issue of item interpretation was examined briefly in an earlier section in the consideration of item omission or revision for five of the instrument's items. Concerns regarding interpretation may be magnified as a limitation in the present study, given the significant number of teacher participants from the Northern Cheyenne and Crow schools who could be classified as Limited English Proficient (LEP) and, in some cases, as English as a Second Language (ESL) speakers. Many adults living on these two reservations were fluent in their native language well before they learned to speak English. Their native languages, Northern Cheyenne and Crow, are not related to English, thus potentially increasing the likelihood of interpretation errors. The challenges of language incompatibilities in conducting research in native communities were discussed in an earlier chapter of this paper. The influence of language incompatibilities on assessment performance for Native people is well researched and documented by scholars such as Nelson - Barber and Trumbull (2007) and Solano – Flores and Nelson - Barber (2001). Likewise, Lee and her colleagues have conducted extensive research on the interaction of language and assessment interpretation in Latino communities in the United States (e.g., Luykx, Lee, Mahotiere, Lester, Hart, & Deaktor, 2001). Their research suggests that a wide variety of "languacultural" factors influence assessment interpretation and performance, noting that "Languages are tightly bound to social and cultural contexts in which they are used; lexical, morphological and grammatical elements embody culturally specific ways of conceptualizing the natural and social world" (Luykx et al., 2001, p. 903). Specifically, they have found interferences in assessments arising from textual organization, and orthographic, semantic, and phonographic factors, as well as cultural beliefs and practices.

The instrument development processes were designed to minimize the language incompatibility issue through the integral involvement of tribal partners in the operationalization of CCI and the generation of the instrument's items. Iterative review of the item wording by multiple tribal partners from the development team and beyond also contributed to minimizing undesirable effects of language incompatibilities. Even so, researchers are advised to keep this factor in mind when using the Revised CCIS. Deeper examination of item interpretation is a recommended focus for further research on the instrument. Think alouds or other methods for assessing subjects' interpretations of items can be used to examine item interpretation. Findings should be used to clarify the wording of items to limit response bias and increase confidence in the validity of subjects' responses.

Frequency of CCI versus the nature of CCI. Also identified as a delimitation of the research is the fact that the instrument was designed only to assess the *frequency* of use of CCI in teachers' practice. The instrument does not probe the nature of the CCI and how it is implemented in teachers' instruction or how it appears to influence students' learning. These are very important ideas that are recommended for future research. Gathering detailed information about the way that CCI is occurring in science education, e.g., the specific content being taught, the way that elders are interacting with students, the kinds of activities that occur on fieldtrips to culturally significant sites, the products that students are generating, the learning objectives that are the focus of instruction, leans toward the use of more qualitative methods such as interviews, focus groups, case studies, teacher journals, and observations of instruction. This would be a worthy follow up study to the one described in this paper. The information gathered regarding the nature of CCI (as opposed to the frequency) would provide a rich and informative picture of CCI that would be valuable in furthering understanding of CCI in American Indian education.

The time and other resources required to complete this type of research would be significant. Resource constraints were the main reason that the current study was delimited to the development of an instrument that assessed frequency only.

Inadequate representation of Culturally Congruent Environment. The element *Culturally Congruent Environment* is somewhat unique from the other two elements of the CCI model (content and pedagogy) because it is a more complex and less tangible subconstruct than the others on some levels and therefore is more difficult to operationalize. The environment or ambience of a classroom is complicated because it is decidedly influenced by interactions and resources that can be observed, defined, and assessed, but is also influenced by things whose explication and assessment are more difficult. Examples of the latter include axiology, spirituality, and subtle cultural norms. Figure 2 attempts to portray the complexity of the environment element of CCI.

The complexity and interdependent nature of a culturally congruent classroom environment was a topic of conversation that emerged organically in the instrument development meetings. Participants questioned the measurability of environment and whether its aspects were something that could be taught to and used by teachers from other cultures who had been raised with their own set of norms, values, and spirituality, for example. It was noted that environment is ubiquitous and somewhat abstract, part of a way of life that is difficult to define and includes the interaction of values, behaviors, emotions, and personalities. One Native teacher on the development team stated that environment is intertwined with and a part of both pedagogy and content in the classroom and that separating it out to measure it would be difficult. It was decided during the two-day development meeting that the work of defining and operationalizing Culturally Congruent Environment in a more thorough manner would have to be done at a later

time in a separate effort, due to the short time left before the instrument would again be administered. It was acknowledged as a delimitation of the study that this element was not likely to be fully represented in the Revised CCIS. Indeed, while factor analyses revealed that there is an instrument scale that aligns with and represents the major element Culturally Important Content and another one that represents Culturally Congruent Pedagogy, there is no one scale that adequately represents the third major theoretical element, Culturally Congruent Environment. Instead this element aligns with all of Scales 3 and 4, along with several items from Scale 2, indicative of its multi-faceted and interdependent nature. Items associated with this element do not exhibit a balanced distribution across the element's traits, thus portending a misrepresentation of the element.

Recommendations for Future Research

Operationalizing and assessing Culturally Congruent Environment. Even in the face of the potential challenges posed by refining the instrument to better operationalize and assess culturally congruent environment, tackling these challenges is recommended for two reasons. First, the nuance and complexity involved in creating a culturally congruent classroom environment makes it a more elusive goal to attain than the other elements of CCI, particularly for cultural outsiders. Better defining and operationalizing CC environment would be a valuable first step toward meeting this goal by helping teachers, especially non-Indian teachers in this case, to understand the kinds of attitudes, beliefs, behaviors and resources that foster a culturally congruent environment supportive of American Indian students' learning.

Second, along with validity, reliability and an internally stable structure, parsimony is a desirable quality in an instrument like the Revised CCIS. Examining the CCI domain representation, the instrument's subscales, and the mix and unequal distribution of items that

align with the environment element of the CCI model suggests that there is room for refinement of the instrument to improve its parsimony. Refinement of the theory underlying the domain model may also be warranted. It would be worthwhile to expand the research to reexamine both the theory underlying the model and the instrument's structure, tinkering with the items to explore whether a single scale for Culturally Congruent Environment is feasible and merited or whether this element of the theory should be modified significantly.

It is anticipated that further operationalizing Culturally Congruent Environment would require a resource intensive effort, with much time devoted to in-depth study of the literature, substantive conversations with minority group stakeholders, the development of additional survey items, and additional collection and analyses of data. Ample time for relationship and trust building should be expected. Researchers who undertake such work are advised to first develop deep understanding and respect for the cultures in which they intend to work. That knowledge should then be deliberately applied to honor norms, axiologies and other aspects of culture to ensure respectful and meaningful interactions with stakeholders. Conversations will likely tread on delicate ground, requiring that researchers work with heightened cultural sensitivity, both in interacting with community members and in transferring their ideas into items for the instrument. It is recommended that any work done in this realm be thoughtfully documented and disseminated to the research community, to provide a resource for guidance on conducting such work in a culturally congruent manner.

Suggestions for improving the representation of environment. One suggestion for modifying the instrument to improve the operationalization of Culturally Congruent Environment is to collapse Scale 3-Classroom Resources into fewer items by developing broader items that encompass several of the current items in the scale. For example, instead of listing

each specific type of classroom resource – books, music, and others – as separate items, one could combine items into a single survey item such as *Artifacts reflecting tribal culture such as books, music, paintings, beadwork, tools, toys, etc.* The inter-item correlations for this scale's items are high relative to those of other scales (in many cases greater then .65), suggesting that there may be some redundancy, further warranting a reduction in the number of original items in this scale.

A suggestion aimed at improving the balance of representation for Culturally Congruent Environment would be to include additional items that more directly address some of the more abstract aspects of the element, such as cultural norms, values and perspectives. These are currently reflected in only a few items in the instrument, suggesting that they are underrepresented as important facets of culturally congruent environment. Because these aspects tend to be more abstract, writing valid items that adequately represent them is challenging. It is recommended that such work be done with tribal partners from the prioritized cultures whose insights and input will likely be better able to capture the nuance of these types of ideas.

Item interpretation and think alouds. Verifying respondents' item interpretation assists the researcher in gauging the clarity of instrument items and provides information useful in refining items to improve the validity of the responses they evoke. In this dissertation study, additional research in which subjects are asked to describe their interpretations of items using a think aloud protocol would have been beneficial in determining whether misinterpretation was a factor influencing teachers' responses, especially on the four items that did not load reliably or that showed low communality. The think aloud technique or another method for checking item interpretation is recommended for future research on the instrument. This method is especially relevant and useful in studies such as this one in which some participants do speak English as

their first language.

Concluding Remarks

Across nearly every ethnic group and academic subject in America's schools, non-White students continue to exhibit lower achievement compared to their White counterparts. The United States Census Bureau's latest prediction forecasts that by 2060 sixty seven percent of the country's population under the age of 18 will be non-White (United States Census Bureau, 2012). With predictions for their growing representation in American schools, the urgency for finding solutions to the achievement differential for ethnically diverse students is also growing.

Many American Indian tribes espouse the use of culturally congruent instruction to better meet the needs of their students and thereby increase their achievement. The United States government and professional organizations for education have likewise advocated for and in some cases mandated the use of CCI as a means to improve American Indian students' achievement. A growing body of research provides evidence of the positive correlation between CCI and improved achievement. Study in this field is becoming a major thrust of educational research, perhaps partly in response to changing demographics in the Unites States and beyond. Given that it is a relatively young field of research, questions for investigation regarding CCI are many including What is CCI? What does it look like in practice? How can proficiency in CCI be developed in teachers? How can it be assessed? and How does it influence student achievement?

This study addressed two research questions related to the assessment of CCI in American Indian contexts, What is a culturally congruent process for developing a valid instrument for assessing the use of CCI in teaching science with Montana American Indian students? and What is the technical quality of such an instrument? Investigating these questions resulted in a culturally congruent instrument development model, a model of CCI composed of

three major elements (content, pedagogy, and environment), an instrument known as the Revised Culturally Congruent Instruction Survey, and evidence of validity for the use of the instrument to draw inferences regarding CCI in Montana tribal contexts. Development model processes included an extensive literature review in conjunction with participatory and culturally congruent collaborative work with Indian and non-Indian K-20 education professionals, tribal culture representatives from five Montana tribal cultures, and professional assessment experts and statisticians. The processes assumed a non- hierarchical structure and deliberately observed cultural norms to assure that all stakeholders would have equitable opportunities for contributing to the work. The design of the development processes helped to ensure the content validity of the instrument for use in the specified context.

The instrument was used in evaluating the CCI of K-8 teachers serving students from the five prioritized tribal cultures. Non-random groups of treatment and comparison group teachers completed the instrument annually over several years. Analyses of the data generated evidence of temporal reliability, internal consistency, and a four factor structure that is similar to the organizational layout of the instrument. Analyses also provided multiple forms of evidence of content, concurrent, convergent and discriminant validity.

The study's delimitation of partially representing the element of Culturally Congruent Environment in operationalizing CCI in the instrument items means that a significant portion of the three-pronged theory on which the study was premised is not addressed in the instrument. Given the elusiveness of developing proficiency in the abstract traits of CCI, particularly for cultural outsiders, further research is recommended to attempt to define and operationalize these traits to finish out the instrument as a more comprehensive measure of CCI.

The study produced substantive evidence of the instrument's ability to generate data from

which valid inferences can be drawn, at least for Montana tribal contexts. Validity is not an all or nothing condition, but rather occurs on a continuum. The gathering of evidence of validity is typically a protracted process involving various techniques conducted over a series of studies and in a range of contexts. Thus the degree of validity associated with an instrument's use typically changes over time as more evidence is gathered, and can vary with the context and purpose for which the instrument is used. The stronger the body of evidence accumulated, the greater the researcher can be assured that the inferences made using the instrument in a given context is accurate and useful. It is hoped that the Revised CCIS will continue to be used in research and that its validity is further investigated.

A word of warning is advised for researchers considering the use of the Revised CCIS. The evidence of the technical quality of the instrument gathered in this dissertation study is only truly applicable within the research contexts in which the Revised CCIS was developed, i.e., in the five tribal communities that participated in the study. Using the instrument in other settings should be done with caution and with recognition of its context specific nature. It is recommended that rather than use the instrument intact in other research contexts, research partners develop their own instrument specific to their context, or at the very least adapt the Revised CCIS to support its valid use elsewhere.

While the context specific nature of CCI means that the Revised CCIS will likely require adaptation if used in contexts outside of the one for which it was designed, it holds significance to the research and education community in providing a template for the operationalization of CCI and its assessment. Likewise, the development process model, in demonstrating the use of culturally congruent practices to equitably engage stakeholders in instrument development, has potential value as a resource for guiding those looking to work with communities to develop an

instrument of this nature. Used properly, the instrument and development model have potential to move the research base forward regarding CCI, worthwhile goals that may assist in the attainment of equitable educational outcomes for all students.

References

ACT Information Brief 2012-28. (August, 2012). *College degree completion rates by race/ethnicity and college readiness.* Retrieved from http://www.act.org/research/researchers/briefs/2012-28.html

Aikenhead, G., & Michell, H. (2011). Bridging cultures: Indigenous and scientific ways of knowing nature. Toronto: Pearson.

American Indian and Alaska Native Education. 64 Fed. Reg. 54622-54624. (October 7, 1999). Retrieved from http://www2.ed.gov/legislation/FedRegister/announcements/1999-4/100799a.html

American Indian and Alaska Native Education (2004). 69 Fed. Reg. 25295-25297. (April 30, 2004). Retrieved from http://edocket.access.gpo.gov/2004/pdf/04-10377.pdf

Almy, S., & Theokas, C. (2010). Not prepared for class: High poverty schools continue to have fewer in-field teachers. Education Trust. Retrieved from http://www.edtrust.org/dc/publication/not-prepared-for-class-high-poverty-schools-continue-to-have-fewer-in-field-teachers

Anderson, G., & Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher, 28* (5), 12–21, 40.

Anderson, G., Herr, K. & Nihlen, A. (1994). *Studying your own school: An educator's guide to practitioner action research*. Thousand Oaks: Corwin Press.

Annie E. Casey Foundation. (2006). *Race matters: Unequal opportunities in education*. Retrieved from http://www.aecf.org/resources/race-matters-1/

Au, K. (2009). Isn't culturally responsive instruction just good teaching? *Social Education, 73* (4), 179–183.

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

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., & Zhang, J. (2013). *The Condition of Education 2013* (NCES 2013-037). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch.

Barnhardt, R., & Kawagley, O. (2005). Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology and Education Quarterly*, *36* (1), 8–23.

Barton, P., & Coley, R. (2009). *Parsing the achievement gap II*. Educational Testing Service Policy Evaluation and Research Center. Princeton, N. J.: Educational Testing Service.

Baum, S., Ma J., & Payea, K. (2013). *Education pays 2013: The benefits of higher education for individuals and society.* College Board. Retrieved from https://trends.collegeboard.org/education-pays

Boykin, A. W., & Bailey, C. (2000). The role of cultural factors in school relevant cognitive functioning: Synthesis of findings on cultural contexts, cultural orientations, and individual differences. (CRESPAR Technical Report No. 43). Baltimore, MD: Center for Research on the Education of Students Placed At Risk, Center for Social Organization of Schools, Johns Hopkins University.

Boykin, A.W., Coleman, S. T., Lilja, A. J., &Tyler, K. M. (2004). Building on children's cultural assets in simulated classroom performance environments: Research vistas in the communal learning paradigm. (CRESPAR Technical Report No. 68). Baltimore, MD:

Center for Research on the Education of Students Placed At Risk, Center for Social Organization of Schools, Johns Hopkins University.

Boykin, A. W., Tyler, K.M., Watkins-Lewis, K., & Kizzie, K. (2006). Culture in the sanctioned classroom practices of elementary school teachers serving low-income African American students. *Journal of Education for Students Placed at Risk*, *11*(2), 161–173.

Cajete, G. (2005). American Indian epistemologies. *New Directions for Student Services*. 109, 68–79.

Cajete, G. (2000). *Native science: Natural laws of interdependence*. Santa Fe, NM: Clear Light.

Cajete, G. (1999). The Native American learner and bicultural science education. In Swisher and Tippeconnic ,III., (Eds.), *Next Steps: Research and Practice to Advance Indian Education* (pp. 135-160). ERIC Clearinghouse on Rural Education and Small Schools: Charleston.

Calabrese-Barton, A., & Lee, O. (2006). National Association for Research in Science Teaching Equity and Ethics Committee: A call to action. *Journal of Research in Science Teaching, 43* (9), 875–878.

Cardell, G. W., Cross, W. C., & Lutz, W. J. (1978). Peer learning among Indian students: Extending counselor influence into the classroom. *Journal of American Indian Education, 17*(2), 7–12.

Carter, R.T. (2000). Reimagining race in education: A new paradigm from psychology. *Teachers College Record, 102*(5), 864–897.

Center for Research on Equity and Diversity in Education. *The CREDE five standards for effective pedagogy and learning*. Retrieved September 5, 2010 from http://crede.berkeley.edu/research/crede/standards.html

Chawla Sahota, P. (2010). *Community-based participatory research in American Indian and*

Alaska Native communities. National Congress of American Indians Policy Research Center. Retrieved from http://www.ncaiprc.org/files/CBPR%20Paper%20FINAL.pdf

Chicago Teachers Union. (2012). *Still separate still unequal*. Retrieved from www.ctunet.com/blog/text/SegregationFinal.pdf

Cohen, J. (1992*).* A power primer. *Psychological Bulletin*, 112 (1), 155-159.

Committee on Equal Opportunities in Science and Engineering. (2009). *Biennial report to Congress, 2007-2008*. National Science Foundation. Retrieved from http://www.nsf.gov/od/oia/activities/ceose/reports/2008CEOSE_BiennialReport.pdf

Conner, R.F. (2004) Developing and implementing culturally competent evaluation: A discussion of multicultural validity in two HIV prevention programs for Latinos. In M. Thompson-Robinson, R. Hopson, & S. SenGupta, (Eds.). *In Search of Cultural Competence in Evaluation. New Directions in Evaluation. 102*, 51–66.

Cullen, J. E. (2008). "Some friends and I started talking" – *A participatory action research project to deconstruct white privilege among student affairs practitioners* (Doctoral dissertation). Retrieved from Proquest. (194001822).

Darling-Hammond, L. (2000). *Teacher quality and student achievement: A review of state policy evidence.* Education Policy Analysis Archives, 8(1). Retrieved from http://depts.washington.edu/ctpmail/publications/

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics*. *88*(9-10), 1197–1720.

Del Basterra Rosario, M., Trubull, E., & Solano-Flores, G. (2011). *Cultural validity in assessment: Addressing linguistic and cultural diversity*. New York: Routledge.

Deloria, V., Jr., & Wildcat, D. R. (2001). *Power and place: Indian education in America.*

Golden, CO: Fulcrum Resources.

Demmert, W., & Towner, J. (2003). A review of the research literature on the influences of culturally based education on the academic performance of Native American students. Portland: Northwest Regional Education Laboratory.

Denzin, N., & Lincoln, Y. (2008). Critical methodologies and Indigenous inquiry. In N. Denzin, Y. Lincoln, and L. Tuhiwai-Smith (Eds.). *Handbook of Critical and Indigenous methodologies* (pp. 1–20). Thousand Oaks: Sage.

DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed.). Applied Social Research Methods Series, Vol. 26. Thousand Oaks: Sage.

Deyhle, D. (1995). Navajo youth and Anglo racism: Cultural integrity and resistance. *Harvard Educational Review*, 65, 403–444.

Education Trust. (2013). Breaking the glass ceiling of achievement for low income students and student of color. Retrieved from http://www.edtrust.org/high_end_gaps

Gay, G. (2010). *Culturally Responsive Teaching: Theory, Research, & Practice*. 2nd edition. New York: Teachers College Press.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106–116.

Gilbert, W. S. (2010). Developing culturally based science curriculum for Native American classrooms. In J. Reyner, W. S. Gilbert, & L. Lockard (Eds.). *Honoring our heritage: Culturally appropriate approaches for teaching Indigenous students*, (pp**.** 43–56). Flagstaff: Northern Arizona University.

Gilbert, W. S. (2005). *Native Science Connections Research Project: Integrating relevant cultural knowledge into the science curriculum for grades 4-6th*. Retrieved from http://www2.ed.gov/rschstat/research/pubs/oieresearch/conference/gilbert_200501.pdf

Gilliland, H. (1999). *Teaching the Native American*. 4th edition. Dubuque: Kendall/Hunt.

Gutierrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*, 19–25.

Hall, M., & Ward Hood, D. (2005). Persuasive language, responsive design: A framework for interculturally responsive evaluation. In S. Hood, R. Hopson, & H. Frierson, (Eds.) *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice* (pp. 59–84). Greenwich: Information Age Publishing.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238.

Herr, K., & Anderson, G. (2005). *The action research dissertation: A guide for students and faculty*. Thousand Oaks: Sage.

Hilberg, R.S., & Tharp, R.G. (2002). Theoretical perspectives, research findings, and classroom implication of the learning styles of American Indian and Alaska Native students. *ERIC Digest, EDO-RC-02-02,* September, 2002.

Hilberg, R.S., Tharp, R.G., & Degeest L. (2000). The efficacy of CREDE's standards-based instruction in American Indian mathematics classes. *Equity & Excellence in Education*, *33*(2), 32–40.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey

questionnaires. *Organizational Research Methods*, *1*(1), 104-121.

Hollins, E. (2008). *Culture in School Learning: Revealing the Deep Meaning*. Routledge. 2nd edition. New York: Routledge.

Hurley, E., Boykin, A.W., & Allen, B. (2005). Communal versus individual learning of a math-estimation task: African American children and the culture of learning contexts. *The Journal of Psychology,* 2005, *139*(6), 513–527.

Indian Self-Determination and Education Assistance Act of 1974. Pub. L. No. 93-638, Stat. 1017 (1975).

Institute of Education Sciences. (2008). *What Works Clearinghouse Procedures and Standards Handbook*. Washington. DC: United States Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1

Institute of Medicine and National Research Council. (2010). *Student Mobility: Exploring the Impact of Frequent Moves on Achievement: Summary of a Workshop***.** Washington, DC: The National Academies Press.

Kirkhart, K. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood, R. Hopson, & H. Frierson, (Eds.) *The role of culture and cultural context: A mandate for inclusion, the discovery of truth and understanding in evaluative theory and practice* (pp. 21–38). Greenwich: Information Age.

Klump, J., & Mcneir, G. (2005). Culturally responsive practices for student success: A regional sampler. Northwest Regional Education Laboratory. Retrieved from http://educationnorthwest.org/webfm_send/296

Kovach, M. (2009). *Indigenous methodologies: Characteristics, conversations, and contexts*. Toronto: University of Toronto Press.

Krathwohl, D. R. (1998). *Methods of educational and social science research.* (2nd ed.). New York, NY: Addison Wesley

LaFrance, J. (2004). Culturally competent evaluation in Indian Country. In M. Thompson-Robinson, R. Hopson, & S. SenGupta, (Eds.). *In Search of Cultural Competence in Evaluation. New Directions in Evaluation: 102*, 39–50.

LaFrance, J., Kirkhart, K. & Nichols, R. (in press). Cultural views of validity: A conversation. In *Continuing the Journey to Reposition Culture and Cultural Context in Evaluation Theory and Practice*. S. Hood, R. Hopson, H. T. Frierson, & K. Obeidat (Eds.). Charlotte: InfoAge Publishing Co.

LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation, 23*(2), 13–31.

LaFrance, J., Nichols, R., & Kirkhart, K. E. (2012). Culture writes the script: On the centrality of context in indigenous evaluation. In D. J. Rog, J. L. Fitzpatrick, & R. F. Conne, (Eds.). *Context: A framework for its influence on evaluation practice. New Directions for Evaluation, 135, (*59–74).

Lankford. H, Loeb, S., & Wyckoff, J. (2002). "Teacher Sorting and the Plight of Urban Schools". *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

LaVeaux, D., & Christopher, S. (2009). Contextualizing CBPR: Key principles of CBPR meet the Indigenous research context. *Pimatisiwin: A Journal of Aboriginal and Indigenous Community Health, 7*(1), 1–25.

Lee, O. (2002). Science inquiry for elementary students from diverse backgrounds. In W.G. Secada (Ed.), *Review of Research in education*, *26* (pp. 23–69).Washington, DC: American Educational Research Association.

Lee, O. (2003). Equity for linguistically and culturally diverse students in science education: A research agenda. *Teachers College Record*, *105*(3), 465–489.

Lee, O. (2005). Science education and student diversity: Synthesis and research agenda. *Journal of Education for Students Placed at Risk, 10*(4), 433–440.

Lee, O., & Buxton, C. (2010). *Diversity and Equity in Science Education*. Teachers College Press.

Lee, O., & Luykx, A. (2006). Science education and student diversity: Synthesis and research agenda. New York: Cambridge University Press.

Lee, O., Luykx, A., Buxton, C., & Shaver, A. (2007). The challenge of altering elementary school teachers' beliefs and practices regarding linguistic and cultural diversity in science instruction. *Journal of Research in Science Teaching*, *44*(9), 1269–1291.

Lipka, J., & Adams, B. (2007). Some evidence for ethnomathematics: Quantitative and qualitative data from Alaska. In Favilli (Ed.), *Ethnomathematics and Mathematics Education, Proceedings of the 10th International Congress of Mathematics Education.* Tipografia Editrice Pisana: Pisa.

Lipka, J., Parker Webster, J., & Yanez, E. (2005). Factors that affect Alaska Native students' mathematical performance. *Journal of American Indian Education, 44*(3), 1–8.

Lipka, J., Sharp, N., Adams, B., & Sharp, F. (2007). Creating a third space for authentic biculturalism: Examples from Math in a Cultural Context. *Journal of American Indian Education, 46*(3), 94–115.

Luyxk, A., & Lee, O. (2007). Measuring instructional congruence in elementary science classrooms: Pedagogical and methodological components of a theoretical framework. *Journal of Research in Science Teaching, 44*(3), 424–447.

Luykx, A., Lee, O., Mahotiere, M., Lester, B., Hart, J., & Deaktor, R. (2007). Cultural and home language influences on children's responses to science assessments. *Teachers College Record*, *109*(4), 897–926.

Mackety, D. M., & Linder-VanBerschot, J. A. (2008). *Examining American Indian perspectives in the Central Region on parent involvement in children's education* (Issues & Answers Report, REL 2008–No. 059). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. Retrieved from http://ies.ed.gov/ncee/edlabs

Matthews, C.E., & Smith, W.S. (1994). Native American related materials in elementary science instruction. *Journal of Research in Science Teaching, 31,* 363–38.

McGee–Banks, C., & Banks, J. (1995). Equity pedagogy: An essential part of multicultural education. *Theory into Practice*, *34* (3), 152–158.

Meriam, L, Brown, R., Cloud, H. R., Dale, E. E., Duke, E., Edwards, H. R., … Spillman, W. J. (1928). The problem of Indian administration. *The Institute for Government Research Studies in Administration*. Johns Hopkins Press: Baltimore. Retrieved from http://www.alaskool.org/native_ed/research_reports/IndianAdmin/Indian_Admin_Problms.html

Merriam – Webster. (2014). Retrieved July 12, 2014 from http://www.merriam-webster.com/

Messick, S. (1990). Validity of test interpretation and use. Princeton, N.J.: Educational Testing Service.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist 50, 741–749.

Milligan, K., Moretti, E., & Oreopoulos, P. (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics*, August, 1167–96.

Mohatt, G., & Erickson, F. (1981). Cultural differences in teaching styles in an Odawa school: A sociolinguistic approach. In H. Trueba, G. Guthrie, & K. Au (Eds.), *Culture and the Bilingual Classroom: Studies in Classroom Ethnography* (pp. 105–119). Rowley, MA: Newbury House.

Montana Office of Public Instruction. (2010). *Montana American Indian student achievement data report*. Montana Office of Public Instruction: Helena.

Moses-Snipes, P., & Snipes, V. (2005). The call: The importance of research on African American issues in mathematics and science education. *The Negro Educational Review*, *56* (2 & 3), 103–105.

Moss, P. A. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, *17*(2), 6–12.

Moss, P. A. (2005). Understanding the other/understanding ourselves: Toward a constructive dialogue about "principles in educational research. *Educational Theory*, *55*(3), 263-283.

Moss, P. A., Phillips, D. C., Erickson, F. D., Floden, R. E., Lather, P. A., & Schneider, B. L. (2009). Learning from our differences: A dialogue across perspectives on quality in education research. *Educational Researcher*, *38*(7), 501–517.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*(2), 159–168.

Native American Languages Act of 1990. P.L. 101-477. Retrieved from http://thomas.loc.gov/cgi-bin/bdquery/z?d101:h5040:

National Center for Education Statistics. (2013a) Digest of Education Statistics. Table 203.50- Enrollment and percentage distribution of enrollment in public elementary and secondary schools, by race/ethnicity and region: Selected years, fall 1995 through fall 2023. Institute of Education Sciences. U.S. Department of Education. Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_203.50.asp

National Center for Education Statistics. (2013b). Digest of Education Statistics. Table 209.10 – Number, highest degree and year of full-time experience of teacher in public and private elementary and secondary schools, by selected teacher characteristics: Selected years, 199-2000 through 2011-2012. Institute of Education Sciences. U.S. Department of Education. Retrieved from http://nces.ed.gov/programs/digest/d13/tables/dt13_209.20.asp

National Center for Education Statistics. (2013c). *The Nation's report card: Trends in academic progress 2012* (NCES 2013 456). Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013456

National Council of Teachers of Mathematics. (2004). *Final report: Achievement Gap Task Force*. Retrieved from https://my.nctm.org/uploadedFiles/About_NCTM/Board_and_Committess/achievement_gap.pdf

National Council of Teachers of Mathematics. (2008). *Equity in mathematics education position statement 2008*. Retrieved from http://www.nctm.org/about/content.aspx?id=13490

National Council of Teachers of Mathematics. (2014). *Access and equity in mathematics education: A position of the National Council of Teachers of Mathematics* Retrieved from http://www.nctm.org/about/content.aspx?id=42501

National Indian Education Association & Center for Indian Education. (2005). *Preliminary report on No Child Left Behind in Indian Country*. Washington, DC: National Indian Education Association. Retrieved from http://www.niea.org/policy/policy-resources.aspx

National Indian Education Association. (2007). *2007 Annual Report*. Retrieved from http://www.niea.org/profile/annualreport2007.pdf

National Science Teachers Association. (2000). *NSTA's position statement on multicultural science education*. Retrieved from http://nsta.org/about/positions/multicultural.aspx

Nelson–Barber, S., Lafrance, J., Trumbull, E., & Aburto, S. (2005). Promoting culturally reliable and valid evaluation practice. In S. Hood, R. Hopson & H. Frierson, (Eds.), *The Role of Culture and Cultural Context: A Mandate for Inclusion, the Discovery of Truth and Understanding in Evaluative Theory and Practice* (pp. 59-84).Greenwich: Information Age Publishing.

Nelson–Barber, S., & Trumbull, E. (2007). Making assessment practices valid for Indigenous American students. *Journal of American Indian Education*, *46*(3), 132–147.

No Child Left Behind Act of 2001. Pub. L. No. 107–110, § 115, Stat. 1425 (2002). Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Northwest Indian Applied Research Institute. (no date). *Traditional American Indian values and behaviors.* Evergreen College, OR. Retrieved May 15, 2014 from http://nwindian.evergreen.edu/curriculum/valuesbehaviors.html

Olson, L. (January 9, 2003). The great divide. *Quality counts 2003: If I can't learn from you*.

Education Week on the Web. Retrieved from http://www.edcounts.org/archive/sreports/qc03/templates/article.cfm@slug=17divide.h22.html

Parsons, E. C. (2008). Learning contexts, Black cultural ethos, and the science achievement of African Americans. *Journal of Research in Science Teaching 45*(6), 665–683.

Penfield, R. D., & Lee, O. (2010). Test-based accountability: Potential benefits and pitfalls of science assessment with student diversity. *Journal of Research in Science Teaching*, 47(1), 6–24.

Pewewardy, C., & Hammer, P. (2003). Culturally responsive teaching for American Indian students. *Eric Clearinghouse on Rural Education and Small Schools*. Retrieved from http://www.ericdigests.org/2005-1/teaching.htm

Phuntsog, N. (2001). Culturally responsive teaching: What do selected United States elementary school teachers think? *Intercultural Education, 12* (1), 51–64.

Ponterotto, J. G., Baluch, S., Greig, T., & Rivera, L. (1998). Development and initial score validation of the teacher multicultural altitude survey. *Educational and Psychological Measurement*, *58*(6), 1002–1016.

Quigley, D. (2001). *Compilation on environmental health: Research ethics issues with Native communities*. Collaborative Initiative for Research Ethics in Environmental Health. Syracuse University.

Rampey, B.D., Dion, G.S., & Donahue, P.L. (2009). *NAEP 2008 trends in academic progress* (NCES 2009–479). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision.

Psychological Assessment, *12*(3), 287.

Sable, J., & Plotts, C. (2010). *Public elementary and secondary school student enrollment and staff counts from the Common Core of Data: School year 2008–09* (NCES 2010-347). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Science Education Resource Center. (2014). Retrieved from http://serc.carleton.edu/NAGTWorkshops/affective/efficacy.html

Schmidt, W., Cogan, L., & McKnight, C. (2010). Equality of educational opportunity: Myth or reality in U.S. schooling? *American Educator, Winter 2010-2011*, 12–19.

Singh, N.K. (2011). Culturally appropriate education: Practical and theoretical implications. In Reyhner, Saskiestewa and Lockard, (Eds.), *Honoring our heritage: Culturally appropriate approaches for teaching Indigenous students* (pp. 11–42). Northern Arizona University.

Siwatu, K. O. (2007). Preservice teachers' culturally responsive teaching self-efficacy and outcome expectancy beliefs. *Teaching and Teacher Education, 23*, 1086–1101.

Skinner, L. (1999). Teaching through traditions: Incorporating languages and culture into curricula. In K. Swisher and J. Tippeconnic, III., (Eds.). *Next steps: Research and practice to advance Indian education* (pp. 107-134*)*. Charleston: ERIC Clearinghouse on Rural Education and Small Schools.

Sleeter, C. (2001). Epistemological diversity in research on preservice teacher preparation for historically underserved children. *Review of Research in Education*, *25*, 209–250.

Solanos-Flores, G. & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching, 38*(5), 553–573.

Sternberg, R., Lipka, J., Newman, T., Wildfeuer, S., & Grigorenko, E. L. (2006). Triarchically-based instruction and assessment of sixth-grade mathematics in a Yup'ik cultural setting in Alaska. *Gifted and Talented International 21(2),* 6–19.

Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 414.

Tribal Education Departments National Assembly. (2011). *Tribal education departments report.* Retrieved May 11, 2013 from http://www.niea.org/data/files/tedna teds report.pdf

Tyler, K., Uqdah, A., Dillihunt, M., Beatty-Hazelbaker, R., Conner, T., Gadson, N., … Stevens, R. (2008). Cultural discontinuity: Toward a quantitative investigation of a major hypothesis in education. *Educational Researcher*, *37* (5), 280–297.

United States Census Bureau. (2012). *2010 census summary file 1 - technical documentation*. Retrieved from http://www.census.gov/prod/cen2010/doc/sf1.pdf

United States Census Bureau. (2012). *2012 National population projections*. Retrieved from http://www.census.gov/population/projections/data/national/2012.html

United States Census Bureau. (2010). Race, Hispanic or Latino, age, and housing occupancy: 2010, Table QT-PL. 2010 Census National Summary File of Redistricting Data. Retrieved from http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?fpt=table

United State Department of Education. (2009). *Investing in Innovation Fund fact sheet*. Retrieved from http://www2.ed.gov/programs/innovation/factsheet.html

United States Department of Education. (2009). *Race to the Top application*. Retrieved from http://www.ed.gov/programs/racetothetop/application.doc

United State Department of Education. (2010). *Tribal leaders speak: The state of American*

Indian education 2010. Retrieved from http://www2.ed.gov/about/inits/ed/indianed/consultations.report.pdf

United States General Accountability Office. (2010). *K-12 education: Many challenges arise in educating students who change school frequently.* GAO-11-40. Retrieved from http://www.gao.gov/products/GAO-11-40

United States Office of Personnel Management. (2014). *Assessment decision guide*. Retrieved from http://apps.opm.gov/adt/content.aspx?page=5_Glossary&AspxAutoDetectCookieSupport=1&JScript=1

Ute Tribe. (2004). *Ute Tribal Education Department plan*. Retrieved September 27, 2010 from http://www.uiteducation.com/index.aspx?c=12

Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. *Psychological Assessment*, *16*(3), 231.

Web Center for Social Research Methods. (2006). Research methods knowledge base. Retrieved June 21, 2014 from http://www.socialresearchmethods.net/kb/measval.php

Wenglinsky, H. (2000). How teaching matters: Bringing the classroom back into discussions of teacher quality. Princeton, NJ: Educational Testing Service.

Wilson, S. (2008). *Research is ceremony: Indigenous research methods*. Halifax: Fernwood Publishing.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838.

Yazzie, T. (1999). Culturally appropriate curriculum: a research based rationale. In K. Swisher

and J. Tippeconnic, III., (Eds.). *Next steps: Research and practice to advance Indian* education (pp. 83-106). Charleston: ERIC Clearinghouse on Rural Education and Small Schools.

Yankton Sioux Tribe. (1995). *Yankton Sioux Tribal Code*. Retrieved September 26, 2010 from http://www.narf.org/nill/Codes/yanktoncode/yanktoncodet15education.htm

York–Barr, J., & Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research, 74*(3), 255–316.

Appendix A

Notes on CCI Survey Ideas Compiled from a Literature Review - December 2005

Cultural competency

Examples of cc content

Examples of cc pedagogy

Curriculum resources

Reservation/cultural fieldtrips

Tribal members worked with teacher/students

Classroom displays

Books, other classroom resources available to students

LEP based instruction

Native language use

Attendance at cultural events

High expectations for all

Alternative assessment

Technology

Formative assessment

Collaborative

Wait time

Private practice/low risk

Rapport with students

Rapport with family

Multiple teachers

Visual

Concrete

Multi modal

Student centered

Practical

Constructivist

Flexible time

Appendix B

First Prototype CCIS

Please indicate how often you included each of the following items in your math/science instruction during the last school year.

1) Traditional stories from local Tribes

Never O Rarely O Sometimes O Often O

2) Content about contemporary local Tribal issues

Never O Rarely O Sometimes O Often O

3) Historical content about local American Indian Tribes

Never O Rarely O Sometimes O Often O

4) A fieldtrip to a cultural site significant to local American Indian Tribes

Never O Rarely O Sometimes O Often O

Please indicate how often you had each of the following items accessible to students in your classroom during the last school year.

5) Age appropriate books about local Tribal cultures

Never O Rarely O Sometimes O Often O

6) Bulletin boards/displays that include cultural content

Never O Rarely O Sometimes O Often O

7) Salish or Kootenai words or phrases posted

Never O Rarely O Sometimes O Often O

8) American Indian music

Never O Rarely O Sometimes O Often O

9) Salish, Pend d' Oreille or Kootenai crafts or art work

Never O Rarely O Sometimes O Often O

10) Pictures or videos that address local tribal cultures

Never O Rarely O Sometimes O Often O

Please indicate how frequently you used each of the following strategies or items in your science instruction during the last school year.

11) Cooperative learning groups

Never O Rarely O Sometimes O Often O

12) Formative assessment with feedback to students

Never O Rarely O Sometimes O Often O

13) A variety of instructional methods that address diverse learning styles

Never O Rarely O Sometimes O Often O

14) Student use of instructional technology

Never O Rarely O Sometimes O Often O

15) Strategies that address Limited English Proficiency

Never O Rarely O Sometimes O Often O

16) Alternative assessment (other than paper and pencil tests and quizzes)

Never O Rarely O Sometimes O Often O

17) Salish or Kootenai language

Never O Rarely O Sometimes O Often O

18) Private teacher-student discussion of student learning

Never O Rarely O Sometimes O Often O

19) Performance based assessment

Never O Rarely O Sometimes O Often O

20) Examination of content for cultural bias

Never O Rarely O Sometimes O Often O

21) Extended wait time

Never O Rarely O Sometimes O Often O

22) Interaction with every student's parents or guardians

Never O Rarely O Sometimes O Often O

23) Working with Tribal elders or other community member as guest teachers

Never O Rarely O Sometimes O Often O

24) Content relevant to Salish, Kootenai, or Pend d'Oreille cultures

Never O Rarely O Sometimes O Often O

25) Examination of instructional methods for cultural bias

Never O Rarely O Sometimes O Often O

26) Visit by a Tribal member to your class to share cultural information

Never O Rarely O Sometimes O Often O

27) Contact a Tribal member to obtain culture related information or resources

Never O Rarely O Sometimes O Often O

Appendix C

Second Prototype CCIS

Please indicate how often you included each of the following items in your science instruction during the 2008-'09 school year.

1) Traditional stories from local Tribes

Never O Rarely O Sometimes O Often O

2) Content about contemporary local Tribal issues

Never O Rarely O Sometimes O Often O

3) Historical content about local American Indian Tribes

Never O Rarely O Sometimes O Often O

4) A fieldtrip to a cultural site significant to local American Indian Tribes

Never O Rarely O Sometimes O Often O

5) Visit by a Tribal member to your class to share cultural information

Never O Rarely O Sometimes O Often O

6) Contact a Tribal member to obtain culture related information or resources

Never O Rarely O Sometimes O Often O

Please indicate how often you had each of the following items accessible to students in your classroom during the 2006-'07 school year.

7) Age appropriate books about local Tribal cultures

Never O Rarely O Sometimes O Often O

8) Bulletin boards/displays that include cultural content

Never O Rarely O Sometimes O Often O

9) Posted words or phrases in local Native languages

Never O Rarely O Sometimes O Often O

10) American Indian music

Never O Rarely O Sometimes O Often O

11) Locally made American Indian crafts or art work

Never O Rarely O Sometimes O Often O

12) Pictures or videos that reflect local Tribal cultures

Never O Rarely O Sometimes O Often O

13) Other (Please specify.) _____

Please indicate how frequently you used each of the following strategies or items in your science instruction during the 2008-'09 school year.

14) Collaborative learning groups

Never O Rarely O Sometimes O Often O

15) Strategies chosen to address diverse learning styles

Never O Rarely O Sometimes O Often O

16) Strategies that assist learners who are Limited English Proficient (e. g., frequent use of graphics, models, other visuals; moving from concrete to abstract; contextualized use of vocabulary)

Never O Rarely O Sometimes O Often O

17) Alternative assessment

Never O Rarely O Sometimes O Often O

18) Local Native language

Never O Rarely O Sometimes O Often O

19) Formative assessment with direct feedback to students

Never O Rarely O Sometimes O Often O

20) Private one on one teacher-student discussion of student learning

Never O Rarely O Sometimes O Often O

21) Examination of content for cultural bias

Never O Rarely O Sometimes O Often O

22) Extended wait time

Never O Rarely O Sometimes O Often O

23) Interaction with every student's parents or guardians

Never O Rarely O Sometimes O Often O

24) Working with Tribal elders or other community member as guest teachers

Never O Rarely O Sometimes O Often O

25) Mentoring of students by adults other than the classroom teacher

Never O Rarely O Sometimes O Often O

26) Opportunities for private practice precede public demonstration of proficiency

Never O Rarely O Sometimes O Often O

27) Practical application of science knowledge by students in classroom activities

Never O Rarely O Sometimes O Often O

28) Art based instructional methods (e.g., metaphors, storytelling, music, etc.)

Never O Rarely O Sometimes O Often O

29) Examination of instructional methods for cultural bias

Never O Rarely O Sometimes O Often O

30) Examination of instructional content for cultural bias

Never O Rarely O Sometimes O Often O

31) Teaching core science content using a local or place based context

Never O Rarely O Sometimes O Often O

32) Teaching American Indian traditional science knowledge along with Western science content

Never O Rarely O Sometimes O Often O

33) Open ended problem based learning

Never O Rarely O Sometimes O Often O

34) Observational learning strategies (e.g., adult or peer modeling, demonstrations, apprenticeships)

Never O Rarely O Sometimes O Often O

35) Local Native language in instruction and interactions with students

Never O Rarely O Sometimes O Often O

Appendix D

Cohort 1 Treatment Teachers Focus Group Question Guide

1. How would you describe the attributes of a "culturally competent" teacher? Whenever possible describe examples of how these attributes translate into classroom practice. (flip chart attributes – practices)

   Look for the following:

   - Knowledge of another culture
   - Awareness of own cultural lens
   - Affectations attitudes –such as “kids can learn”

2. What does a culturally competent science classroom and/or lessons look like? Can you give some examples?

3. What is the most challenging or difficult aspect of cultural competent teaching? How can we deal with these challenges?

4. Has the PD project influenced your understanding of attributes/practices that demonstrate cultural competency?

   A) How did it change your understanding?
   B) What are you doing in your teaching practice now that you did not do before joining the project?

5. This is a list generated by a group of Native American teachers and cultural experts who are part of the project. How does it compare with your list? Probe for thoughts on similarities or differences.

6. The PD project involved many components (summer on-site workshops, culture camps, school year meetings, online instruction, teacher lesson study, working with higher education faculty and cultural experts, etc.) What has been the more significant elements for you and explain your reasons for choosing these?

   Follow up – Project PD elements include: science content, pedagogical skills/knowledge, cultural instruction, leadership development .Write these on flip chart and mark each one that is chosen by a member of the group – can have two votes

7. What elements are most important to include in a model for professional development designed to build cultural competency?

Appendix E

Revised CCIS

Section 1: Curriculum Content

Think about your **science instruction** during the 2008-2009 school year. Circle the number in the column that best represents **the percentage of science lessons** in which you used each of the following types of content in your during the 2008-2009 school year.

e.g., For the first item, if you included a traditional story in 25% of the science lessons you taught during 2008-2009 school year, you would circle the number 3 in the "Sometimes" column.

	Never	Seldom (1 to 20%)	Sometimes (21 to 40%)	Often (41 to 60%)	Very Often (61 to 80%)	Almost Always (>80%)
1) A traditional story from a Montana Indian tribe	1	2	3	4	5	6
2) Contemporary issues relevant to Montana Indian tribes	1	2	3	4	5	6
3) Historical content about Montana Indian tribes	1	2	3	4	5	6
4) A fieldtrip to a site significant to Montana Indian tribes	1	2	3	4	5	6
5) Traditional science knowledge from Montana Indian tribes	1	2	3	4	5	6
6) Science content tied to a place based context relevant to a Montana Indian tribe	1	2	3	4	5	6
7) Other cultural content – Please specify here:	1	2	3	4	5	6

Section 2: Instructional Strategies

Think about your **science instruction** during the 2008-2009 school year. Circle the number in the column that best represents the **percentage of science lessons** in which you used each of the following instructional strategies during the 2008-2009 school year.

	Never	Seldom (1 to 20%)	Sometimes (21 to 40%)	Often (41 to 60%)	Very Often (61 to 80%)	Almost Always (>80%)
8) Had students work in collaborative groups	1	2	3	4	5	6
9) Used extended wait time in conversations with students	1	2	3	4	5	6
10) Encouraged students to assume responsibility for their learning - e.g., students made choices about how they studied a topic, how they were assessed, etc.	1	2	3	4	5	6
11) Local tribal elders or other tribal community members were guest teachers	1	2	3	4	5	6
12) Used teaching strategies that support Limited English Proficient or Second Language learners (e. g., used graphics, models, other visuals; moved from concrete to abstract; made frequent contextualized use of vocabulary)	1	2	3	4	5	6
13) Used alternative forms of assessment like authentic assessment, or performance based assessment (instead of multiple choice, fill in the blank, e.g.)	1	2	3	4	5	6
14) Provided specific formative feedback to each student	1	2	3	4	5	6
15) Used metaphors, analogies, or symbols to represent science content	1	2	3	4	5	6

Section 2: Instructional Strategies - Continued

	Never	Seldom (1 to 20%)	Sometimes (21 to 40%)	Often (41 to 60%)	Very Often (61 to 80%)	Almost Always (>80%)
16) Used local Native language in instructional interactions with students	1	2	3	4	5	6
17) Provided ample opportunity for students to engage in private practice before publicly demonstrating their proficiency	1	2	3	4	5	6
18) Used science activities in which students designed solutions to problems relevant to the their community	1	2	3	4	5	6
19) Supported mentoring of students by adults other than the classroom teacher or paraprofessionals	1	2	3	4	5	6
20) Used art based teaching methods (e.g., storytelling, music, drawing, painting, poetry, drama, etc.)	1	2	3	4	5	6
21) Used observational learning strategies (e.g., adult or peer modeling, demonstrations, apprenticeships)	1	2	3	4	5	6
22) Was flexible with time (e.g., changed scheduling of instruction to meet individual students' needs)	1	2	3	4	5	6
23) Other instructional strategies for teaching ethnically diverse students– Please specify here:	1	2	3	4	5	6

Section 3: Classroom Resources Accessibility

Think about your **classroom environment** during the 2008-2009 school year. Circle the number in the column that best represents **the percentage of school days** that each of the following resources were **accessible** to students in your classroom during the 2008-2009 school year.

	Never	Seldom (1 to 20%)	Sometimes (21 to 40%)	Often (41 to 60%)	Very Often (61 to 80%)	Almost Always (>80%)
24) Age appropriate books about Montana tribal cultures	1	2	3	4	5	6
25) Bulletin boards or displays that included content from Montana Indian cultures	1	2	3	4	5	6
26) Posted words or phrases written in local Native languages	1	2	3	4	5	6
27) Music from Montana Indian tribes	1	2	3	4	5	6
28) Tools, crafts or art work made by members of Montana Indian tribes	1	2	3	4	5	6
29) Pictures or videos of Montana Indian cultures	1	2	3	4	5	6
30) Games and toys from Montana Indian cultures	1	2	3	4	5	6
31) Web sites or software about Montana Indian cultures	1	2	3	4	5	6
32) Other culturally relevant resources – Please specify here:	1	2	3	4	5	6

Section 4: Additional Education Related Practices

This last section of the survey asks you to again think about the 2008-2009 school year. Circle the number in the column that best represents **how many times during the school year** you engaged in each of the practices listed below.

	Never	1 to 2 times per year	3 to 4 times per year	5 to 6 times per year	7 to 8 times per year	9 or more times per year
33) Communicated with every student's parents or guardians to discuss their student's progress	1	2	3	4	5	6
34) Held a private conference with each student to discuss their progress	1	2	3	4	5	6
35) Examined instructional methods for cultural bias	1	2	3	4	5	6
36) Examined instructional content for cultural bias	1	2	3	4	5	6
37) Examined your classroom environment and management for cultural compatibility with your American Indian students	1	2	3	4	5	6
38) Consulted with tribal elders, culture committees, or other tribal community members about content relevant to Montana Indian tribes	1	2	3	4	5	6
39) Consulted with tribal elders or other tribal community members about classroom management or instructional strategies	1	2	3	4	5	6
40) Examined your science curriculum to see how well it addresses the "Essential Understandings About Montana Indians"	1	2	3	4	5	6

41) Other education related practices that address ethnic diversity – Please specify here:	1	2	3	4	5	6