*Journal of the Scholarship of Teaching and Learning*, Vol. 12, No. 3, September 2012, pp. 142 – 164.

# Looking deeper than the gradebook: Assessing cultural diversity attitudes among undergraduates

**Robert Lake and Kent Rittschof**[1]

*Abstract: Identification of college students' attitudes about diversity issues is an important part of the assessment of student development across many fields of study. This article discusses an action research approach and classroom application strategies stemming from a survey of 88 pre-service teacher candidates on their attitudes toward homosexuality, race, social class, and women's equality, following a university course on diversity. Data were examined using a Rasch model approach to determine and compare linear attitude measures and resulting change from pretest to posttest. Findings included endorsement difficulty changes among diversity issues as well as changes in student attitudes, primarily though not exclusively in concert with the course curriculum. Teaching approaches including the use of personal narratives were considered relative to findings. Areas for enhancement of instructional strategies were identified. Implications for teacher-scholars on examining linear measures of student attitudes and improving instruction on diversity issues in higher education were discussed.*

*Keywords: diversity, attitude change, Rasch model, personal narratives*

## I. Attitudes on Diversity Among University Students.

Educators around the world are faced with numerous challenges associated with effectively promoting the learning engagement of a diverse population of students. As part of many university programs in the United States attempting to help future educators understand cultural diversity, one or more undergraduate courses specifically dealing with diversity issues are typically required. Goals associated with such diversity courses usually include greater understanding of the many types of student diversity, and the many implications of cultural and other differences within educational settings. Toward addressing these goals, awareness of commonly held misunderstandings about diversity is often central to the curriculum.

While teacher education programs usually examine whether teacher candidates have sufficiently learned the required material from diversity courses, these programs often overlook whether the candidate's attitudes have changed in ways that coincide with the research-based information provided in the course and the general awareness expected of many school teachers. This article describes a study that begins to address this neglected aspect of teacher candidate assessment within the context of a University course dealing with multiculturalism. Furthermore, considering that teacher candidates study across college or university settings in the physical sciences, social sciences, or humanities, this examination of teacher candidates has relevance to many fields of study. That is, attitudes of future educators are influenced by the context of

[1] Department of Curriculum, Foundations, & Reading, Georgia Southern University, Box 8144 Statesboro, GA 30460, boblake@georgiasouthern.edu, kent_r@georgiasouthern.edu

pluralistic and diverse ways of seeing, knowing, and being in the world as they study across different domains.

*A. Theoretical Perspective.*

This inquiry is framed in the literature of multicultural education in America. A broad survey of the major works in this field reveals several evolutionary periods of this discipline (Banks, 1993). The first period emerged during the Jim Crow period through the work of Dubois (1902) and Woodson (1933) who were instrumental in initiating the study of Black history. The next major period of research arose out of the civil rights movement with the Coleman Report (1966) with whom the terms "social capital" and "white flight" had their origin. The civil rights movement was a catalyst for the emergence of ethnic studies (Banks, 1973) as well as the women's rights, gay and lesbian rights, Chicano rights, and disability rights movements to name a few. The term multicultural education began to be used in the last half of the 1970's (Grant, 1977; Hilliard, 1974; Klassen & Gollnick, 1977). During the culture wars of the 1980's, multicultural education was aligned with minority language advocacy, (Nieto, 1986) and brought direct challenges to issues of race, class, and gender in schools (Sleeter & Grant, 2009). By the close of the 20th century, multicultural education was fully immersed in the literature of social justice, critical educational practices, and equality issues for all minorities.

The conceptual framework of the institution's education college where the present study took place informs the choice of the literature and theoretical perspective of the curriculum under investigation. The following excerpt from the conceptual framework appears on the syllabus of the course of central interest:

> We believe that Reflective Educators for Diverse Learners, as the theme for our conceptual framework, considers all learners and represents a vision of professional practice for undergraduate students, graduate students, and faculty, joining together to form a community of learners. Therefore, we believe that all educators, at all levels, must acknowledge the multifaceted nature of their work and engage in an informed pedagogy that both recognizes and celebrates the diversities of contemporary life. *Reflective Educators for Diverse Learners* is the framework that permeates various orientations to the foundation of education, students' reflections upon their educational experience, observations of teachers in practicum, and the portraiture of schools (Georgia Southern University College of Education, 2006).

In addition to gaining further perspective from some of the key literature in the field of multicultural education by leading theorists such Banks (2006), Sleeter and Grant (2009), and Neito and Bode (2008), the instructor also adapted an "additive approach" (Valenzuela, 1999). This method draws on each student's "funds of knowledge" (González, Moll, & Amanti, 2005) as a starting point and lens for recognizing, affirming and valuing the cultural wealth (Martin, 2002) and diversity of any and all of those who are "other." This includes ethnicity, religious or non-religious worldviews, socio-economic status, and sexual orientation. For example, one of the first assignments given is a personal culture presentation that includes personal samples of home language use including maxims, metaphors and accents, celebrations, music and food. This assignment serves as a "home base" for exploring and affirming diversity.

The theoretical framework and course content was further informed by Maxine Greene's (1995) notion of social imagination through the addition of personal and humanizing narratives

of those that are considered "other." For example, stories of the lives of Muslim teachers, children raised by gay couples, stories of racial violence and hate crimes, historical narratives, and movies based on true stories were all continually woven into the course content. Greene (1995) says that, "I have learned the value of connective details. Without them, it is extraordinarily difficult to overcome abstraction in dealing with other people. A fearful oversimplification takes over in the blankness" (p. 95). Thus, teacher candidates were encouraged to share and reflect upon personal narratives as a means of developing those connective details. Examples from students' work that that follow were provided with their permission.

In one example a student asked if he could share his personal situation with the class. He opened with "I have three dads." Needless to say he had everyone's attention right away. He explained that he lives with his father and his father's gay partner and that he really respects the gay partner because of the interest the partner takes in his life and how well the partner communicates with him when compared with his biological father. He also stays with his mother and stepfather quite often. One reason this story had impact within the class context was that the student was very well liked by all in the class and had outstanding leadership qualities. He concluded with:

> I have to admit that now that I am at the completion of this class, I now have a more open mind to others around me. I noticed that just because the girl sitting next to me is the same race as me doesn't mean she feels the same way on certain topics. For instance one day our class was discussing our relationships with our parents. The girl next to me seemed to be a mama's girl that likes to be the girly girl. It turned out that she plays rugby and talks to her dad on a daily basis (S. Mincey, personal communication, December 12, 2010).

Our framework for understanding was also influenced by several decades of inquiry on preservice teachers' beliefs on diversity and the major themes within the body of research. Castro (2010), for example tracked themes in research from the middle 1980s through the late 2000s and converged on the following four themes in research as most current: a) a lack in understanding multicultural issues, b) contradictory attitudes/perceptions concerning diverse populations and social justice, c) importance of personal background on attitudes, beliefs and multicultural concepts, and d) instructional practices that foster changes in preservice teachers' beliefs about diversity, social justice, or multicultural education. We find that each of these research themes is relevant to our current investigation. We are particularly interested in major factors that influence change in college students' attitudes. Garmon's (2004, 2005) well considered research into change factors led to identification of *disposition* related influences and *experience* related influences on attitude change. Disposition factors include a) openness, b) self-awareness / self-reflectiveness, and c) commitment to social justice. Experience factors include a) intercultural experiences, b) educational experiences, and c) support group experiences. Each of these factors are also relevant to our current investigation though the educational experiences factor is most directly examined in the present inquiry.

*B. Action Research Questions.*

The overarching question of interest was whether students who successfully completed an undergraduate diversity course changed their attitudes about diversity issues addressed in the course. The issues investigated included two gender categories of *homosexuality* and *women's*

*equality,* as well as the two broad categories of *race* and *social class*. We were interested in whether students would demonstrate change in their reported diversity attitudes across four issue categories that coincided with the information provided in the course. Specifically, we sought to determine whether changes in attitude would tend toward agreement with statements consistent with course curriculum and toward disagreement with statements that were inconsistent with course curriculum.

A secondary but crucial question of interest was whether a pre-experimental single-group approach to examining change with a widely used instrument would generate valid, meaningful analysis on attitudes through repeated measures change data (Dimitrov & Rumrill, 2003; Morris & DeShon, 2002), despite the limitation of no control group for comparison. That is, by incorporating a contemporary Item Response Theory (IRT) analytic approach to investigate change we examined whether a single-group pre/post study design would yield measures that could inform teaching and continuous improvement of instruction. The importance of utilizing a single-group design was to investigate using an action research approach that can be frequently and easily repeated by an instructor, and that avoids the logistical constraints of control-group experimental research designs such as the exclusion of participants from treatment or alternately, the locating of non-treatment participants each semester. By using this single-group pre-experimental design it should be emphasized that generalizing beyond the particular group examined would not be supported. Although a lack of control group also prevents specific comparisons of whether any change would occur in the absence of treatment, the steps taken to support honest responding and our use of an IRT Rasch (1960/1980) scale approach with effect size adjustments were implemented to maximize meaningful data reflecting change. The design was also based on the assumption stemming from Garmon's work (2004, 2005) and our practical experiences that significant changes in diversity attitudes are not likely to occur easily by themselves. As Gay, Mills, and Airasian (2006) have noted relative to pretest-posttest design decisions, "Certain prejudices, for example, are not likely to change unless a concerted effort is made" (p. 253). With these ideas in mind we designed this research process in support of a valid, sustainable approach that assists with continuous improvement of instructional effort.

In particular, by taking advantage of a set of Rasch methods for constructing scaled linear measures and determining change in attitudes we were able to examine our research question beyond the broad issue of whether expected change takes place. We sought to examine and compare the relative degree to which change in diversity attitudes may occur both overall and for a variety of subtopics from within the four general categories of homosexuality, women's equality, race, and class. By specifying and examining these measures of change we sought to identify instructional areas that may need to be addressed differently. This measurement approach, in comparison to traditional investigations using raw scores and percentages, allows us to simultaneously take into account both the level of attitudes and the relative difficulty participants have with endorsing particular attitudes. Thus we are able to improve upon levels of certainty about the meaning behind the attitude assessment results and better examine potential sources of change, bias, and inconsistent responding that can affect the measurement of attitudes (Curtis, 2004). Specifically, Rasch measurement includes several diagnostic indices that permit item-specific and person-specific scrutiny on data reliability and measurement fit, for example. In addition, any measured changes in attitudes were assumed to reflect influences that include instruction, but that may not be limited to instruction, such as maturation or other courses.

Furthermore, we recognize the potential complexity of human attitudes, particularly with respect to controversial topics. For instance, an attitude toward an issue may include both an

automatic, "implicit" attitude as well as a different "explicit" attitude (Wilson, Lindsey, & Schooler, 2000). Although most attitude measures do not differentiate between such implicit and explicit attitudes we maintain that studying student attitudes while remaining aware of complexities and possible limitations can nonetheless provide valuable insights that support the development of effective teaching.

## II. Method.

### *A. Context and Design.*

The study took place during the Fall 2009/ Spring 2010 semesters within a pre-service education course called *Exploring Socio-Cultural Perspectives on Diversity in Educational Context*. The institution where the course was delivered was a university located in the heart of the rural American South, a region that is generally characterized by traditionally conservative values when compared to metropolitan areas. The course was a live, face-to-face semester requirement that met twice weekly for 90 minutes. Participants were simultaneously enrolled in two other required pre-service education courses including *Exploring Teaching and Learning* and *Investigating Critical and Contemporary Issues in Education*.

A pre-experimental single group pretest-posttest design was used. The first author conducted a pre and post semester survey of students' attitudes toward diversity. The survey data was used to construct attitude measures then calculate measured attitude change. Attitude measures reflected the relative agreement with statements that were consistent with research on diversity as well as statements that were inconsistent with research on diversity.

### *B. Participants.*

Undergraduate junior level college students (N = 88) who were enrolled in and successfully passed a three-credit course on diversity issues at a medium sized university in the Southeastern United States participated as part of a class activity. All students were enrolled in the course as prerequisite to entering one of several teacher certification programs at the institution. This course is part of a block of three courses that are taken in conjunction with a fifty-hour field placement component. There were two sections of students in the fall of 2009 and one section in the spring of 2010. An aggregate of the gender and ethnicity of the student participants was as follows: There were 69 white females, 14 African-American females, two African-American males, and thirteen white males.

### *C. Instruments.*

The Human Relations Attitude Inventory (HRAT; Koppelman & Goodhart, 2005) assesses cultural attitudes. It consists of a 64 item survey on the topics of *homosexuality, race, social class,* and *women's equality* (see Table 3 for specific sub-topics). Each item consisted of a statement followed by a five point Likert scale that included the levels of "strongly agree," "agree," "uncertain," "disagree," "strongly disagree." The instrument is intended for assessment of students who studied the Koppelman and Goodhart (2005) textbook, *Understanding human differences: Multicultural education for a diverse America,* used in the course. An example item statement from the instrument that was inconsistent with course curricula is as follows:

"Minorities do not achieve as much in our society because they do not aspire to achieve as much as white people do." An example item statement that was consistent with course curricula is as follows: "Racial segregation in our schools and neighborhoods remains a problem." We found no prior validity studies on the HRAT within searches of international databases of the literature. Therefore, both the qualitative and quantitative analyses of validity and reliability within this investigation were crucial to our interpretations. The complete HRAT instrument is available online at http://wps.ablongman.com/ab_koppelman_humandiff_2/77/19966/5111404.cw/content/index.html.

A content validity evaluation of the instrument was first conducted by the two authors of this article in order to identify items from the instrument that were most relevant to the course content pertinent to the investigation. In addition, the two authors screened for items that tended to express overgeneralizations of course information and yield data from items that would not address the research question well. Mutual agreement was used to identify 46 of the 64 items that possessed appropriate levels of content validity for this investigation. These 46 items consisted of 31 items that were statements reflecting *course inconsistent* views of diversity. The remaining 15 items were statements reflecting *course consistent* views of diversity.

In addition, three items from the standard end of course student evaluation instrument were aggregated to provide additional context to the findings. These three items addressed the amount of perceived learning and the change in interest level following the course.

### *D. Procedure.*

The researchers administered the entire HRAT within classrooms during the first week of class and again during the last week of class. Students were given one hour to complete the paper and pencil inventory. Procedures were designed to encourage and support the same level of honest responding on both the pretest and posttest, and minimize responding based primarily on social desirability biases. Students read the following instructions: "For each statement, select the response most representative of your own thinking and select the space corresponding to that response. Make each response a separate and independent one. Please respond to all statements. Respond as honestly as possible and work through the inventory as quickly as possible. Do not include your name when submitting the form. Note: Reference to "minority" or "minorities" in this inventory is to racial minorities in the U.S. (i.e., African Americans, Hispanic Americans, Asian/Pacific Island Americans, and American Indians) and does not include white ethnic groups and/or religious minorities." Students were also told that their responses would have no bearing on their grade in the course.

## III. Results.

### *A. Measurement Properties.*

The 31 items reflecting *course inconsistent* perspectives were analyzed separately from the 15 items reflecting *course consistent* perspectives in order to examine whether students responded reliably on the issues regardless of the framing of the questions. In addition, we did not assume that ratings of "strongly disagree" on any *course inconsistent* statements necessarily equated to ratings of "strongly agree" on any *course consistent* statements, for example. Measurement

validity characteristics were examined using an IRT Rasch (1960/1980) model approach. Rasch analysis is a contemporary latent trait approach that allows an examination of both the items and the students on a common interval measurement scale to gauge the comparative differences among both students and the items (Smith & Smith, 2006). Rasch measures of attitudes take into account the levels of difficulty in endorsing each survey item statement, unlike traditionally used counts and averages of ratings. Using the Rasch model, ordinal raw score ratings from the Likert instrument were converted to an interval scale of logistic units, or logits, that are scaled with the mean at zero. Fit analyses were then conducted using the *infit* and *outfit* procedures to help examine unidimensionality, an important measurement validity requirement (Bond & Fox, 2007) addressing the question of whether a single identifiable construct (i.e., diversity attitudes) was measured by all the selected diversity items and categories of issues those items represent. These analyses were conducted first separately on each pretest and posttest data sets for both the *course inconsistent* and the *course consistent* subsets of items in order to maximize the diagnostic potential. Measurement change between pretest and posttest were examined in logits and effect sizes, in conjunction with variability, error, and reliability indices to help emphasize magnitude and direction. The Winsteps (Linacre, 2011) computer program was used for analysis.

Both the pretest and the posttest scores of *course inconsistent* items were approximately normal in distribution, with skewness of .17 (*SE* = .26) and kurtosis of -.34 (*SE* = .51) for the pretest and skewness of .37 (*SE* = .26) and kurtosis of -.02 (*SE* = .51) for the posttest. Likewise for *course inconsistent* item score normality with skewness of .-.21 (*SE* = .26) and kurtosis of -.41 (*SE* = .51) for the pretest and skewness of .18 (*SE* = .26) and kurtosis of -.52 (*SE* = .51) for the posttest. Initial examination of standardized fit statistics revealed one item that misfitted the Rasch model on both the pretest, *z* infit = 2.5, *z* outfit = 2.7, and posttest, *z* infit = 3.8, *z* outfit = 3.8. Standardized weighted infit and unweighted outfit levels that are above 2.0 indicate underfit to the Rash model, resulting from an improbable pattern of responding on that item. Item number 7, the misfitting item read: "One's gender has little to do with one's educational opportunity." Though misfit alone was not used to determine inappropriateness of any items, this consistent underfit on both pretest and posttest and its potentially ambiguous wording were used to decide on removal of this single item (Bohlig, Fisher, Masters, & Bond, 1998), leaving 45 items with more favorable measurement characteristics on in this modification of the 64 item HRAT.

Pretest and posttest distributions of difficulty measures for both the *course consistent* and the *course inconsistent* item subsets all showed productive matches (i.e. correspondence) of item difficulties and person attitude distributions for the majority of students and items. Distributions for both items and students were primarily within 1 and -1 logits, indicating that overall differences among most students were not extreme, for the most part. Table 1 depicts item and person statistics summaries. Instrument reliability levels were relatively consistent and sufficiently high, either at or above the .92 level. Error rates for items (average of .12 to .14) and persons (.21 to .34) were at consistent levels for the respective number of items under examination. Fit statistics indicated good overall model fit though the combined analysis identified a potential misfit of the measurement model for 3 more items (infit, outfit > 2.0) that will be examined further with future samples of students to determine whether misfit items exist across student samples. Overall, the modified HRAT consisting of 45 items showed appropriate measurement properties with this type of student sample and the diversity course being examined. Measurement properties of scaled measures, reliability, error, infit, and outfit yielded suitable levels for meaningful analysis using the modified instrument.

*B. Likert Scale Statistics.*

Prior to examining measures derived from the raw data, an overview of the Likert scale counts provides important preliminary perspective on the data. While these raw counts are highly informative, comparisons of counts do not reflect the precise magnitude differences among items and among students that influence the measurement scale for attitudes. Table 1 (bottom section) shows overall counts of ratings from *Strongly Agree* to *Strongly Disagree*, which indicate that most students were in disagreement with *course inconsistent* statements and most students were likewise in agreement with *course consistent* statements, both before and after the course. Many students were uncertain about statements. In addition, when compared with pretest counts, posttest counts showed 5.9% average decreases in "uncertain" ratings, 12.0 % overall increases in "disagree" or "strongly disagree" categories for *course inconsistent* statements, and 9.7 % overall increases in "agree" or "strongly agree" categories for *course consistent* statements. Counts also indicate the numbers of posttest ratings that do not coincide with course information.

Counts on issue specific items also yielded posttest shifts to the "disagree" and "strongly disagree" categories on *course inconsistent* statements for 9.7% of ratings with *homosexuality* items, 8.6% of ratings with *race* items, 13.1% of ratings with *social class* items, and 12.4% of ratings with *women's equality* items. Similarly, with *course consistent* statements, counts yielded posttest shifts to the "agree" and "strongly agree" categories for 19.3% of ratings with *homosexuality* items, 4.3% of ratings with *race* items, 12.8% of ratings for *social class* items, and 8.4% of ratings with *women's equality* items. Averaging these pretest to posttest shifts toward course consistency in attitude ratings across statement types yields changes of 14.5% for *homosexuality,* 6.5% for *race*, 13.0% for *social class*, and 10.4% for *women's equality*. Although these general percentage shift trends provide a broad view of the change in attitudes, they do not account for the relative endorsement difficulty of the individual sub-issues within the four categories, nor do they represent precise measures of change. Discussion of endorsement difficulty and measures of change among sub-issues and as their implications follow.

*C. Measures of Difficulty Endorsing Diversity Issues.*

Linear measures constructed from a second combined Rasch analysis were used to conduct valid comparisons of the two survey administrations. Analyses included combining pretest and posttest data in two distinct ways, known as "racking and stacking" the data, to place pretest and posttest data on common frames of reference and measurement scales prior to making item and person comparisons (Wright, 2003). Racked data allowed a focus on items and associated diversity issue measures of endorsement difficulty while stacked data allowed a focus on student attitudes and changes. Although items such as these that assess attitudes are different than items assessing ability, we use the term 'difficulty' considering that higher levels reflect ideas that are not as agreeable to the students as a whole, or more difficult to endorse. This should not be confused with being more difficult to answer correctly, considering that there are no incorrect answers in attitude survey items.

Measures of endorsement difficulty are on the logit scale with zero as the mean score. Endorsement difficulty statistics are provided in Table 2 in aggregate and Table 3 by item both as context for the findings on student attitude and for future comparisons. These statistics on specific sub-topics addressed by each item can be usefully considered as instructional revision is planned. For example, items with small levels of measured change, as shown in the far right

column, represent sub-topics or questions that were interpreted in a similar way both before and after the course regardless of the changes in viewpoints among students. Very little change in endorsement difficulty between the two assessments is normal, indicating consistent measurement. However, when change occurs we expect positive change (increased difficulty) with course inconsistent items and negative change (decreased difficulty) with course consistent items due to the most probable influence of the course on the way issues are interpreted. That is, we expect the items that are consistent with course material to become more easily endorsed by students who might, as a result of the course, develop a stronger rationale for endorsing those items, stemming from what they learned. Similarly, items that are inconsistent with the course might become more difficult to endorse as a result of a weakened rationale for the perspective, stemming from what was learned. Relatively large change can reflect a measurement problem known as differential item functioning (DIF), indicating an item that is not functioning repeatedly in relative unison with the other items (Smith & Smith, 2004). This can result from the differences in the ways perspectives in particular issues are being considered at different times.

**Table 1. Item and person statistics on pretest and posttest.**

| Statistic | Course Inconsistent | | Course Consistent | |
|---|---|---|---|---|
| | Pretest | Posttest | Pretest | Posttest |
| **Items** | | | | |
| Number of Items | 30 | 30 | 15 | 15 |
| N (Participants) | 88 | 88 | 88 | 88 |
| Mean Raw Score | 247.7 | 227.9 | 307.9 | 323.5 |
| SD Raw Score | 48 | 45.1 | 41.7 | 37.4 |
| Max Raw Score | 347.0 | 328.0 | 386.0 | 380.0 |
| Min Raw Score | 149.0 | 146.0 | 249.0 | 250.0 |
| Item Reliability | .94 | .94 | .92 | .94 |
| **Persons** | | | | |
| Number of Items | 30 | 30 | 15 | 15 |
| N (Participants) | 88 | 88 | 88 | 88 |
| Mean Raw Score | 84.5 | 77.6 | 52.5 | 55.1 |
| SD Raw Score | 13.0 | 13.5 | 4.8 | 5.2 |
| Max Raw Score | 112.0 | 110.0 | 64.0 | 69.0 |
| Min Raw Score | 50.0 | 52.0 | 41.0 | 46.0 |
| Item Reliability | .86 | .87 | .62 | .72 |
| **Likert Rating Counts** | | | | |
| Strongly Agree | 174 | 125 | 173 | 230 |
| Agree | 641 | 540 | 581 | 652 |
| Uncertain | 661 | 515 | 326 | 248 |
| Disagree | 851 | 1036 | 212 | 161 |
| Strongly Disagree | 313 | 424 | 28 | 29 |

Although the majority of items across the four main categories of topics had relatively little change, as shown in Table 3, a few items had large unexpected change. For instance, nine of the ten *course inconsistent* items dealing with homosexuality did not show large amounts of change from pretest to posttest. However, the *course inconsistent* item on the specific sub-topic of homosexual fantasizing (#30) showed a relatively large unexpected decrease in endorsement difficulty (-.58 logits), suggesting a possible change in the way many students viewed this item and possibly the corresponding issue. Connecting this type of specific data to the instructional context of that particular sub-topic within the course material and discussion of homosexuality can be a useful part of considering possible enhancements.

Items that were at, near, or above one standard deviation (.50 logits) of unexpected change in difficulty measures are shown in bold on the far right column of Table 3 to highlight use of the item data for informing instructional revision. While the category of homosexuality had one item with large unexpected change, within the race category none of the items showed unexpected change at or above one standard deviation. Still, item #61 on racial segregation showed .38 logits of unexpected change making some reexamination of this sub-topic worth considering. Within the social class category only item #40 on stereotypes of the working class led to large unexpected change of .92 logits. Notable also in this category was item #48 on affecting children which had .40 logits of change in the unexpected direction, also making that sub-topic worth reexamining. Within the category of women's equality three of twelve items showed large unexpected change, including item #39 on working hard, at -.51 logits, #47 on hating men, at -.48 logits, and # 59 on sexist attitudes, at .75 logits. Each of these noted issues were scrutinized closely with respect to future instruction following this analysis.

Similarly, large change in the expected direction indicates that the perception of that item or its represented issue changed more than most from pretest to posttest, but in the direction suggested by course content. As an example, within the category of women's equality, item #19 on victims of sexism, led to large change at -1.07 logits that was supported by the course. Using this type of data can likewise be a constructive part of course reflection as recent revisions and new approaches are evaluated for their possible influence. Furthermore, in addition to examination of change, the relative difficulty among items and their subtopics shown before instruction and after instruction, without regard to change may also serve to inform instructors' ideas about how new students perceive the various issues. As an example, item #56 dealing with adults on welfare was relatively difficult for student to endorse at .92 logits, prior to the course. Instructors may benefit from being aware of that type of finding.

Using the data from Table 3, average change from all 45 items relative to the course consistent direction showed that the greatest amount of overall post-course change among these items occurred for the category of homosexuality at .45 logits, followed by women's equality at .22 logits, then race at .09 logits and lastly social class at -.01 logits. Pretest and posttest data columns on Table 3 regarding the difficulty to endorse confirm that race and social class subtopics were also the more challenging issues with respect to the course content. Generally, however, these data on endorsement difficulty indicate the importance of focusing on sub-issues within the broader categories of issues in order to gain specific awareness of the most likely influences of the course experience on students.

**Table 2. Item measures on pretest and posttest rack analysis.**

| Statistic | Course Inconsistent | Course Consistent |
|---|---|---|
| **Items Racked** | | |
| Mean Item Measure | .00 | .00 |
| SD Measure | .48 | .52 |
| Max Measure | .84 | .92 |
| Min Measure | -.93 | -1.33 |
| | | |
| Mean Error | .12 | .14 |
| SD Error | .01 | .02 |
| Max Error | .15 | .22 |
| Min Error | .09 | .10 |
| | | |
| Mean Z Infit | .00 | .1 |
| SD Z Infit | 1.0 | .5 |
| Max Z Infit | 2.0 | 1.4 |
| Min Z Infit | -2.2 | -1.0 |
| | | |
| Mean Z Outfit | .1 | .2 |
| SD Z Outfit | 1.1 | .6 |
| Max Z Outfit | 3.2 | 1.7 |
| Min Z Outfit | -2.2 | -.9 |
| | | |
| Item Reliability | .94 | .93 |
| Item Separation | 3.82 | 3.56 |

*E. Measures of Attitude Change on Diversity Issues.*

Linear measures of student attitudes from stacked data were compared. These analyses allowed us to answer the following research questions: a) whether student attitude measures primarily did or did not correspond with course information, b) whether there was change between pretest and posttest attitude measures, and d) what the relative size of the overall effects were. Comparisons of attitudes were again made relative to *course consistent* and *course inconsistent* items separately as shown in aggregate on Table 4.

For *course inconsistent* items, pretest to posttest measures indicated 60 students (68%) changed their attitudes about diversity toward correspondence with the course information, 4 students (5%) showed no change in attitude, and 24 students (27%) changed their attitudes toward the opposite direction of course information. However, 15 of those 24 students (17% overall) who's attitudes changed away from course information had posttest attitudes that primarily corresponded with the course information, but to a lesser degree than their attitudes on the pretest. Thus, only 12 students (14%) showed attitudes that both changed in the opposite direction of course information and were primarily not in correspondence with the course information. Also, 75 students (85%) either changed their attitudes as hypothesized or remained primarily correspondent with the diversity course content. Figure 1 shows the locations of student attitude measures plotted for pretest by posttest on *course inconsistent* items. In Figure 1, negative measures reflect more course consistent attitudes, as they represent lower agreement

**Table 3. Posttest fit statistics for the HRAT instrument, followed by pretest, posttest and change measures for difficulty to endorse.**

| Course Inconsistent Items on *Homosexuality* | Z Infit | Z Outfit | Pre | Post | Change |
|---|---|---|---|---|---|
| 10. unnatural | .5 | .4 | -0.19 | -0.07 | 0.12 |
| 14. gay rights | -1.5 | -1.2 | 0.11 | 0.64 | 0.53 |
| 26. many partners | -1.4 | -1.5 | 0.23 | 0.52 | 0.29 |
| 30. fantasize | 5 | .7 | 0.30 | -0.28 | **-0.58** |
| 38. promiscuous | .5 | 1.1 | 0.33 | 0.05 | -0.28 |
| 46. child molesters | -.3 | -.6 | 0.72 | 0.68 | -0.04 |
| 54. same sex relationships | 1.5 | 1.6 | 0.17 | 0.13 | -0.04 |
| 62. proves | .3 | .5 | 0.40 | 0.52 | 0.12 |
| 66. violent crimes | -1.2 | -1.2 | 0.15 | 0.51 | 0.36 |
| 70. a choice | 1.6 | 1.4 | -0.72 | -0.74 | -0.02 |
| | | | | | |
| Course Consistent Items on *Homosexuality* | | | | | |
| 22. mental illness | -.3 | .0 | 0.58 | -0.25 | -0.83 |
| 34. many contributions | .2 | .6 | 0.44 | -0.44 | -0.88 |
| | | | | | |
| Course Inconsistent Items on *Race* | | | | | |
| 09. do not achieve | -.1 | -.3 | 0.73 | 0.84 | 0.11 |
| 17. skin color | 1.8 | 2.6 | -0.91 | -0.81 | 0.10 |
| 33. affirmative action | -.3 | -.4 | -0.16 | -0.06 | 0.10 |
| 41. all-white communities | -.4 | -.3 | 0.23 | 0.31 | 0.08 |
| 45. stopped complaining | -1.6 | -1.8 | -0.14 | -0.07 | 0.07 |
| 57. same opportunity | .2 | .1 | -0.48 | -0.30 | 0.18 |
| 65. victims of racism | 2.0 | 2.8 | -0.93 | -0.91 | 0.02 |
| | | | | | |
| Course Consistent Items On *Race* | | | | | |
| 21. institutional racism | 1.4 | 1.4 | 0.49 | 0.65 | 0.16 |
| 37. get hired | -.1 | -.1 | 0.47 | 0.22 | -0.25 |
| 53. cultural racism | -.9 | -.9 | -0.03 | -0.64 | -0.61 |
| 61. racial segregation | -.5 | -.4 | -0.13 | 0.25 | 0.38 |
| | | | | | |
| Course Inconsistent Items on *Social Class* | Z Infit | Z Outfit | Pre | Post | Change |
| 12. will power | .1 | .7 | 0.07 | 0.24 | 0.17 |
| 20. want to work | -.9 | -.5 | 0.03 | 0.00 | -0.03 |
| 24. welfare assistance | .1 | .1 | 0.23 | 0.06 | -0.17 |
| 36. homeless | -.3 | -.1 | 0.28 | 0.75 | 0.47 |
| 44. dependent | .0 | .0 | -0.90 | -0.49 | 0.41 |
| 52. tax dollars | -.8 | -.9 | -0.46 | -0.63 | -0.17 |
| | | | | | |
| Course Consistent Items on *Social Class* | | | | | |
| 32. poverty | -.3 | -.3 | -0.58 | -0.50 | 0.08 |
| 40. stereotype working class | .9 | 1.7 | -0.44 | 0.48 | **0.92** |
| 48. affecting children | .3 | .4 | -0.28 | 0.12 | 0.40 |
| 56. adults on welfare | .3 | .7 | 0.92 | 0.07 | -0.85 |
| | | | | | |
| Course Inconsistent Items on *Women's Equality* | | | | | |
| 15. feminists | -.2 | .0 | -0.29 | 0.44 | 0.73 |

| | | | | | |
|---|---|---|---|---|---|
| 23. discrimination | .1 | .2 | -0.58 | -0.20 | 0.38 |
| 31. sexism | -.8 | -.9 | -0.76 | 0.15 | 0.91 |
| 39. work hard | 1.2 | 1.3 | -0.07 | -0.58 | **-0.51** |
| 47. hate men | -1.1 | -.9 | 0.46 | -0.02 | **-0.48** |
| 55. paid about the same | .0 | .5 | 0.28 | 0.50 | 0.22 |
| 63. rapes are perpetrated | .6 | 1.4 | 0.40 | 0.30 | -0.10 |
| | | | | | |
| Course Consistent Items on *Women's Equality* | | | | | |
| 11. sex role stereotypes | -.2 | -.1 | 0.75 | 0.37 | -0.38 |
| 19. victims of sexism | -.2 | -.2 | -0.26 | -1.33 | -1.07 |
| 43. occupations | -.4 | -.2 | -0.51 | -0.70 | -0.19 |
| 59. sexist attitudes | -.1 | .2 | -0.14 | 0.61 | **0.75** |
| 67. violence against women | -.5 | -.4 | 0.14 | -0.33 | -0.47 |

*Note:* Items are labeled by instrument number and an identifying topic term. Higher measured change corresponds with increased difficulty to endorse, and lower measured change corresponds with decreased difficulty to endorse. The complete instrument is available at http://wps.ablongman.com/ab_koppelman_humandiff_2/77/19966/5111404.cw/content/index.html

with statements that were not consistent with the course. Hence, student attitude locations to the left of the diagonal represent movement in the course consistent direction. This type of plot can be a very effective means of visualizing the differences and similarities in change across a group of students.

Measures for students illustrated on Figure 1 are also on the logit scale with zero as the mean score. To further assist with interpretation relative to the level of agreement and disagreement, .02 logits was the measured point corresponding with an average uncertainty level (Likert rating of 3) for *course inconsistent* items. Thus, measures above .02 shown in Figure 1 tended toward the agreement range while those below tended toward the disagreement range. The mean pretest measure of attitudes toward *course inconsistent* items was -.21 logits, 95% CI [-.33, -.10] (*SD* = .55), and the mean posttest score was -.50 logits, 95% CI [-.63, -.38] (*SD* = .58), a statistically significant difference of -.29 logits, 95% CI [-.40, -.19] (*SD* = .50), $t$ (174) = 3.42, $p$ = .001, favoring average change that continued toward greater course correspondent attitudes. After correcting for the correlation between means ($r$ = .61) using Morris and DeShon's (2002) equation, Cohen's $d$ = .45, a medium effect size for course correspondent attitude change (Cohen, 1988).

For *course consistent* items Figure 2 shows that pretest to posttest change in attitude paralleled that of *course inconsistent* items, but in the opposite direction as expected. Measures indicated 58 students (66%) changed their attitudes about diversity toward correspondence with the course information, 3 students (3%) showed no change in attitude, and 27 students (31%) changed their attitudes in the opposite direction of course information. However, 18 of the 27 (20% overall) who's attitudes changed away from course information had posttest attitudes that remained primarily correspondent with the course information, but to a lesser degree. Thus, only 9 students (10%) showed attitudes changing in the opposite direction of course information and also primarily not corresponding with the course information. In all, 76 students (86%) either changed their attitudes as hypothesized or remained primarily consistent with the diversity course content.

The measured point corresponding with an average uncertainty level (Likert rating of 3) for *course consistent* items was -.33 logits. The mean pretest score for attitudes on *course consistent* items was .36 logits, 95% CI [.26, .47] (*SD* = .51), and the mean posttest score was .68 logits, 95% CI [.53, .82] (*SD* = .67), a statistically significant difference of .31 logits, 95% CI

[.19, .44], $t$ (87) =-5.02, $p$ = .001. Again, correcting for the correlation between means ($r$ = .54) according to Morris and DeShon (2002), Cohen's $d$ = .52, also a medium effect size though slightly higher than that for course inconsistent items.

**Table 4. Person measures on pretest and posttest stack analysis.**

| Statistic | Course Inconsistent | Course Consistent |
|---|---|---|
| **Persons Racked** | | |
| Mean Person Measure | -.36 | .52 |
| SD Measure | .58 | .61 |
| Max Measure | .95 | 3.04 |
| Min Measure | -1.88 | -.65 |
| | | |
| Mean Error | .21 | .33 |
| SD Error | .01 | .05 |
| Max Error | .27 | .54 |
| Min Error | .20 | .28 |
| | | |
| Mean Z Infit | -.1 | -.1 |
| SD Z Infit | 1.8 | 1.4 |
| Max Z Infit | 6.7 | 3.8 |
| Min Z Infit | -4.3 | -4.2 |
| | | |
| Mean Z Outfit | -.1 | -.1 |
| SD Z Outfit | 1.8 | 1.4 |
| Max Z Outfit | 6.4 | 3.4 |
| Min Z Outfit | -4.0 | -4.0 |
| | | |
| Person Reliability (Cronbach Alpha) | .87 | .62 |
| Person Separation | 2.62 | 1.54 |

*F. Student Learning and Interest Level Perspectives.*

Three items from the end of course student evaluations were examined to provide student perspective on their learning and interest levels. These perspectives were considered an important part of the context for interpreting the diversity attitudes findings. The first item examined was, "Compared to other courses of similar credit value: How much did you learn from this course?" An aggregate of the four course sections were examined. Data were similar across all sections showing that overall 8% responded with either "much less" or "less," 26% responded with "about the same," and 66% responded with "more" or "much more." The other two items examined related to course interest. The first of this pair of items was "What was your interest in this subject matter before taking this course?" and the second was "What was your interest in this subject matter after taking this course?" The percentages of before-course versus after-course interest reported respectively were 21% (before) versus 6% (after) who reported "no interest at all" or "mildly interested," 42% (before) versus 20% (after) who reported "average," and 36% (before) versus 74% (after) who reported "interested" or "very interested." Overall, self-reports of post-course learning and change in interest levels showed some variation among

students but were predominantly supportive of perceived learning relative to other courses and increased interest in the subject following the course.

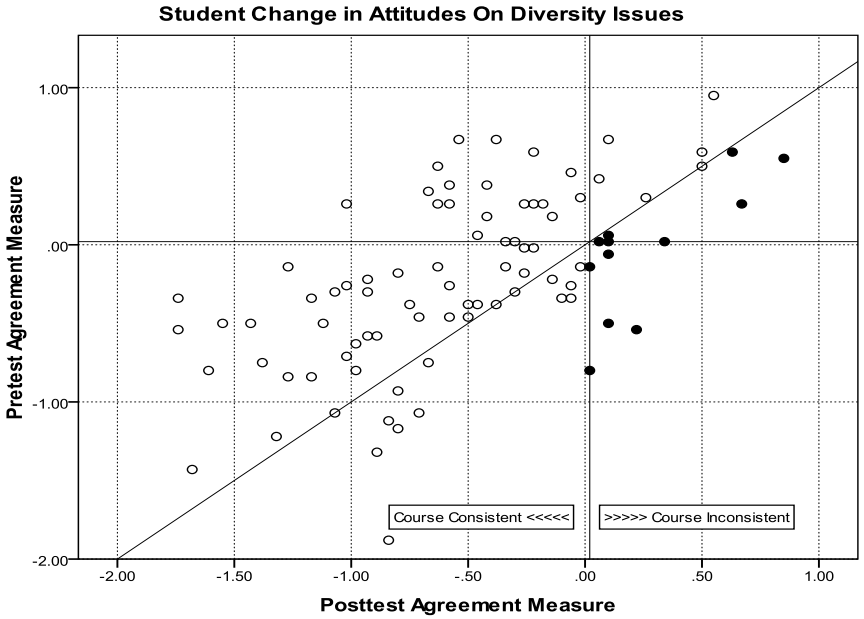


**Figure 1. Pretest to posttest change in attitudes on course inconsistent diversity statements.** Lower measures represents course consistent attitudes. Students located to the left of the diagonal line showed measured change toward increasingly course consistent diversity attitudes. Students located to the left of the vertical line had mostly course consistent attitudes regardless of change direction, and vice versa. Students who were both to the right of the vertical line and to the right of diagonal line (darkened circles) changed away from course consistent attitudes. Student locations adjacent to or at the diagonal line showed little or no change.

*G. Summary of Findings.*

Instrument measurement characteristics were favorable though imperfect and possibly improvable with the 45 selected items from Koppelman and Goodhart's (2005) HRAT survey. Survey items were a good match for this student sample, but there was some redundancy among items toward the center of the distributions. Also, for this type of student group there appeared to be need for a few items that are at both difficulty level extremes than any of the current items. Fit statistics suggest the need for continued evaluation of a few existing items (see Table 3 infit and outfit columns) as other student samples are assessed. While there was good evidence for the overall unidimensionality of this portion of the HRAT instrument, endorsement difficulty changed slightly from pretest to posttest and varied among items and their corresponding issues. Clearly, change was not limited simply to student attitudes, but also appeared to occur with regard to the student interpretations of some issues and questions themselves, relative to others. These findings support the benefit of further research on the dynamics of interpreting individual diversity attitude items and issues across time and experiences. On the whole, the Rasch statistics as shown in Tables 2, 3, and 4 provided strong reasons to consider the linear measures that were constructed from the Likert data, while imperfect, to support valid, useful comparisons of interest.

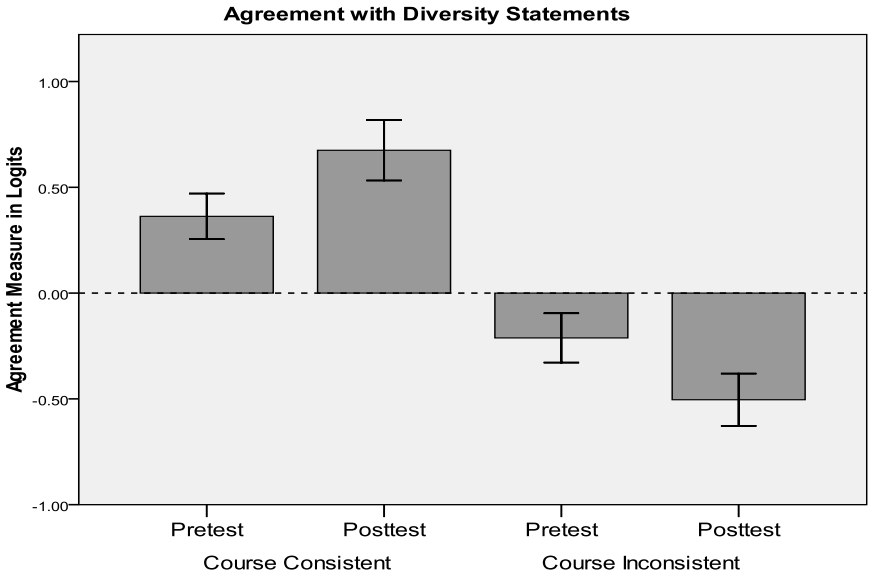


**Figure 2. Comparisons of pretest and posttest measures of agreement (and 95% confidence intervals) with diversity item statements that are course consistent and course inconsistent.** The mean measure for each comparison is set at zero logits. Effect sizes for changes illustrated were $d$ = .52 standard deviation units for course consistent change and $d$ = .45 standard deviation units for course inconsistent change.

Change in course consistent attitudes was shown by large majority of students who reported different levels of item endorsement following the course, the pretest to posttest changes in student measures, and the effect sizes that were just below and just above one half of one standard deviation on *course consistent* and *course inconsistent* item types respectively. That is, most students demonstrated measurable change in their reported diversity attitudes between the beginning and the end of the diversity course in line with the information provided in the course. Findings also identified a number of students whose attitudes changed in opposition to the course information. This outcome corresponded with the instructor's recognition that a few students within each class continually exhibited resistance to many of the issues of diversity, which, at the very least offered us a vantage point into the vigorous challenge by some students to the intent of the diversity curriculum. By isolating the data from these nine students (Figure 1), the nature and magnitude of their resistance, relative to the attitudes of the rest of the class were more easily examined.

Whether questions were framed to be consistent or inconsistent with course information did not appear to greatly affect their reported attitudes, a finding of importance to future investigations and the use of reverse coding of survey data. One of the benefits of having questions framed differently is in verifying whether each student completed the survey in a consistent rather than careless or random manner, regardless of whether other possible threats to internal validity such as a pretesting, maturation, or history influence student responses. The fact that the course consistent versus the course inconsistent question framing effect sizes differed by only $d$ = .07 provides support for the relative consistency of the measures.

Item endorsement difficulty measures indicated that a few sub-issues warranted particular attention toward future instructional revision, as indicated by apparent differential item

functioning. Thus, measures of endorsement difficulty were shown to be a useful tool for examining changes on sub-issues that can occur among students following instruction. With difficulty measures using the same scale as attitude measures, change magnitude for difficulty could be considered from a similar frame of reference as that of attitudes.

## IV. Discussion.

As Neito and Bode (2008) have emphasized, our students' understanding of diversity supports effective education in a changing world. Among other reasons for this is the increasing interdependence of our communities, which requires that students better understand one another, including varieties among people in a global society. In this investigation we sought to track the component of growth in student understanding that was reflected in attitudes toward acceptance of human differences. The evidence of student growth shown in this investigation is a type often overlooked by standard assessment practices in higher education. The Rasch measures provided a means to specify whether and how the diversity course influences went beyond learning the required material on diversity research. Most students in the course changed the way they think about diversity issues, and to a substantial degree. While these students' attitudes typically became more accepting of diversity among people, findings also provided evidence for the need to adjust the way several sub-topics across the categories of race, homosexuality, social class, and women's equality are taught within the course. For instance, the results of this survey have been used to bring a greater focus on the race category in subsequent semesters using the guiding framework described at the beginning of this article. The instructor recognized that a greater variety and better balance of personal narratives than previously achieved in the classroom would logically be instrumental in helping students develop the connections that lead to change. To illustrate, the following vignette is from student who was enrolled in a semester after the results of the survey were interpreted. In this excerpt from an assignment the student comments on the story of a classmate who up until recently was an "unregistered alien." The classmate was, in fact, one of the highest achieving students from the prior year, and her story appeared to touch many in the class profoundly.

> What I view as the turning point was when a female student in class opened up about the Mexican coyotes. I had never heard this term before, and in all honesty I just assumed that people who crossed the border merely had to walk across a fence when a guard's back was turned and they were in. It was seeing the raw human struggle that changed me. All of a sudden, the term illegal alien was no longer some abstract concept attached to a subhuman, taco eating fiend, it was someone's mother. It was a she, and that started a change in me. (N. Adams, personal communication, December 12, 2010)

### *A. Changes in Student Perspectives on the Difficulty to Endorse.*

Beyond the general trends that express consistency with the targeted goals of the curriculum, some of specific topics of instruction that were made salient in Table 3 bring to light some areas of focus that may call for an adjustment in future curricular design as well as areas that appear to be addressed well. One of the most unexpected areas of focus had to do with issues of women's equality. For example, item 39 reads, "Women shouldn't be given the rights feminists are demanding; women must first work hard and earn them." With this item there was a -0.51 logit

change (relatively easier to endorse) toward greater inconsistency with the targeted goal of instruction, which was surprising given the fact that the majority of students who took this survey were women. Another example in this topic area is expressed in the misinterpretation of feminists as "women who hate men" (Item 47). There was a -0.48 logit change also in an opposite direction (easier to endorse) of the instructional orientation with this item. Related to this finding the instructor has already chosen as an example the kind of personal narrative mentioned earlier for use in dispelling feminist stereotypes (Kress, 2012). Furthermore, with item 59 which reads, "Most men in our society are not aware of their sexist attitudes," there was a change of .75 logits (more difficult to endorse) also in the direction inconsistent with the instructional orientation on this topic. On the other hand item 19 on sexism which reads "Both females and males are victims of sexism" showed a relatively large decrease in difficulty to indorse of -1.07 logits, consistent with the more expected influence of the course orientation. This change likely occurred because of the salient course material highlighting that men as well as women are victims of sexual harassment.

*B. Resistance and Entrenchment Among Some Students.*

Of additional interest to the instructor was the evidence provided by data that exposure to diversity issues led to a small portion of students possibly becoming more deeply entrenched, or polarized (Kuhn & Lao, 1996), in their views that are contrary to the information and discussion provided by the course. This finding suggests the need for closer analysis of attitudes on specific issues for this type of attitude entrenched subgroup. The fact that these data could help identify the number of students (Figure 1) who most clearly resisted the curriculum and compare that measured resistance level to those who did not resist using the same measurement scale allowed a useful means of understanding student viewpoints more thoroughly. Additionally, if further investigations indicate this possible entrenchment effect to be common across different populations, deeper examination of the reasons is a next step toward improving instruction on diversity. The various dispositional and experiential factors identified by Garmon (2004, 2005) appear to be useful for consideration within focused inquiries on attitude polarization and entrenchment that may result from student engagement with diversity topics.

*C. Benefits, Challenges, and Reflections for Future Planning.*

Overall, despite the lack of an experimental control group, the analyses of change were of real value to the instructional improvement process of this course, in part because these repeated measures were calibrated on the same scale for improved comparison. Additionally, honesty in responding, encouraged through the survey instructions and procedures, was supported by data. The students who provided responses in resistance to the curricula served as one indicator that there was not apparent pressure to respond dishonestly yet in a socially desirable manner according to what students might assume the instructor or institution would prefer. However, it is crucial to be aware with this type of change analysis that effects observed can be due to influences beyond the instruction, such as other experiences at either pretest or posttest.

The survey administration and interpretation were found to be sustainable in terms of the effort and time needed during a semester. Data entry and analysis were considered the most time consuming aspects for attempting to sustain the use of this type of study each semester. Access to a systematic analyses process, preferably with a campus testing and assessment office, would

perhaps allow the best means to sustaining this type of pretest-posttest process involving the construction of Rasch measures from the survey data. Such a resource was not available at the institution where this study occurred. Still, encouraging is the fact that this type of analysis is becoming increasingly accessible to educators through personal computing tools and software now available (Bond & Fox, 2007; Linacre, 2011).

Having reflected on success and challenges faced with pedagogical strategies used and the data from this investigation, the course instructor has initial evidence to support that the course consistent attitude changes observed among most students can result from a) deliberate effort to create a non-threatening class environment that attempts to welcome and value all perspectives, b) direct teaching that addresses misinformation, and c) using personal narratives toward the goal of creating empathy across differences in gender and sexual identity. More specifically, the instructor sets the tone in the very beginning of the semester by assuring students that the purpose of the course is not to make students feel guilty about their backgrounds in terms of race, class, gender, or ethnicity. In addition, the more controversial topics such as homosexuality are covered toward the end of the semester, after there has been a sense of group trust established. One way this is achieved is through two configurations of class meetings. One configuration involves the instructor presenting material in a whole class setting with much interaction from the students. Many narratives like those mentioned in the beginning of this article come out of this type of meeting. The second configuration of class meetings involves the students working in groups of three or four while responding to group study-guide questions as they collaboratively prepare a single group document. This smaller setting can encourage more openness and trust in a way that the students tend to appreciate. Also during the semester, each student is required to interview in-depth someone who is "other" than them. Though these strategies were considered crucial to instructional effectiveness, it is important to consider whether these strategies or other course characteristics encouraged the attitude entrenchment found among a small number of students. This question will be the topic of subsequent investigation.

Finally, the course instructor has used the measures of change from this study alongside results from course examinations of required topics. By examining both performance and attitude data together he can carefully consider where he might place greater focus on specific areas of instruction with the intent of more effectively using multiple forms of evidence to inform future practices.

## References

Banks, J. (2006). *Cultural Diversity and Education: Foundations, Curriculum and Teaching.* Boston: Pearson, Allyn & Bacon.

Banks, J. (1993). Multicultural education: Historical development dimensions and practice. *Review of Research in Education, 19*(1), 3-49.

Banks, J. (1973). *Teaching Ethnic Studies.* Washington, D.C.: National Council for the Social Studies.

Bell, L. (2002). Sincere fictions: The pedagogical challenges of preparing white teachers for multicultural classrooms. *Equity and Excellence in Education, 35*(3), 236-244.

Bohlig, M., Fisher, W.P. Jr., Masters, G.N., and Bond, T. (1998). Content validity and misfitting items. *Rasch Measurement Transactions, 12*(1), 607.

Bond, T.G., and Fox, C.M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Castro, A.J. (2010). Themes in the research on preservice teachers' views of cultural diversity: Implications for researching millennial preservice teachers. *Educational Researcher, 39*(3), 198-210.

Cochran-Smith, M. (1995). Color-blindness and basket weaving are not the answers: Confronting the dilemmas of race, culture and language diversity in teacher education. *American Education Research Journal, 32*(3), 493-522.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coleman, J. S. (1966). Equal schools or equal students? *The Public Interest, 1,* 70-75.

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts, and implications. *International Education Journal, 5*(2), 125-143.

Delpit, L. (1995). *Other People's Children: Cultural Conflict in the Classroom.* New York: New Press.

Dimitrov, D., and Rumrill, P. (2003). Pretest-posttest designs and the measurement of change. *Work, 20*, 159-165.

Du Bois, W.E.B. (1903). *The Souls of Black Folk.* New York: Bantam Classic.

Dugard, P., & Todman, J. (1995). Analysis of pre-test-post-test control group designs in educational research. *Educational Psychology, 15*(2), 181-199.

Garmon, M.A. (2005). Six key factors for changing preservice teachers' attitudes/beliefs about diversity. *Educational Studies, 38*(3), 275-286.

Garmon, M.A. (2004). Changing preservice teachers' attitudes/beliefs about diversity: What are the critical factors? *Journal of Teacher Education, 55*(3), 201-213.

Gay, L. R., Mills, G. E., and Airasian, P. (2006). Educational research: Competencies for analysis and application. Upper Saddle River, NJ: Pearson.

Georgia Southern University College of Education. (2006). *Conceptual framework*. Retrieved April 1, 2012, from http://coe.georgiasouthern.edu/pdfs/cfram.pdf

Grant. C. (1977). *Multicultural Education: Commitments, Issues and Applications.* Washington, D.C. Association for Supervision and Curriculum Development.

Greene, M. (1995*). Releasing the Imagination: Essays on Education, the Arts, and Social change*. San Francisco: Jossey-Bass.

González, N., Moll, L., and Amanti, C. (Eds.). (2005). *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*. Mawah, NJ: Lawrence Erlbaum Associates.

Hilliard, A.G., III. (1974). Restructuring teacher education for multicultural imperatives. In W. A. Hunter (Ed.), *Multicultural Education Through Competency Based Teacher Education* (pp. 40-55). Washington, DC: American Association of Colleges for Teacher Education.

Klassen, F, and Gollnick, D. (1977). *Pluralism and the American Teacher.* Washington, D.C.: American Association of Colleges for Teacher Education.

Koppelman, K.L., and Goodhart, R.L. (2005). *Understanding Human Differences: Multicultural Education for a Diverse America*. Boston: Pearson.

Kress, T. (2012). Rethinking feminism: Towards a critical-feminist pedagogy for solidarity and social movement. In Pruyn, M., Malott, C. & Orelus, P. (Eds.) *Paths to Gender Justice in Education: Theories & Practices*. Charlotte, NC: Information Age Publishing.

Kuhn, D., and Lao, J. (1996). Effects of evidence on attitudes: Is polarization the norm? *Psychological Science, 7*(2), 115-120.

Larke, P. (1990). Cultural Diversity Awareness Inventory: Assessing the sensitivity of preservice teachers. *Action in Teacher Education, 12*(3), 23-30.

Linacre, J.M. (2011). *Winsteps® Rasch measurement computer program*. Beaverton, OR: Winsteps.com.

Martin, J.R. (2002). *Cultural Miseducation: In Search of a Democratic Solution.* New York: Teachers College Press.

Menchaca, M. (1997). Early racist discourses: The roots of deficit thinking. In R. Valencia (Ed.) *The Evolution of Deficit Thinking: Educational Thought and Practice,* (pp. 13-40). Washington D.C.: Falmer Press.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*(9), 74-149.

Morris, S.B., and DeShon, R.P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

Nieto, S. (2000). Placing equity front and center: Some thoughts on transforming teacher education for a new century. *Journal of Teacher Education, 51*(3), 180-187.

Neito, S., and Bode. P. (2008) *Affirming Diversity: The Socio-Political Context of Multicultural Education*. Boston: Pearson.

Nieto, S. (1986). Equity in education: The case for bilingual education. *Bulletin of the Council on Interracial Books for Children, l7* (3 & 4), 4-8.

Pearl, A. (1997). Cultural and accumulated environmental deficit models. In R. Valencia (Ed.), *The Evolution of Deficit Thinking: Educational Thought and Practice.* (pp. 132-159). London: Falmer Press.

Pollock, M. (2004). *Colormute: Race Talk Dilemmas in an American School.* Princeton, NJ: Princeton University Press.

Rasch, G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.

Reynolds, W., and Webber, J. (2004). Introduction. In Reynolds, W., & Webber, J. (Eds.), *Expanding Curriculum Theory* (1-18). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schofield, J. (1997). Causes and consequences of the colorblind perspective. In J. Banks & C. M. Banks (Eds.), *Multicultural Education: Issues and Perspectives.* (pp. 251-271). Needham Heights, MA: Allyn & Bacon.

Sleeter, C. (1995). White preserves students and multicultural education coursework. In J. Larkin & C. Sleeter (Eds.), *Developing Multicultural Teacher Education Curricula.* (pp. 17-30). Albany, NY: State University of New York Press.

Sleeter, C.E. and Grant, C.A. (2009). *Making Choices for Multicultural Education: Five Approaches to Race, Class and Gender*. New York: Wiley.

Smith, E.V., and Smith, R.M. (2004). *Introduction to Rasch Measurement: Theory, Models, and Applications*. Maple Grove, MN: JAM Press.

Valencia, R., and Solorzano, D. (1997). Contemporary Deficit Thinking. In R. Valencia (Ed.), *The Evolution of Deficit Thinking: Educational Thought and Practice.* (pp. 160-210). London, Washington, D.C.: Falmer Press.

Valenzuela, A. (1999). *Subtractive Schooling. U.S.-Mexican Youth and the Politics of Caring.* Albany, NY: State University of New York Press.

Wilson, T.D., Lindsey, S., and Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.

Woodson, C. (1933). *The Miseducation of the Negro*. New York: AMS Press.

Wright, B.D. (2003). Rack and stack: Time 1 vs. time 2. *Rasch Measurement Transactions, 17(1),* 905-906.

Wright, B.D., and Stone, M.H. (1979). *Best Test Design*. Chicago: MESA Press.

Zeichner, K. (1996). Educating teachers for cultural diversity. In K. Zeichner, S. Melnick, & M. Gomez (Eds.), *Current Reforms in Preservice Teacher Education.* (pp. 133-175). New York: Teachers' College Press.