



JoSoTL



JOURNAL OF THE SCHOLARSHIP OF TEACHING AND LEARNING

Volume 12, Number 3, September 2012

josotl.indiana.edu

Volume 12	Number 3	September 2012
Vathsala Sagayadevan & Senthur Jeyaraj	The role of emotional engagement in lecturer-student interaction and the impact on academic outcomes of student achievement and learning	1
Frances G. Smith	Analyzing a college course that adheres to the Universal Design for Learning (UDL) Framework	31
Annette Kujawski Taylor & Patricia Kowalski	Students' misconceptions in Psychology: How you ask matters...sometimes	62
Jana Hackathorn, Kathryn Cornell, Amy M. Garczynski, Erin D. Solomon, Katheryn E. Blankmeyer, & Rachel E. Tennial	Examining exam reviews: A comparison of exam scores and attitudes	78
Erin Malone & Amie Spieth	Team-based learning in a subsection of a veterinary course as compared to standard lectures	88
Karen S. Meaney, Jeff Housman, Arnaldo Cavazos, & Michelle L. Wilcox	Examining service-learning in a graduate physical education teacher education course	108
Jennifer M. Johnson & Glenn W. Lambie	A multicultural personal growth group as a pedagogical strategy with graduate counseling students	125
Robert Lake and Kent Rittschof	Looking deeper than the gradebook: Assessing cultural diversity attitudes among undergraduates	142
Richard L. Mehrenberg	Book Review Team Teaching: Across the Disciplines, Across the Academy	165
	JoSoTL Mission	167
	Submission Guidelines	168
	Editorial Board	169
	Style Sheet	170

Erratum

Citation:

Taylor, A.K., & Kowalski, P. (2012). Students' misconceptions in psychology: How you ask matters...sometimes. *Journal of the Scholarship of Teaching and Learning*, (12)3, 62-72.

As of January 6, 2015, the above article was reduced in length by 5 pages due to the removal of an appendix. Subsequent articles in this issue were not repaginated.

The role of emotional engagement in lecturer-student interaction and the impact on academic outcomes of student achievement and learning

Vathsala Sagayadevan¹ and Senthu Jeyaraj²

Abstract: Engagement has been studied as a multidimensional construct consisting of three subtypes: behavioral, cognitive, and emotional (Fredricks, Blumenfeld, & Paris, 2004). Among these, behavioral engagement has received the most and emotional engagement, the least attention (Fredricks et al., 2004). The current study thus aimed to examine the relationship between lecturer-student interaction, emotional engagement (specifically affective reactions expressed within the classroom), and academic outcomes (such as, student achievement and learning) in a sample of 140 undergraduate psychology students (M = 24, F = 116). Participants were randomly assigned to one of the experimental conditions (i.e., good vs. poor lecturer-student interaction) and completed the Lecturer-Student Interaction (LSI) questionnaire, Class-related Emotions Questionnaire (CEQ), Perception of Learning (POL) Questionnaire, and two measures of academic achievement. Individuals who shared a good interaction with their lecturer reported higher levels of emotional engagement compared to those who shared poor interactions with their lecturers. In addition, while emotional engagement failed to mediate the pathway between lecturer-student interaction and academic achievement, it was found to partially mediate lecturer-student interaction and student learning. The present findings highlight the significance of emotional engagement in enhancing learning outcomes in students.

Keywords: emotional engagement, teacher-student interaction, student learning, academic achievement

Engagement, defined as ‘energy in action’ (Appleton, Christenson, Kim, & Reschly, 2006, p. 428) represents the connection between an individual and the activity in which one is involved (Ainley, 2004; Appleton et al., 2006). The study of engagement is valued for both its positive academic outcomes (Appleton et al., 2006; Fredricks, Blumenfeld, & Paris, 2004) and psychosocial benefits (Reddy, Rhodes, & Mulhall, 2003). In general, high levels of engagement are associated with enhanced achievement (Barkatsas, Kasimatis, & Gialamas, 2009; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996; Wigfeld & Eccles, 2000), effective learning, acquisition of knowledge and skills (Furlong, Whipple, Jean, Simental, Soliz, & Punthuna, 2003; Ladd & Dinella, 2009) as well as better emotional functioning (Skinner, Furrer, Marchand, & Kindermann, 2008). In addition, it also serves as a protective factor against student dropout and involvement in risky activities (Connell, Spencer, & Aber, 1994; Finn, 1989; Finn & Voelkl, 1993; Jimerson, Campos, & Greif, 2003; Skinner et al., 2008; Skinner, Wellborn, & Connell, 1990). While engagement is regarded as more crucial for influencing various outcomes among

¹ vathsala.sagayadevan@my.jcu.edu.au

² senthu.jeyaraj@jcu.edu.sg

at-risk individuals (e.g., ethnic minorities) (Birch & Ladd, 1997), it has been shown to hold equal relevance for the general student population (Klem & Connell, 2004).

I. Engagement Subtypes.

Past research has studied engagement as a multidimensional construct consisting of three main subtypes: behavioral, emotional, and cognitive engagement; each of which have their own distinct markers (Appleton et al., 2006; Jimerson et al., 2003; Ladd & Dinella, 2009; Sinclair, Christenson, Lehr, & Anderson, 2003). Some of the markers of behavioral engagement include class participation (Chapman, 2003; Jimerson et al., 2003) and exertion of effort on task (Skinner et al., 2008); whereas those of cognitive engagement consist of self-regulation and use of learning strategies (Chapman, 2003; Fredricks et al., 2004). Likewise, emotional engagement can be identified through markers such as identification with the academic institution (Finn, 1989) and expression of affective reactions (e.g., interest) in the classroom (Fredricks et al., 2004; Skinner et al., 2008). The emotional subtype however, has received less attention in comparison to the behavioral and cognitive subtypes (Fredricks et al., 2004). This is partly attributed to the lack of conceptual clarity associated with this subtype (Fredricks et al., 2004). Appleton et al. (2006) for instance, used the term psychological engagement to examine achievement outcomes among high school students though the former included markers (e.g., identification) similar to that of emotional engagement.

A. Emotional engagement as a predictor.

In general, past studies have provided evidence for the role of engagement in influencing academic outcomes such as achievement and learning (Handelsman, Briggs, Sullivan, & Towler, 2005; Skinner et al., 2008; Skinner et al., 1990). Achievement, defined as an indicator of individuals' academic ability is usually assessed through grades on exams and standardized achievement tests (McLean, 2001); whereas learning represents the process through which an individual acquires knowledge (McLean, 2001). Despite this distinction between the two academic outcomes, studies have mainly focused on the role of engagement in influencing achievement of individuals as opposed to learning (e.g., Furrer & Skinner, 2003). In particular, the dominant focus has been on both the behavioral and cognitive subtypes given evidence for their pivotal role in influencing academic achievement (Fredricks et al., 2004; Ladd & Dinella, 2009). Ladd, Birch, and Bus (1999) for instance, found higher levels of cooperative and independent participation (i.e., behavioral engagement) to be predictive of higher achievement among kindergarten children. Similarly, certain markers of cognitive engagement (e.g., self-regulation, persistence, and effort) have also been found to predict academic achievement (Miller et al., 1996).

Emotional engagement as a predictor of academic achievement however has yielded mixed evidence with primary support for its role emerging from studies examining this subtype as part of a general or combined measure (i.e., with behavioral or cognitive engagement) of the construct (Fredricks et al., 2004). For instance, a combined scale of emotional and behavioral engagement was found to predict school performance in African-American youths (aged 10 – 16 years) (Connell et al., 1994). However, given that the separate effects of these subtypes were not examined, evidence for the role of emotional engagement in predicting school performance is inconclusive.

Additional support for this subtype has also been gleaned from research examining achievement emotions (Pekrun, Goetz, Titz, & Perry, 2002). These emotions defined as affective reactions expressed within the classroom (i.e., also an indicator of emotional engagement) comprise of emotions such as enjoyment, hope, and anger which have been linked to achievement outcomes (Frenzel, Thrash, Pekrun, & Goetz, 2007; Pekrun et al., 2002). In support of this, Pekrun, Goetz, Frenzel, Barchfeld and Perry (2011) found positive emotions such as enjoyment to be positively related to the Grade Point Average (GPA) of undergraduate psychology students. However, the associative nature of this finding once again failed to support the role of emotional engagement as a predictor of academic achievement.

Contrary to the aforementioned findings, other studies have found some evidence implying the role of emotional engagement as a predictor of learning as opposed to academic achievement. For instance, Handelsman et al. (2005) developed a student engagement instrument comprising of emotional, skills, performance, and participation/interaction engagement. Although all four subtypes were associated with academic achievement (e.g., assignments grades); only skills (similar to cognitive engagement), performance, and participation/interaction engagement (similar to behavioral engagement) emerged as significant predictors of these outcomes (Handelsman et al., 2005). Emotional engagement in contrast, was predictive of intrinsic outcomes associated with learning (e.g., valuing learning in its own right) (Handelsman et al., 2005). Likewise, Ainley and Ainley (2011) found students' enjoyment of science (i.e., achievement emotions) among other factors to positively predict an interest in learning more about science topics. The current state of evidence thus, reflects limited understanding with respect to the role of emotional engagement in predicting academic outcomes such as achievement and learning.

B. Emotional engagement as an outcome.

Among the various contextual factors posited to influence student engagement, teacher-student interaction (also studied as teacher support and teacher-student relationship) has received substantial support for being the strongest predictor of engagement and the most significant contributor of academic outcomes (e.g., Lam et al., 2012). Past studies examining various aspects of this contextual factor have generally identified a good quality teacher-student interaction to be characterized by high levels of emotional (Fraser & Fisher, 1982; Patrick, Ryan, & Kaplan, 2007), academic, autonomy support, (Skinner & Belmont, 1993) and provision of structure (Jang, Reeve, & Deci, 2010); which have in turn been positively associated with engagement of individuals. Skinner et al. (2008) for instance, found student reports of teacher support (i.e., involvement, structure, autonomy support) to be predictive of increases in emotional engagement and declines in emotional disaffection across the year. Likewise, a longitudinal study by Skinner and Belmont (1993) found teacher involvement (similar to emotional support) to predict emotional engagement in elementary school students. However, both these studies were conducted on young children for whom teacher support is generally regarded as critical (Birch & Ladd, 1997). Hence, whether this contextual factor plays as much an important role among older students (e.g., tertiary students) is yet to be established through further research.

C. Emotional engagement as a mediator.

Given the established links between teacher-student interaction and engagement as well as engagement and academic outcomes, the Self-System Model of Motivational Development (SSMMD) provides a relevant theoretical framework in examining the pathway linking teacher-student interaction, engagement, and academic outcomes (Skinner et al., 2008). According to this model, features of a particular context (e.g., characteristics of teacher-student interaction) are posited to influence the three basic psychological needs of individuals (i.e., need for competence, autonomy, and relatedness) (Ryan & Deci, 2000; Skinner et al., 2008). The extent to which these needs are fulfilled is in turn expected to predict the level of engagement of individuals, which then predicts their academic outcomes (Skinner et al., 2008).

Consistent with this, Hughes and Kwok (2007) found the quality of teacher-student relationship to indirectly predict math and verbal scores of first-grade students through engagement. Likewise, Klem and Connell (2004) found reports of teacher-support (e.g., provision of structure) to have an indirect influence on achievement scores of students through engagement. However, given that both these studies defined engagement of individuals in predominantly behavioral terms (e.g., participation, effort) (Hughes & Kwok, 2007; Klem & Connell, 2004), the applicability of these findings to subtypes such as emotional engagement is unclear.

D. Insights gained from the review of literature on engagement and academic outcomes.

In view of existing findings, it is evident that the lack of conceptual clarity associated with emotional engagement has resulted in this subtype being less researched in comparison to the behavioral and cognitive subtypes (Fredericks et al., 2004). In addition, the tendency to incorporate the emotional subtype into a general or combined measure of engagement has yielded mixed evidence with respect to its role in influencing academic achievement and learning in individuals. As such, the current study defines emotional engagement in terms of “affective reactions of students expressed within academic settings”, given that this aspect has been associated with academic outcomes in general (Frenzel et al., 2007; Pekrun et al., 2002, p. 93). Based upon this conceptually specific definition, the study aimed to explore the unique role of the emotional subtype in influencing achievement and learning of individuals. Furthermore, it sought to clarify if this subtype was more influential in affecting learning than academic achievement of individuals. Moreover, given that past studies have predominantly focused on middle and high school students (e.g., Appleton et al., 2006) within traditional classroom settings (e.g., Reeve & Tseng, 2011), the present study used the SSMMD as a general framework to examine the links between lecturer-student interaction, emotional engagement, and academic outcomes among tertiary students.

E. Hypotheses of the current study.

Three hypotheses were proposed in an attempt to address the gaps identified through the review. Given that past studies have indicated a positive association between the quality of teacher-student interaction (i.e., characterized by emotional, academic, autonomy support, structure) and engagement (e.g., Hughes & Kwok, 2007), hypothesis one aimed to investigate if students who share a good interaction with their lecturer report higher levels of emotional engagement

compared to those who share a poor interaction with their lecturer. Second, in line with studies showing teacher support to be positively related to academic achievement (e.g., Lam et al., 2012), hypothesis two sought to investigate whether students who share a good interaction with their lecturer are more likely to have higher achievement in their respective modules compared to those who shared a poor interaction. Finally, given the inconclusive evidence with respect to the emotional subtype in influencing academic outcomes, hypothesis three aimed to investigate the role of emotional engagement between lecturer-student interaction and academic achievement, as well as lecturer-student interaction and student learning.

II. Method.

A. Design.

The study is comprised of a pilot study and the research study. The independent variable (IV) is the quality of lecturer-student interaction and this consists of two levels: good and poor lecturer-student interaction. The dependent variables (DVs) are academic achievement and perceived learning; whereas the mediator is emotional engagement (MV).

B. Participants.

One-hundred and forty undergraduate psychology students (M = 24, F = 116) from James Cook University, Singapore campus were recruited for the current study; 20 of whom participated in the pilot study. The age of participants ranged from 18 to 53 years ($M = 22.79$, $SD = 4.78$). Participants were recruited primarily through convenience sampling using posters and announcements made during the lectures and tutorials. Individuals who required course credits were awarded 4 credits for the pilot study and 2 credits for the actual study.

C. Instruments.

Materials utilised in the study included an information sheet, an informed consent form, the Lecturer-Student Interaction (LSI) Questionnaire (Appendix 1), the Class-related Emotions Questionnaire (CEQ) (Appendix 2), Perception of Learning (POL) questionnaire (Appendix 3), and achievement measures (Appendix 4).

Manipulation Check Questions and Demographic Information. Demographic information requested from participants included their age and gender. In addition, two manipulation check questions “How do you usually feel when you attend this lecturer’s lesson?” and “What are some of the characteristics that you would associate with this lecturer?” were included to ensure that the experimental manipulation was successful (Refer to Appendix 1). In general, participants assigned to the *good* lecturer-student interaction condition were expected to report more positive than negative feelings when attending the lecturer’s lesson, and associate more positive than negative characteristics to the lecturer compared to those assigned to the *poor* lecturer-student interaction condition.

Lecturer-student Interaction (LSI) Questionnaire. The 9-item questionnaire, encompassing four aspects of lecturer-student interaction: autonomy, emotional, academic support and provision of structure was used to measure the quality of lecturer-student interaction (IV). Two versions of the questionnaire, one for each experimental condition (i.e., good and

poor) were used. Both questionnaires were identical in all aspects except for minor differences in the instructions. The term 'good' in the instructions "Think of one lecturer in James Cook University who has taught you last semester, with whom you perceive you have a good interaction (i.e. lecturer-student interaction)" was replaced with 'poor' to distinguish the two experimental conditions (Refer to Appendix 1). Two items were used to assess each emotional support, academic support and provision of structure; whereas three items, one of which is reverse-coded, was used to measure autonomy support.

Emotional support. Questions on emotional support were adapted from the Instrumental help subscale (Cronbach's $\alpha = .95$) of the Teacher-Student Relationship Inventory (Ang, 2005) and the Teacher-student relationship subscale (Cronbach's $\alpha = .88$) of the Student Engagement Instrument (SEI) (Appleton et al., 2006). An example of an item would be "My lecturer cares about me as a person and not just as a student".

Academic support. Items tapping into teacher provision of academic support were adapted from the Teacher Support Scale (Cronbach's $\alpha = .95$) by Metheny, McWhirter, and O'Neil (2008). An example of an item would be "My lecturer was willing to help me learn".

Autonomy Support and Provision of Structure. Questions measuring autonomy support and provision of structure were adapted from the Observer's Rating measure used by Reeve, Jang, Carrell, Jeon, and Barch (2004). An example of an item tapping into autonomy support would be "My lecturer was open to student discussions and opinions voiced by students" whereas an item measuring provision of structure would be "My lecturer directed the attention of students to important content (e.g. exam-related material) in the lecture".

All items were rated on a 5-point Likert scale with anchors 1 (*Strongly Disagree*) and 5 (*Strongly Agree*). An Exploratory Factor Analysis (EFA) was conducted with the results indicating items loading on a single factor. Hence, the questionnaire was used as a composite measure of the quality of lecturer-student interaction. The pattern and structure matrix for the one-factor solution of Lecturer-Student Interaction (LSI) Questionnaire is reflected in Appendix 5.

A total score was obtained by summing up the nine items on the scale (possible score range: 9 - 45) such that higher scores were indicative of higher quality lecturer-student interaction. Items in the adapted scale had good internal consistency with a Cronbach alpha coefficient of .89.

Class-related Emotions Questionnaire (CEQ). The twelve-item CEQ was adapted from the Class-related Emotions scale of the Achievement Emotions Questionnaire (AEQ) (Pekrun et al., 2002) and was used as a measure of students' emotional engagement (MV) during lecture. The questionnaire assessed six emotions (i.e., enjoyment, boredom, hopelessness, anger, hope, anxiety) frequently expressed in academic settings. Six of the items were reverse-coded. Items were categorized into emotions experienced before (5 items), during (5 items) and after class (2 items) to facilitate recall of emotions experienced at the instance when taking the module. An example of an item would be "I felt frustrated during the lecture". All items were rated on a 5-point Likert scale with anchors, 1 (*Strongly Disagree*) and 5 (*Strongly Agree*) and responses were summed to obtain a single emotional engagement score (possible range: 12 - 60). Higher scores were reflective of greater levels of emotional engagement whereas lower scores were indicative of lower levels of engagement. Both the original CEQ subscales (Cronbach's alpha range: .84 - .93) (Pekrun et al., 2002) and the adapted CEQ (Cronbach's alpha = .95) had good internal consistency.

Perception of Learning (POL) Questionnaire. A five-item Perception of Learning (POL) questionnaire was used to assess perceived learning (DV) by students during the lecture. Items were adapted from the Course Experience Questionnaire (McInnis, Griffin, James, & Coates, 2001) and the Online Learning Beliefs, Emotions, and Behaviors Survey (OLBEBS) (Artino, 2009). An example of an item assessing perceived learning is “My lecturer encouraged me to relate what I was learning in the module to what I already know”. All items were rated on a 5-point Likert scale with anchors 1 (*Strongly Disagree*) and 5 (*Strongly Agree*). An overall perception of learning score was obtained by summing up the individual items, with scores ranging from 5 to 25. Both the original (OLBEBS) and the adapted questionnaire had good internal consistency with a Cronbach alpha value of .88 (Artino, 2009) and .91 respectively.

Achievement measures. Two grades; one obtained in the module taught by the lecturer from *one* semester ago (i.e., Module grade) and the average grade obtained *two* semesters ago were used as measures of achievement. Both grades were obtained with respect to the current semester in which participants were completing the questionnaires. The module grade was used as the measure of the DV whereas the average grade was used as a control variable. Letter grades were converted to a 5-point Likert scale with anchors 1 (*N*) and 5 (*HD*)³.

D. Procedure.

Pilot study. A total of 20 psychology students participated in the pilot study. Participants were randomly assigned to either the ‘good’ ($n = 10$) or the ‘poor’ ($n = 10$) lecturer-student interaction condition. They were then provided with the questionnaires and asked to complete them in the following order: Information sheet, an informed consent form, the Lecturer-Student Interaction (LSI) questionnaire, the Class-related Emotions Questionnaire (CEQ), and achievement measures. The Perception of Learning (POL) questionnaire was not included as part of the pilot study. Instructions (as printed on the questionnaire) to the two groups were as follows: “Think of one lecturer in James Cook University who has taught you (last semester), with whom you perceive you have a GOOD/POOR interaction (i.e. lecturer-student interaction). Answer the following questions keeping this lecturer in mind.” Following this instruction, participants completed the manipulation check questions before proceeding to complete the remaining questionnaires. Participants were not required to identify the lecturer and were assured confidentiality of their responses. Upon completion of the questionnaires, participants were requested to provide their opinions regarding the clarity and comprehensiveness of the measures. Participants were then debriefed about the study and thanked for their participation. The pilot study session lasted for approximately 45 minutes.

Following the pilot study, reliability analyses were conducted on both the Lecturer-Student Interaction (LSI) questionnaire and the Class-related Emotions Questionnaire (CEQ). Both questionnaires had good internal consistency, with a Cronbach alpha value of .80 and .93 respectively. Minor changes were made to the wording and structure of the questions based on participants’ opinions for clearer comprehension. For example, the term ‘hopeless’ in one of the CEQ items “The thought of having to attend his/her lecture made me feel hopeless” was replaced by the synonym ‘discouraged.’ This change was warranted given that the majority of the pilot study participants perceived the term ‘hopeless’ to have a strong negative connotation. Similar changes were made to few other questionnaire items. One item from the Lecturer-Student

Interaction (LSI) questionnaire, “My lecturer tried to incorporate students’ interests into the lecture” was excluded and replaced by two others (Items 5 and 6 as reflected in Appendix 1). Modifications to the questionnaire items during the pilot study and justification for these changes are reflected in Appendix 6.

Research study. Procedures for the actual study differed slightly from the pilot in that participants either completed an online ($n = 95$) or a paper-and-pencil version of the questionnaire ($n = 25$). Conducting the study online was not expected to affect the results obtained given evidence for little difference in the accuracy of responses between online and paper-and-pencil methods (Bates & Cox, 2008; Bressani & Downs, 2002; Wharton, Hampl, & Winham, 2003). The online questionnaire elicited mandatory responses for informed consent before inviting the participants to complete the questionnaires. The duration of the actual study was approximately 20 minutes. The online questionnaire, hosted by Survey Gizmo can be viewed at the following link: (<http://edu.surveygizmo.com/s3/822881/Lecturer-Student-Interaction-and-Achievement-1>).

III. Data Analysis and Results.

A. Checking of assumptions.

Only data obtained from the actual study was used for further analysis. All analyses were performed using SPSS Version 18.0. Initial screening revealed no missing data. No serious violations of normality, homoscedasticity, linearity, and multicollinearity were observed. A review of scatterplots, histograms and boxplots revealed no extreme univariate outliers with standard deviations greater than ± 3.30 (Tabachnick & Fidell, 2007). One multivariate outlier (ID 53) was detected using Mahalanobis distance and was removed. Statistical analyses were performed on the remaining 119 ($M = 23$, $F = 96$) data sets.

B. Descriptives, manipulation check questions and correlations between variables.

A preliminary analysis was conducted on the manipulation check questions to gauge the success of the experimental manipulation (i.e., assignment to the ‘good’ and ‘poor’ lecturer-student interaction condition). A comparison of percentage values between the two experimental conditions found individuals who shared a good interaction with the lecturer (hereby referred to as ‘good’ interaction condition) to report more positive (e.g., inspired) than negative feelings (e.g., bored) when attending the lecture compared to those who shared a poor interaction (hereby referred to as ‘poor’ interaction condition) (Figure 1).

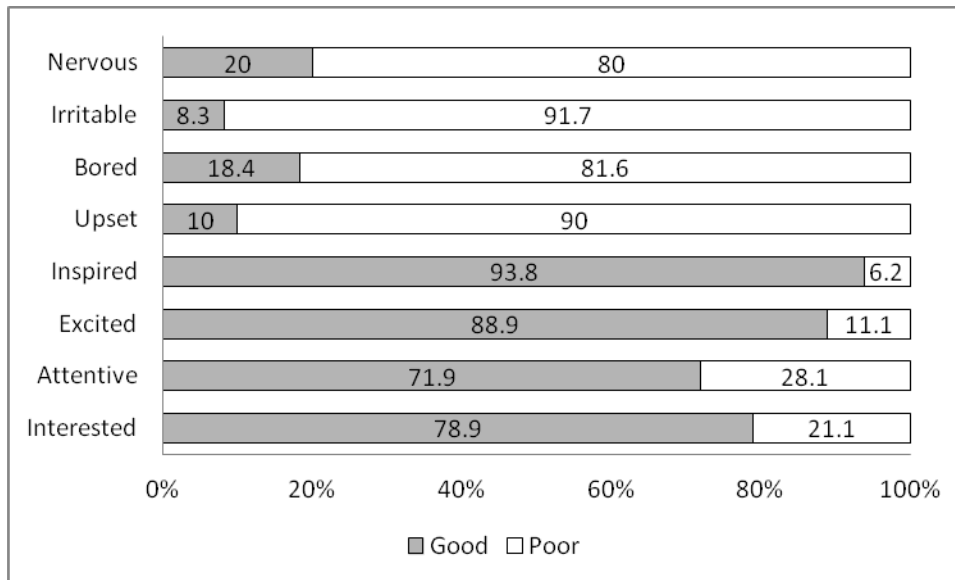


Figure 1. Percentage checked on manipulation questions (feelings) between ‘poor’ and ‘good’ interaction conditions.

Likewise, individuals in the ‘good’ interaction condition associated more positive (e.g., approachable) than negative (e.g., boring) characteristics to the lecturer compared to those in the ‘poor’ interaction condition (Figure 2), with the exception of the characteristics *demanding* and *strict*. Individuals in both the experimental conditions were equally likely to associate the characteristic ‘demanding’ with their lecturer whereas, those in the ‘good’ as opposed to the ‘poor’ interaction condition were more likely to associate the characteristic of ‘strict’ (56% vs. 44%) with their lecturer. Given these results, the manipulation check was deemed successful.

In addition, a correlation analysis conducted revealed predominantly moderate to strong relations between the variables; with the main variables under study: Lecturer-student interaction (LSI), perceived learning (POL), emotional engagement (CEQ), and achievement (module grade) being significantly correlated with one another. A summary of the correlations, means, and standard deviations of the variables used in the study are presented in Table 1 and Table 2 respectively.

C. Quality of lecturer-student interactions and emotional engagement scores.

An independent samples *t*-test was conducted to assess the mean difference in emotional engagement scores between individuals in the poor and good lecturer-student interaction conditions. Individuals in the ‘good’ interaction condition ($M = 43.28$, $SD = 8.91$) had significantly higher emotional engagement scores compared to individuals in the ‘poor’ interaction condition ($M = 31.86$, $SD = 9.38$); $t(117) = 6.81$, $p < .001$ (two-tailed), 95% CI [8.10, 14.74], $\eta^2 = .28$.

D. Quality of lecturer-student interaction and academic achievement.

Likewise, an independent samples *t*-test was also performed to compare the academic achievement of individuals between the ‘good’ and ‘poor’ interaction conditions. Individuals

who shared a good interaction with their lecturer ($M = 3.33, SD = .86$) did not significantly differ in terms of academic achievement (i.e., grades obtained in the module taught by the lecturer) compared to those who shared a poor interaction with their lecturer ($M = 3.03, SD = .93$); $t(117) = 1.83, p = .07$ (two-tailed), 95% CI [-.03, .62].

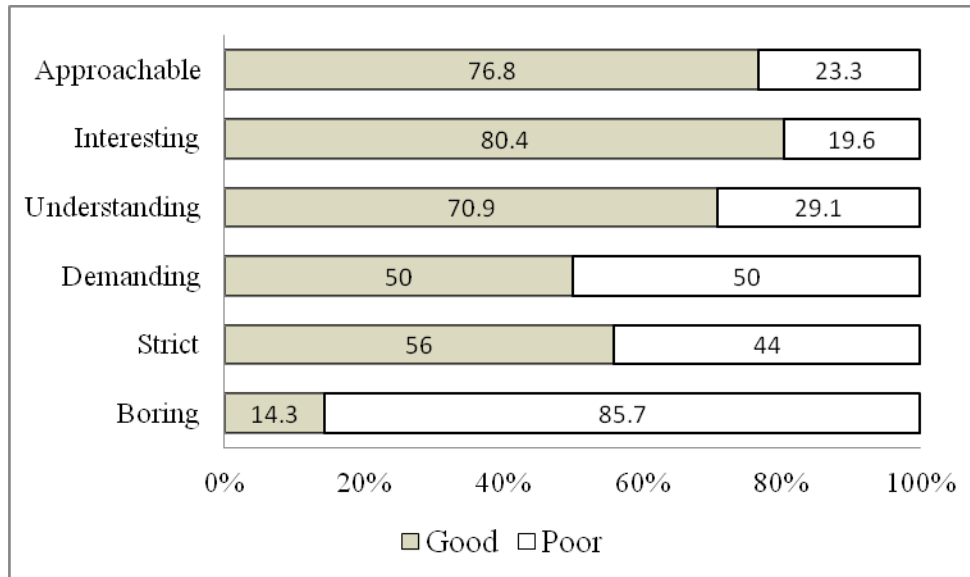


Figure 2. Percentage checked on manipulation questions (characteristics) between ‘poor’ and ‘good’ interaction conditions.

E. Role of emotional engagement between lecturer-student interaction and academic achievement.

A mediation analysis was conducted to test the viability of emotional engagement (MV) as a mediator of lecturer-student interaction (IV) and academic achievement of individuals (DV). In line with Baron and Kenny’s (1986) 4-step approach; for mediation to occur, the (1) IV should significantly predict the DV, (2) the IV should significantly predict the MV, (3) the MV should significantly predict the DV when the IV is controlled for and (4) the effect of the IV on the DV should reduce (partial mediation) or become insignificant (full mediation) when the MV is included (Baron & Kenny, 1986; Preacher & Hayes, 2004). Three separate multiple regression analyses were conducted to test these pathways, with the effects of average grade and age being controlled in every step, given that these variables significantly correlated with the outcome (i.e., academic achievement) (Tucker et al., 2002).

As illustrated in Figure 3, (1) the quality of lecturer-student interaction (IV) significantly predicted academic achievement of individuals (DV); $\beta = .34, p < .001$, and (2) the quality of lecturer-student interaction significantly predicted emotional engagement; $\beta = .80, p < .001$. However, (3) when controlled for the IV, emotional engagement failed to predict academic achievement of individuals; $\beta = .14, p = .16$ (ns). Step 3 of Baron and Kenny’s (1986) 4-step approach was not fulfilled and thus, emotional engagement was found to not mediate the relationship between lecturer-student interaction and academic achievement.

Table 1. Summary of intercorrelations between LSI, CEQ, POL, age, gender, module grade and average grade.

Variables	1	2	3	4	5	6	7	8	9	10	11
1. LSI	-	.72**	.90**	.90**	.82**	.80**	.78**	.40**	.09	.03	.01
2. Emotional Support	.72**	-	.56**	.53**	.38**	.53**	.52**	.21**	.06	.04	.06
3. Academic Support	.90**	.56**	-	.77**	.72**	.67**	.67**	.38**	.06	-.05	.01
4. Autonomy Support	.90**	.53**	.77**	-	.66**	.68**	.71**	.42**	.11	.10	-.04
5. Structure	.82**	.38**	.71**	.66**	-	.71**	.73**	.30**	.08	-.05	.02
6. CEQ	.80**	.53**	.67**	.68**	.71**	-	.80**	.37**	.08	.01	-.001
7. POL	.78**	.52**	.68**	.71**	.73**	.80**	-	.33**	.06	-.01	-.11
8. Module Grade	.40**	.21**	.38**	.42**	.30**	.37**	.33**	-	.70**	.05	.18*
9. Average Grade	.09	.06	.06	.11	.08	.08	.06	.70	-	.01	.17
10. Gender	.03	.04	-.05	.10	-.05	.01	-.01	.05	.01	-	-.05
11. Age	.01	.06	.01	-.04	.01	-.001	-.11	.18*	.17	.05	-

* $p < .05$, ** $p < .001$, two-tailed.

Table 2. Means and standard deviations of variables.

Variables	Good (<i>n</i> = 60)		Poor (<i>n</i> = 59)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
LSI Total	34.05	4.99	27.25	5.85
Emotional Support	5.88	1.62	4.75	1.64
Academic Support	8.17	1.17	6.73	1.64
Autonomy Support	11.92	1.82	9.73	2.13
Structure	8.25	1.34	6.05	1.93
CEQ	43.28	8.91	31.86	9.38
Perception of Learning	19.25	2.61	14.24	4.06

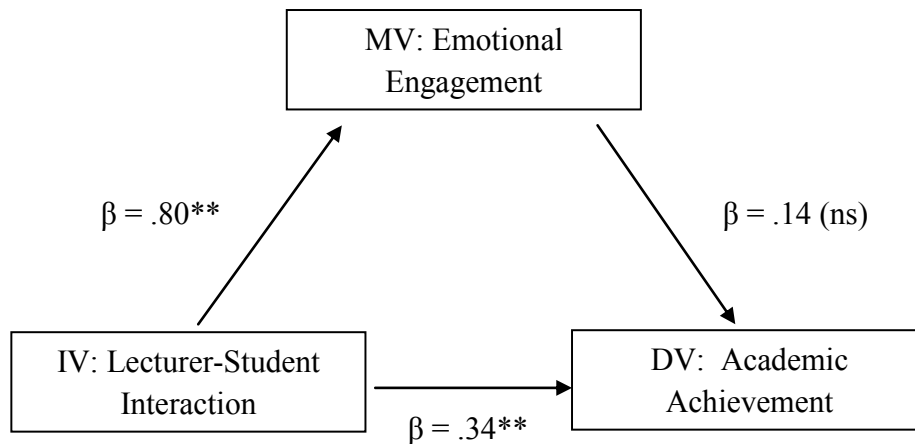


Figure 3. Emotional engagement as a mediator between lecturer-student interaction and academic achievement of individuals. Coefficients reflected are standardized beta coefficients. * $p < .05$, ** $p < .001$, two-tailed.

F. Role of emotional engagement between lecturer-student interaction and perceived learning.

A second mediation analysis was conducted to examine emotional engagement as a mediator between lecturer-student interaction and perceived learning by students. Consistent with Baron and Kenny's (1986) 4-step approach, (1) the quality of lecturer-student interaction (IV) significantly predicted perceived learning (DV), $\beta = .78$, $p < .001$; (2) the quality of lecturer-student interaction significantly predicted emotional engagement of individuals (MV), $\beta = .80$, $p < .001$; (3) when controlled for the IV, emotional engagement significantly predicted perceived learning, $\beta = .48$, $p < .001$ and lastly, (4) the inclusion of the MV in the model weakened the effect of lecturer-student interaction on perceived learning, $\beta = .40$, $p < .001$ (Figure 4). Hence, emotional engagement was found to partially mediate the pathway between lecturer-student interaction and perceived learning by individuals. In addition, a bootstrap estimate of indirect effect was conducted at 95% confidence interval on 5000 bootstrap samples

(Preacher & Hayes, 2004). Since zero was not in the 95% CI [.13, .35], lecturer-student interaction was also found to have a significant indirect effect on perceived learning through emotional engagement (Preacher & Hayes, 2004).

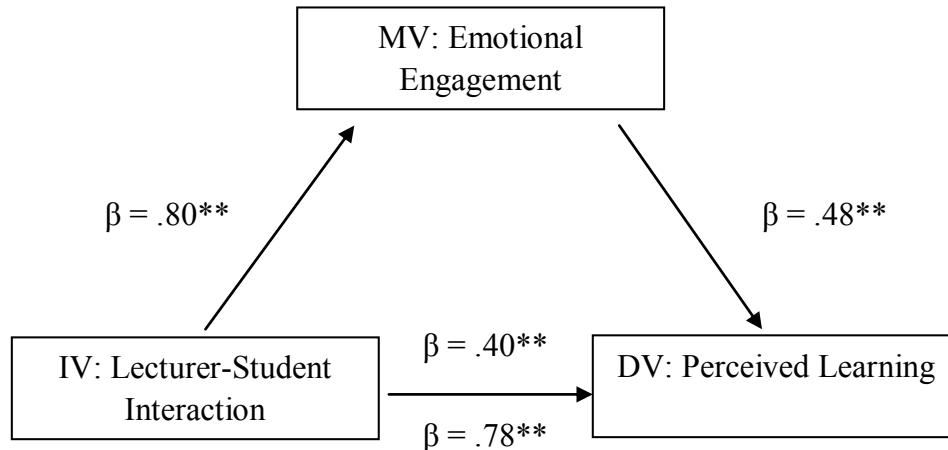


Figure 4. Emotional engagement as a mediator between lecturer-student interaction and perceived learning of individuals. Coefficients reflected are standardized beta coefficients.

* $p < .05$, ** $p < .001$, two-tailed.

IV. Discussion.

The current study investigated how the quality of lecturer-student interaction (i.e., good vs. poor) influenced the level of emotional engagement of students during lectures and how that in turn affects student achievement (i.e., module grade) and learning in the module taught by the lecturer.

A. Findings that address hypothesis one.

In line with hypothesis one, students who shared a good interaction with their lecturer reported significantly higher levels of emotional engagement during the lecture compared to those who shared a poor interaction. This finding offered some support for the Self-System Model of Motivational Development (Skinner et al., 2008). According to the model, features of a context (i.e., characteristics of lecturer-student interaction in the current study) are posited to influence the three basic psychological needs of individuals (i.e., need for competence, relatedness, autonomy), which then serves to predict engagement of individuals (Skinner et al., 2008). Although the lecturer-student interaction measure used in the current study was a global one, it encompassed the four aspects of emotional, academic, autonomy support, and structure. Drawing upon the findings of Hughes and Kwok (2007), it is likely that individuals in the ‘good’ interaction condition received higher levels of support (as characterized by the aforementioned four aspects) in comparison to those in the ‘poor’ interaction condition. Furthermore, past studies have shown each of these four aspects play a role in fulfilling the individual needs for relatedness, autonomy, and competence (Jang et al., 2010; Skinner et al., 2008). As such, in line with the

SSMMD, these higher levels of support experienced by individuals in the ‘good’ interaction condition are likely to have satisfied the three basic needs of individuals to a greater extent, which then could have accounted for the higher levels of emotional engagement reported by these individuals.

B. Findings that address hypothesis two.

Contrary to hypothesis two, students who shared a good interaction with their lecturer did not differ in terms of achievement from those who shared a poor interaction. This finding was at odds with past studies such as Lam et al. (2012) and Birch and Ladd (1997) that have shown a supportive, non-conflictual relationship with teachers to promote academic achievement among individuals. A few reasons may account for this finding. Past research has predominantly focused on a younger student population (i.e., elementary, middle, and high school) among whom teacher-student interaction has been found to be crucial for achievement and school adjustment (Birch & Ladd, 1997). However, the current study examined the role of lecturer-student interaction among university students. Hence, it is possible that this factor is not as critical for achievement among this population given that contact time with teachers in university settings is much lesser compared to students in middle and high schools. Instead, given that these students are expected to engage in more independent learning, this latter factor could also be accounting for a substantial portion of variance in achievement.

Second, the grade obtained in the respective module taught by the lecturer is used as a measure of achievement in the present study. Given that tutors, as opposed to lecturers, account for a larger proportion of a given grade for a module, it is likely that interaction with the lecturer may not be significantly reflected in student grades. Rather, the effect of lecturer-student interaction on academic achievement might be moderated by an interaction with the tutor.

C. Findings that address hypothesis three.

With respect to hypothesis three, emotional engagement failed to mediate the pathway between lecturer-student interaction and academic achievement of individuals but was a partial mediator between lecturer-student interaction and perceived learning. In particular, emotional engagement failed to predict academic achievement of individuals. Past studies have defined the emotional subtype in several ways (e.g., identification, school liking) (Fredricks et al., 2004; Ladd & Dinella, 2009). Furthermore, the role of emotional subtype pertaining to academic outcomes in past studies is confounded due to the inclusion of behavioral and cognitive subtypes (Fredricks et al., 2004).

Given that the current study used a ‘pure’ measure of emotional engagement (defined as affective reactions expressed by students within academic settings), the non significant outcome between emotional engagement and academic achievement can be viewed from the perspective that by adopting a clear definition of the concept, greater insight has emerged into the significant impact of emotional engagement on student learning. This finding helps deepen current understandings, which can better inform the pedagogical practice of faculty members, such that they impact student outcomes.

An alternate explanation to that proposed above, could be that emotional engagement failed to predict academic achievement due to this subtype having a more central role in predicting other academic outcomes. In line with this reasoning, emotional engagement was

found to significantly predict perceived learning by individuals. This latter finding that emotional engagement plays a more influential role in learning as opposed to academic achievement corroborates with the findings of Handelsman et al. (2005), whereby this subtype emerged as a significant predictor of intrinsic learning outcomes (e.g., developing a positive attitude towards learning) but not as a predictor of academic achievement.

D. Contributions of findings that enhance current understandings of emotional engagement.

Findings from the current study thus contribute to existing research in two main ways. First, examining emotional engagement as an attribute on its own rather than as part of a general measure of engagement provides some insight into the role of this subtype's impact (or lack of it) on different academic outcomes (i.e. no significant impact on academic achievement, but significant impact on student learning outcomes). Second, majority of the studies that investigate the concept of engagement have focused on achievement measures (Appleton et al., 2006; Fredricks et al., 2004). While achievement might be an important outcome in itself, it may not be a good indicator of engagement given that individuals who have high achievement may not necessarily be more engaged (Zyngier, 2008). Hence, showing that emotional engagement has a significant role in predicting learning as opposed to academic achievement suggests the possibility that the various subtypes may have differential roles in influencing academic outcomes. Thus, promoting emotional engagement (e.g., interest) of students through varying instructional methods (e.g., using crossword puzzles to test understanding of subject content) (Raines, 2010) for instance, might be useful in creating a learning environment that promotes a positive attitude and openness towards learning (Jang, 2008) which in turn may facilitate cognitive (e.g., exertion of mental effort) and behavioral aspects (e.g., participation in class) of engagement. Given evidence from past studies for behavioral and cognitive engagement as strong predictors of achievement (e.g., Ladd & Dinella, 2009), this then is likely to contribute to academic achievement of individuals (i.e., good grades).

E. Contributions of findings that enhance current understanding of the role of teacher-student interactions in relation to engagement and academic outcomes.

Past studies have mainly studied teacher-student interactions among middle and high school students given their importance in influencing achievement, well-being, and school adjustment (e.g., Barkatsas et al., 2009) among this population. However, the current study was able to provide some evidence for the importance of teacher-student interaction in influencing level of engagement in a university population. Furthermore, showing that interactions with the lecturer has an indirect effect on learning through emotional engagement as opposed to academic achievement suggests that at higher levels of education, lecturers might have a more pivotal role in influencing student learning rather than academic achievement.

F. Limitations of the current study and future directions.

In applying these findings however, it is important to take into consideration some of the limitations in the present study. Most of the scales utilised in this study (Ang, 2005; Appleton et al., 2006; Metheny et al., 2008; Reeve et al., 2004) were validated on middle and high school students, given that this has been the main population under study. Although this brings the

validity of these items for university students into question, the pilot study attempted to address this limitation by ensuring applicability of the questionnaire items for this population.

Secondly, while mediation suggested some evidence for the directional influence from lecturer-student interaction to emotional engagement to academic outcomes, the cross-sectional design of the current study does not allow for firm conclusions. Furthermore, Furrer and Skinner (2003) among others, have suggested reciprocal relationships between teacher-student interaction and engagement such that individuals who participate actively (behavioral engagement) and display more positive emotions (emotional engagement) are likely to receive more support from teachers. Given this, future studies can look into exploring these reciprocal relationships using a longitudinal design (Fredricks et al., 2004).

G. Conclusion.

Past studies on student engagement have primarily focused on the behavioral and cognitive subtypes of engagement given their established roles in influencing academic achievement of individuals. Emotional engagement on the other hand, has received less research attention due to its associated lack of conceptual clarity. Furthermore, the tendency to examine this subtype as part of a general measure has yielded mixed evidence with respect to its role in influencing academic outcomes. The current study attempted to address these limitations by examining emotional engagement as a 'pure' construct in relation to the separate academic outcomes of learning and achievement. In addition, the present study also investigated the characteristics of teacher-student interactions in relation to engagement and academic outcomes of tertiary students. While current findings build on existing literature in demonstrating the crucial role of lecturers in influencing the aforementioned outcomes in tertiary level students, further research is recommended to validate these findings in this particular population.

In conclusion, this study not only offers new insight regarding the importance of emotional engagement on academic outcomes such as student learning, but it also provides insights for staff on the significance of their interactions on student learning and provides a platform for incorporating in their pedagogical practices aspects of emotional, academic, autonomy support and provision of structure for enhancing emotional engagement and maximising learning. In encouraging faculty to do so, the university needs to ensure that the evaluation instruments given to students for assessing faculty on their pedagogical practice should accurately reflect the quality of the staff in addressing and achieving emotional engagement.

Appendix

Appendix 1: Lecturer-Student Interaction Questionnaire and Manipulation Check Questions

Think of 1 *lecturer* in James Cook University who has taught you in SP52 (2011), with whom you perceive you have a **GOOD/POOR** interaction* (i.e. lecturer-student interaction). Answer the following questions keeping this lecturer in mind. Please note that you will NOT be required to identify the lecturer. All information will be kept CONFIDENTIAL.

Manipulation Check Questions

1. How do you usually *feel* when you attend this lecturer's lesson? (Tick all that are applicable)

- | | |
|-------------------------------------|---|
| <input type="checkbox"/> Interested | <input type="checkbox"/> Bored |
| <input type="checkbox"/> Attentive | <input type="checkbox"/> Irritable |
| <input type="checkbox"/> Excited | <input type="checkbox"/> Nervous |
| <input type="checkbox"/> Inspired | <input type="checkbox"/> Others (Please state): |
| <input type="checkbox"/> Upset | _____ |

2. What are some of the characteristics that you would associate with this lecturer? (Tick all that are applicable)

- | | |
|--|---|
| <input type="checkbox"/> Understanding | <input type="checkbox"/> Demanding |
| <input type="checkbox"/> Boring | <input type="checkbox"/> Interesting |
| <input type="checkbox"/> Strict | <input type="checkbox"/> Others (Please state): |
| <input type="checkbox"/> Approachable | _____ |

Lecturer-Student Interaction (LSI) Questionnaire

Below are some statements that might be descriptive of your lecturer. With reference to the **same** lecturer, circle the option that is most descriptive of him/her.

1. My lecturer cares about me as a person and not just as a student. (Emotional Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

2. If the need arose, I would be comfortable with sharing personal matters (e.g. family problems) with my lecturer. (Emotional Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

3. My lecturer was available if I needed to clarify my doubts about the lecture. (Academic Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

4. My lecturer was willing to help me learn. (Academic Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

5. My lecturer initiated interaction and discussions that helped me learn the subject material. (Autonomy Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

6. My lecturer was open to student discussions and opinions voiced by students. (Autonomy Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

7. My lecturer's teaching style can be described as rigid (i.e. strict) and controlling. (Autonomy Support)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

8. My lecturer had a clear goal of what he/she wanted to achieve during the lecture. (Structure)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

9. My lecturer directed the attention of students to important content (e.g. exam-related material) in the lecture.(Structure)

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Appendix 2: Class-related Emotions Questionnaire (CEQ)

Answer the following questions keeping in mind the **same lecturer** whom you thought of previously. The statements below describe feelings that you may have experienced **before**, **during** and **after** a lecture taught by this lecturer. All information will be kept CONFIDENTIAL.

Before Class

The following questions pertain to feelings you may have experienced **BEFORE** attending a lecture taught by this lecturer. Please circle the most appropriate option.

1. The thought of attending lecture made me lethargic.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

2. I looked forward to learning during his/her lecture.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

3. I felt motivated to attend his/her lecture.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

4. Even before the lecture, I used to worry whether I will be able to understand the material.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

5. The felt discouraged at the thought of having to attend his/her lecture.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

During Class

The following questions pertain to feelings you may have experienced **DURING** the lecture. Please circle the most appropriate option.

1. I enjoyed listening to his/her lecture.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

2. I was tempted to walk out of the lecture because it was boring.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

3. I felt frustrated during the lecture.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

4. I was focused during his/her lecture that I did not realise time passing.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

5. I felt bored during lecture and therefore I had problems staying alert.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

After Class

The following questions pertain to feelings you may have experienced **AFTER** the lecture. Please circle the most appropriate option.

1. I looked forward to the next lecture (of the same module) when the lecture was over.
Strongly Disagree Disagree Neutral Agree Strongly Agree
1 2 3 4 5

2. I felt glad that I had attended the lecture.
Strongly Disagree Disagree Neutral Agree Strongly Agree
1 2 3 4 5

Appendix 3: Perception of Learning (POL) Questionnaire

Please also answer the following questions. Circle the most appropriate option.

1. My lecturer was able to stimulate my interest in the module.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

2. My lecturer explained the subject material such that it had practical value for me (i.e. can relate to everyday experiences).

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

3. My lecturer encouraged me to apply the skills that I have learnt in the course to other modules that I was taking.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

4. My lecturer encouraged me to relate what I was learning in the module to what I already know.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

5. My lecturer encouraged me to develop my own academic interests as far as possible.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

Appendix 4: Academic Achievement Measures

Please circle the FINAL grade that you obtained in the *module* taught by this lecturer (If unable to remember, please log on to e-student to access your grade). Note that the experimenter will NOT have access to any of your grades.

N P C D HD

Please also circle the *AVERAGE* grade that you obtained 2 semesters ago. (If unable to remember, please log on to e-student to access your grade). Note that the experimenter will NOT have access to any of your grades.

N P C D HD

Appendix 5: The Pattern and Structure Matrix for One-Factor Solution of Lecturer-Student Interaction (LSI) Questionnaire

Item	Pattern		Structure		Communalities
	Coefficients		Coefficients		
	Component 1	Component 2	Component 1	Component 2	
8. Struc 1	.991		.891		.832
9. Struc 2	.906		.854		.740
5. Aut Spt 1	.805		.859	.485	.748
4. Acad Spt 2	.783		.870	.549	.786
3. Acad Spt 1	.650		.780	.583	.674
6. Aut Spt 2	.645		.765	.557	.640
2. Emo Spt 2		.840		.797	.642
7. Aut Spt 3		.650	.361	.679	.465
1. Emo Spt 1		.621	.553	.744	.611

Note. Major loadings for each items are bolded.

Appendix 6: Modifications to Pilot Study Questionnaires The Lecturer-Student Interaction

Emotional Support.

Pilot: My lecturer was ~~interested in me~~ as a person and not just as a student.

Actual: My lecturer (cares about me) as a person and not just as a student. (Qn 1)

Change: The term 'interested in me' was changed to 'cares about me' as the former was considered misleading.

Academic Support

Pilot: My lecturer was ~~not~~ available if I needed to clarify my doubts about the lecture.

Actual: My lecturer (was available) when I needed to clarify my doubts about the lecture (Qn 3)

Change: To clarify understanding.

Autonomy Support

Pilot: ~~My lecturer tried to incorporate students' interests into the lecture.~~

Actual: My lecturer initiated interaction and discussions that helped me learn the subject material. (Qn 5)

Actual: My lecturer was open to student discussions and opinions voiced by students. (Qn 6)

Change: Item was replaced by two new ones to be more specific as participants perceived the term 'students' interests' to be too broad.

Pilot: My lecturer's teaching style can be described as ~~rigid~~ and controlling.

Actual: My lecturer's teaching style can be described as (rigid (i.e. strict)) and controlling. (Qn 7)

Change: To improve clarity of the word 'rigid'.

Provision of Structure

Pilot: My lecturer was ~~organized and~~ had a clear goal of what he/she wanted to achieve during the lecture.

Actual: My lecturer (had a clear goal) of what he/she wanted to achieve during the lecture. (Qn 8)

Change: Item was double-barreled'.

Pilot: My lecturer directed the attention of students to important content in the lecture.

Actual: My lecturer directed the attention of students to important content (e.g. exam-related material) in the lecture. (Qn 9)

Change: To improve clarity of the term 'important content'.

Classroom-related Emotions Questionnaire

Before Class.

Pilot: The thought of attending lecture made me ~~nervous~~.

Actual: The thought of attending the lecture made me feel (lethargic). (Qn 1)

Change: The word 'nervous' was not applicable in lecture settings for most students.

Pilot: I looked forward to learning ~~a lot~~ during his/her lecture.

Actual: I looked forward to learning during his/her lecture. (Qn 2)

Pilot: The thought of having to attend his/her lecture made me feel ~~hopeless~~.

Actual: I felt (discouraged) at the thought of having to attend his/her lecture.

Change: 'Hopeless' perceived to have a strong negative connotation. Replaced by synonym.

During Class.

Pilot: I was tempted to walk out of the lecture because it was ~~so~~ boring.

Actual: I was tempted to walk out of the lecture because it was boring. (Qn 2)

Pilot: I was ~~engrossed~~ in his/her lecture that I ~~do not~~ realise time passing.

Actual: I was (focused) during his/her lecture that (I did) not realize time passing. (Qn 4)

Change: The term 'engrossed' was considered 'too strong' and thus replaced.

After Class.

Pilot: I looked forward to the next lecture when the lecture was over.

Actual: I looked forward to the next lecture (of the same module) when the lecture was over. (Qn 1)

Change: To clarify understanding.

References

- Ainley, M. (2004, November). *What do we know about student motivation and engagement?* Paper presented at the annual meeting of the Australian Association for Research in Education, Melbourne.
- Ainley, M., and Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*, 4-12. doi:10.1016/j.cedpsych.2010.08.001
- Ang, R.P. (2005). Development and validation of the Teacher-Student Relationship Inventory: Using exploratory and confirmatory factor analysis. *The Journal of Experimental Education, 74*(1), 55-73.
- Appleton, J.J., Christenson, S.L., Kim, D., and Reschly, A.L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*, 427-445. doi:10.1016/j.jsp.2006.04.002
- Artino, A.R. (2009). Online learning: Are subjective perceptions of instructional context related to academic success? *Internet and Higher Education, 12*, 117-125.
- Barkatsas, A., Kasimatis, K., and Gialamas, V. (2009). Learning secondary mathematics with technology: Exploring the complex interrelationship between students' attitudes, engagement, gender and achievement. *Computers & Education, 52*, 562-570. doi:10.1016/j.compedu.2008.11.001
- Baron, R.M., and Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.
- Bates, S.C., and Cox, J. (2008). The impact of computer versus paper pencil survey, and individual versus group administration, on self reports of sensitive behaviors. *Computers in Human Behavior, 24*, 903-916.
- Birch, S.H., and Ladd, G.W. (1997). The teacher-child relationship and children's early school adjustment. *Journal of School Psychology, 35*(1), 61-79.
- Bush, R.N. (1942). A study of student-teacher relationships. *The Journal of Educational Research, 35*(9), 645-656.
- Bressani, R.V., and Downs, A.C. (2002). Youth independent living assessment: Testing the equivalence of web and paper/pencil versions of the Ansell-Casey Life Skills Assessment. *Computers in Human Behavior, 18*(4), 453-464.
- Chapman, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation, 8*(13). Retrieved from <http://PAREonline.net/>

Connell, J.P., Spencer, M.B., and Aber, J.L. (1994). Educational risk and resilience in African-American youth: Context, self, action, and outcomes in School. *Child Development*, 65(2), 493-506.

Finn, J.D. (1989). Withdrawing from school. *Review of Educational Research*, 59(2), 117-142.

Finn, J.D., and Voelkl, K.E. (1993). School characteristics related to student engagement. *Journal of Negro Education*, 62(3), 249-268.

Fredricks, J.A., Blumenfeld, P.C., and Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.

Fraser, B.J., and Fisher, D.L. (1982). Predicting students' outcomes from their perceptions of classroom psychosocial environment. *American Educational Research Journal*, 19(4), 498-518.

Frenzel, A.C., Thrash, T.M., Pekrun, R., and Goetz, T. (2007). Achievement emotions in Germany and China: A cross-cultural validation of the Academic Emotions Questionnaire-Mathematics (AEQ-M). *Journal of Cross-Cultural Psychology*, 38(3), 302-309. doi: 10.1177/0022022107300276

Furlong, M.J., Whipple, A.D., Jean, G.S., Simental, J., Soliz, A., and Punthuna, S. (2003). Multiple contexts of school engagement: Moving toward a unifying framework for educational research and practice. *The California School Psychologist*, 8, 99-113.

Furrer, C., and Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148-162. doi: 10.1037/0022-0663.95.1.148

Handelsman, M.M., Briggs, W.L., Sullivan, N., and Towler, A. (2005). A measure of college student course engagement. *Journal of Educational Research*, 98(3), 184-191.

Hughes, J.N., and Kwok, O.M. (2007). Influence of student-teacher and parent-teacher relationships on lower achieving readers' engagement and achievement in the primary grades. *Journal of Educational Psychology*, 99(1), 39-51. doi: 10.1037/0022-0663.99.1.39

Jang, H. (2008). Supporting students' motivation, engagement, and learning during an uninteresting activity. *Journal of Educational Psychology*, 100(4), 798-811. doi: 10.1037/a0012841

Jang, H., Reeve, J., and Deci, E.L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, 102(3), 588-600. doi: 10.1037/a0019682

Jimerson, S.J., Campos, E., and Grief, J.L. (2003). Toward an understanding of definitions and measures of school engagement and related terms. *The California School Psychologist*, 8, 7-27.

Klem, A.M., and Connell, J.M. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, 74 (7), 262-273.

Ladd, G.W., Birch, S.H., and Buhs, E.S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development*, 70 (6), 1373-1400.

Ladd, G.W., and Dinella, L.M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? *Journal of Educational Psychology*, 101(1), 190-206. doi: 10.1037/a0013153

Lam, S.F., Jimerson, S., Kikas, E., Cefai, C., Veiga, F.H., Nelson, B., ... Zollneritsch, J. (2012). Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. *Journal of School Psychology*, 50, 77-94. doi:10.1016/j.jsp.2011.07.004

McInnis, C., Griffin, P., James, R., and Coates, H. (2001). Development of the Course Experience Questionnaire (CEQ). Retrieved from http://www.dest.gov.au/archive/highered/eippubs/eip01_1/01_1.pdf

McLean, M. (2001). Can we relate conceptions of learning to student academic achievement? *Teaching in Higher Education*, 6, 399-413.

Metheny, J., McWcWhirter, E.H., and O' Neil, M.E. (2008). Measuring perceived teacher support and its influence on adolescent career development. *Journal of Career Assessment*, 16, 218-237.

Miller, R.B., Greene, B.A., Montalvo, G.P., Ravindran, B., and Nichols, J.D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, 21, 388-422.

Patrick, H., Ryan, A.M., and Kaplan, A. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology*, 99(1), 83-98. doi: 10.1037/0022-0663.99.1.83

Pekrun, R., Goetz, T., Frenzel, A.C., Barchfeld, P., and Perry, R.P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36, 36-48. doi:10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., and Perry, R.P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91-105.

Preacher, K.J., and Hayes, A.F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717-731.

Raines, D.A. (2010). An innovation to facilitate student engagement and learning: Crossword puzzles in the classroom. *Teaching and Learning in Nursing*, 5, 85-90. doi:10.1016/j.teln.2008.08.003

Reddy, R., Rhodes, J.E., and Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and Psychopathology*, 15, 119-38. doi: 10.1017.S0954579403000075

Reeve, J., Jang, H., Carrell, D., Jeon, S., and Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, 28(2), 147-169.

Reeve, J., and Tseng, C.M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267. doi:10.1016/j.cedpsych.2011.05.002

Ryan, R.M., and Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. doi: 10.1037/110003-066X.55.1.68

Sinclair, M.F., Christenson, S.L., Lehr, C.A., and Anderson, A.R. (2003). Facilitating student engagement: Lessons learned from Check & Connect longitudinal studies. *The California School Psychologist*, 8, 29-41.

Skinner, E.A., and Belmont, M.J.(1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571-581.

Skinner, E., Furrer, C., Marchand, G., and Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology*, 100(4), 765-781. doi: 10.1037/a0012840

Skinner, E.A., Wellborn, J.G., and Connell, J.P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology*, 82(1), 22-32.

Tabachnick, B.G., and Fidell, L.S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Pearson Education.

Trickett, E.J., and Moos, R.H. (1973). Social environment of junior high and high school classrooms. *Journal of Educational Psychology*, 65(1), 93-102.

Tucker, C. M., Zayco, R. A., Herman, K. C., Reinke, W. M., Trujillo, M., Carraway, K.,... Ivery, P. D. (2002). Teacher and child variable as predictors of academic engagement among low-income African American children. *Psychology in the Schools*, 39(4), 477-488. doi: 10.1002/pits.10038

Wharton, C.M., Hampl, J.S., Hall, R., and Winham, D.M. (2003). PCs or paper-and pencil: Online surveys for data collection. *Journal of the American Dietetic Association*, 103(11), 1458-1460. doi: 10.1016/j.jada.2003.09.004

Wigfield, A., and Eccles, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81. doi:10.1006/ceps.1999.1015

Zyngier, D. (2008). (Re)conceptualising student engagement: Doing education not doing time. *Teaching and Teacher Education*, 24, 1765-1776. doi:10.1016/j.tate.2007.09.004

Analyzing a college course that adheres to the Universal Design for Learning (UDL) framework

Frances G. Smith¹

Abstract: Universal design for learning (UDL) offers an educational framework for a college instructor that can maximize the design and delivery of course instruction by emphasizing multiple representations of materials, varied means for student expression, content and knowledge, and multiple ways to motivate and engage student learning. Through a UDL lens, learner variability is anticipated and considered as a strength in the instructional planning process. The present study examined the reflective practice of one faculty member as she applied the UDL framework to her graduate class. Study participants were engaged in action research that both explored the faculty's use of the UDL framework to design and deliver an introductory graduate research methods course and, student perspectives of the application of this approach. Both faculty and student responses were favorable towards the implementation of the UDL instructional practice. Results suggest that when faculty use the UDL framework to help design courses, goals are more clearly aligned with instructional practices; there is a positive relationship to student interest and engagement; and students are positively engaged in the course.

Keywords: Universal design for learning, teaching and learning, UDL, research methods, postsecondary education, universal design, higher education

I. Introduction.

Universal design for learning (UDL) offers a framework for a college instructor that can expand opportunities in the delivery of course instruction. Utilizing multiple formats, varied instructional methods, and flexible features of digital technologies, UDL can enhance learning experiences for all students. Research from the learning sciences supports the fact that learner variability is developmental, systematic, and context-dependent (Fischer & Bidell, 2006; Rose & Fischer, 2009; Rose & Gravel, 2010). This means that college instructors can expect to have a range of learners in classes that vary across multiple dimensions including their background preparation, their learning situated in context of the class; and their learning based on age and development (Rose & Fischer, 2009).

UDL applied to teaching and learning provides a lens that focuses targeted approaches on supporting student's affective, strategic and recognition learning networks (Rose & Meyer, 2002; Rose & Meyer, 2006; Rose & Gravel, 2010). The UDL framework places the burden to adapt on the curriculum rather than the learner (Rose & Meyer, 2006). Instructors can improve educational outcomes for a range of learners by considering the three central principles of UDL in the design of instructional goals, methods, classroom materials and assessments (Hitchcock & Stahl, 2003; Rose & Strangman, 2007).

¹ 2011-12 UDL Postdoctoral Fellow, CAST & Lynch School of Education, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, frances.smith@bc.edu

Research on how people learn (Bransford, Brown, & Cocking, 1999, 2000; Bransford, Vye, Stevens, Kuhl, Schwartz, Bell, Meltzoff et al. 2006) and the learning brain (Raz & Buhle, 2006; Rose, 2001, 2005; Rose & Meyer, 2002) verifies individual differences in approaches to learning. Three key findings have been determined to promote student learning (Bransford et al. 2000; 2006). First, students come to a classroom with preconceived ideas about how knowledge works and their initial understanding needs to be engaged. Learning transfer is heightened or hampered by the orientation of this prior knowledge. Likewise, students of differing cultural orientations may have difficulty with some school practices that conflict with those of their current community.

Second, to develop competence in an area, students need a deep understanding of the context and facts (Bransford et al. 2000). Distinctions are evident between expert and novice learners. Experts are able to notice, organize, and interpret information more successfully than novices. Experts have developed the skills to quickly recognize patterns in information and organize knowledge around key concepts. Scaffolding instruction through means that can highlight critical features, key questions and big ideas can be invaluable towards teaching expertise in novice learners (Rose & Gravel, 2010).

Finally, the third finding involves the need for students to develop a metacognitive approach to learning so that they can self-assess, understand, and appreciate their strengths and differences. Instruction and assessment situations that offer frequent feedback, assist students in self-discovery of their strengths, and encourage their ownership in their learning process encourage metacognition (Bransford et al. 2000; 2006). Providing multiple instructional approaches facilitates the acquisition of learning and the engagement of the learner. UDL offers a framework for educators' design of instruction and assessment that can heighten gaps in student's prior knowledge, scaffold and support their learning, and facilitate metacognition (Rose & Meyer, 2002; Rose & Gravel, 2010).

A. Federal Definition for UDL.

The inclusion of specific UDL language in the Higher Education Opportunity Act (2008) makes this an especially important conversation for faculty in teacher education. As defined, UDL is “a scientifically valid framework for guiding educational practice that provides flexibility in the ways information is presented, in the ways students respond or demonstrate knowledge and skills, and in the ways students are engaged” (20 U.S.C. § 1003(24)). The HEOA emphasizes the need for teacher preparation programs to prepare future teachers with instructional practices that include the use of research-based instructional strategies and technologies consistent with the principles of UDL.

The focus of the UDL framework has revealed positive results in K-12 settings (Abell, Jung, & Taylor, 2011; Coyne, Pisha, Dalton, Zeph, & Cook-Smith, 2012; Dalton, Pisha, Eagleton, Coyne, & Deysher, 2002; Dolan, Hall, Banerjee, Chun, & Strangman, 2005; Kortering, McClannon, & Braziel, 2008) particularly for students with disabilities. However, the research on the affective, strategic, and recognition networks maps to adult learners as well. Applying this framework to college courses can result in instruction that leads to positive student academic and affective outcomes (Schelly, Davies, & Spooner, 2011).

B. Universal Design for Learning.

The central tenets of UDL align with research-based practices that draw from the neurosciences, instructional design practices, and the learning sciences, which promote effective strategies and approaches for learner success (National Center on UDL, 2011a; Rose & Meyer, 2002; Rose & Gravel, 2010). Evolving research from the learning sciences has confirmed that faculty can expect students with learner variability in their classes (Fischer & Bidell, 2009). According to UDL researchers (Rose, 2001, 2005; Rose & Meyer, 2002; Rose & Strangman, 2007) individuals receive and interpret information through three primary neural networks: (a) recognition, (b) strategic, and (c) affective. Learning through these neural pathways is defined by the recognition network that involves the “what” of learning and allows individuals to identify, recognize and see patterns; the strategic network that emphasizes the “how” of learning and enable individuals to set goals, develop plans and act on these approaches; and the affective network that engages the why of learning. Three guiding principles are necessary to comprise a UDL approach (a) providing multiple, flexible methods of presentation that give students various ways to acquire information, (b) providing flexible methods of expression that offer students alternatives for acquiring and demonstrating their knowledge, and (c) offering options for engagement to help students get and stay interested as they are appropriately challenged (Rose & Gravel, 2010; Rose & Meyer, 2006).

The UDL framework is grounded in extensive research from the neurosciences, learning theories and teaching approaches (Rose & Gravel, 2010) that support the three UDL principles and nine categories offering additional guidelines for considering the customized design of instructional environments. Through the UDL guidelines, strategies can be considered to promote expertise in learning; learning that is intentional, purposeful and planned (National Center on UDL, 2012c). A table depicting the UDL guidelines is presented in Figure 1.

I. Provide Multiple Means of Representation	II. Provide Multiple Means of Action and Expression	III. Provide Multiple Means of Engagement
Perception	Physical action	Recruiting interest
Language, expressions, and symbols	Expression and communication	Sustaining effort and persistence
Comprehension	Executive function	Self-regulation

Figure 1. The UDL Guidelines. Retrieved from The National Center on Universal Design for Learning. (2012c). Universal design for learning guidelines – Version 2.0. Wakefield, MA: Author. Copyright 2012 by CAST.

Developing Clear Goals. A core first step in UDL underscores the importance of developing clear goals that align with meaningful and attainable objectives. Goals need to be analyzed with the true intent in mind. (Coyne, Ganley, Hall, Meo, Murray, & Gordon, 2006). As Csikszentmihalyi (1990) notes, clear goals are essential for assuring optimal learning experiences. Through a UDL lens, writing clear goals also clarifies the importance of separating the intent of the goal and its outcome from the means to acquire the goal offering opportunities for providing multiple options. For example, if a stated goal emphasizes the importance of reading and writing text to achieve understanding, various learners may be at a disadvantage. Without more

specificity, students who write with their voice or read text by listening will have difficulty achieving this goal.

Multiple Means of Representation. Once clear goals are established, the development of a learning environment that is rich with varied learning opportunities is necessary. The first UDL principle encourages the use of multiple methods to present information. For example, an instructor might provide a lecture and use a PowerPoint presentation with examples that expand upon their discussion. Instructional environments that capitalize on the flexibility of digital mediums offer opportunities to strengthen specific guidelines addressed within the first UDL principle and provide ways to represent vocabulary terms or mathematical notation and symbols for example. Embedded hyperlinks to unfamiliar terminology can be represented by a definition, video presentation or concept map (National Center on UDL, 2011b).

Learners also vary widely in their learning experiences. Past experiences may include gaps in prior knowledge, different cultural or regional learning experiences or, difficulties challenged by a disability. Providing options that strengthen comprehension are key to the UDL approach and assure that opportunities to activate background knowledge are weaved throughout. Instructional approaches should include multiple opportunities to highlight patterns, critical features and relationships (National Center on UDL, 2012c). Many of today's learning management systems (LMS) offer an array of digital means to embed hyperlinks that can support these important functions. These are important examples of UDL in action and that maximize opportunities for understanding (Rose & Meyer, 2006)

Multiple Means for Action and Expression. Students enter a classroom with a range of capabilities, preferences and approaches. Diversity in the college classroom is more pronounced today than ever before (National Center for Education Statistics, 2010a). Research from the learning sciences confirms that learners, in general, vary widely in how they capitalize upon instructional information in a learning situation (Rose & Fischer, 2009). There is a noticeable distinction in the strategic neural networks of a novice versus an expert learner. Depending on the task, expert learners have developed a strategic approach that facilitates their success in learning. They know how to set clear goals, outline effective steps to obtain their goals, employ effective strategies and monitor their ongoing progress until they reach that goal. In contrast, novice learners have not yet developed these facilities in learning (Bransford et al. 2000; Rose & Gravel, 2010). Many may also be "several steps behind" in their learning development due to variation in background experiences or current capabilities. The second UDL principle encourages educators to be mindful of these differences and offer multiple opportunities for students to gain, express and demonstrate their understanding (National UDL Center, 2011b).

Multiple Means for Engagement. Learners also bring varied ways and preferences for how they engage their learning. They, like many novice learners, have not yet mastered the skills needed to monitor their learning progress, adjust their plans, or determine how to maximize the classroom experience for their benefit. These students will benefit from strategies and supports that build and enrich these skills. The third UDL principle encourages educators to consider this learning area and offer multiple opportunities for students to see relevance and value. Options that further heighten the importance of goals, personalize information and encourage self-assessment for the learner are important (Rose & Gravel, 2010).

The aim of this study was to (1) investigate the practice of one instructor as she redesigned her course using the UDL framework as an overarching lens; (2) explore the ongoing development of a new UDL survey measure that explored student perspectives of UDL in college classrooms; and (3) consider the perspectives of students when UDL is part of a

postsecondary course. The research participants included both the faculty and students in an introductory research methods course.

II. Literature Review.

A. Research in Universal Design for Learning.

Research exploring the application of UDL has been limited and especially in the context of college classrooms. While this research base is growing, the evidence of impact on learning is small. Studies on UDL in higher education have often been of a descriptive nature or focused on the application of universal design principles (Center for Universal Design, 2005) and how these universal design approaches can be infused into instruction or teacher training (McGuire, Scott, & Shaw, 2006) rather than on actual implementation of UDL principles for instruction.

As a framework, UDL encourages educators to consider instructional strategies and technologies that can enhance student learning and engagement. Instruction that is delivered under a UDL lens offers expanded opportunities to personalize and deepen the learning process (Rose & Meyer, 2006; Russell, 2010). The importance of personalization has been underscored by the U.S. Department of Education's *National Education Technology Plan* which defines the term "personalization" as referring "to instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners" (U.S. Department of Education, 2010, p. 12). Spooner, Baker, Ahlgrim, Delzell, Browder, and Harris (2007) found that the inclusion of UDL principles in general and special education teacher training resulted in improved lesson plan designs and the possibility of reaching a broader range of students. When UDL was included in training for college instructors, their course designs were more effective and student's perceptions were more positive (Schelly, Davies, & Spooner, 2011). Recently, Abell, Jung and Taylor (2011) confirmed that students in middle and high school settings showed higher perception scores for both their classroom personalization and participation when UDL approaches were included in the classroom setting. In this particular study, personalization was encouraged by student opportunities to interact with the teacher; and participation was encouraged through engagement. As these authors note, personalization and participation are central to the components of UDL that encourage targeted instructional approaches to scaffold learner performance and provide opportunities for choice and engagement (p. 178).

B. Technology and Learning in Higher Education.

The use of technology in higher education is increasing as members of the millennial generations enter college (Dahlstrom, deBoor, Grunwald, & Vockley, 2011; Howe & Strauss, 2003; Oblinger & Oblinger, 2005)—bringing new digital approaches and expectations for the classroom instructor (Levin & Arafeh, 2002; Prensky, 2010). Oblinger and Oblinger (2005) note that among these "net generation" students, 20% began using computers between the ages of five and eight. The millennial generation are defined as those born from 1982 to the (Howe & Strauss, 2003). These students are developing greater digital literacy and are more comfortable in Web-based environments that focus on expression through audio, video and graphics. The affordances provided through today's digitally rich Web 2.0 environments offer students

multiple ways (and often preferences) to communicate with others (Prensky, 2010). Many are skilled (and often schooled) in using the web as a medium of expression through websites, blogs and web spaces that showcase their work.

The use of instructional approaches that infuse digital tools and resources, such as those embraced by UDL, are essential to engage today's college students. The flexible features of digital media offer many opportunities for allowing students to interact with the content, connect in conversation with others, and demonstrate their understanding. Researchers have found that "effective use of dynamic media can lead to increased student engagement" (Bull & Garofalo, 2009, p. 41). According to the *2011 National Study of Undergraduate Students and Information Technology*, students identify technology as key to "making learning a more immersive, engaging, and relevant experience" (Dahlstrom et al. 2011, p. 4). Over thirty-three percent of students in this study highlighted technology as important to their learning in college. Comments such as "makes learning more creative, makes learning more relevant to real life, makes learning more engaging, and extends learning beyond the classroom" were specifically noted (p. 11). Technology offers college students a medium for convenience, increased productivity and a way to stay connected with others. Infusing digital media into classroom instruction, supported through a UDL lens, capitalizes upon ways to customize students' learning experiences and preferences.

C. Practitioner Inquiry and Reflective Practice.

Educators that engage in reflective practice expand opportunities to enrich their teaching (Cochran-Smith & Lytle, 2010). Teaching is both a craft and an art. According to Eble (1986) the acts of teaching are multifaceted and often include a measure of sharing expertise, orchestrating a successful performance, and seeking multiple means to incite and encourage learning. In some ways, teaching viewed as an art might be compared to many of the creative professions. Take the example of a gourmet chef. Their ability to develop a signature dish depends upon their capability to become experts of various skills, consider and combine ingredients into a complete finale, and deliver this dish in a manner that tempts all the senses. Similarly, skillful teachers proceed in much the same manner as they plan their instruction and consider the different styles of learning in their classrooms and individual student needs. For skillful educators, this art is one that requires constant inquiry and reflection to provide optimal learning experience for students. Effective teachers are mindful of the importance a practice of inquiry and reflection that play integral parts in their teaching and learning.

Cochran-Smith and Lytle (2010) confirm that practitioner researchers that engage in inquiry opportunities about their teaching practice are inviting opportunities for creative ways to alter and adjust their instruction, consider the impact on their students, and continuously monitor their effectiveness. Reflective educators constantly seek opportunities to adjust and improve their practice, reconsider their inquiry stance and redevelop their classrooms into rich and meaningful learning spaces (Cochran-Smith & Lytle, 2010). The success of this practice is predicated on the notion that "teacher is inquirer rather than teacher as expert" (G. Maimon, personal communication, February 21, 2012). Often their classrooms become "research laboratories," where new approaches are implemented and students become active co-researchers in the practice. "Educational research will not have any practical value if it does not affect teaching and learning in classrooms, no matter how brilliant the design or how magnificent the result" (Wang, Kretschmer, & Hartman, 2010, p. 105).

The UDL framework underscores the importance of reflective educational practice; re-evaluating how instructional practices are addressing learner preferences is central to this brain-based approach (Rose & Gravel, 2010; Rose & Meyer, 2002). Monitoring student progress, evaluating the curriculum with learner variability in mind, and maximizing effective uses of technology are important tenets of the UDL approach.

III. Methodology.

This was an action research study of one instructor's reflective practice to explore the effectiveness of using a UDL framework with a research methods in education course at a large, urban, research-oriented university in the southeastern United States. The author of this article partnered with the instructor to explore her educational classroom practice through a framework of UDL.

Action research studies are characterized by the ability for researchers to work with practitioners in the creation of knowledge through multiple sources (Huang, 2010). As a form of inquiry, action research allows for a constant and iterative process of reflection and action that informs the practice of the educator. Through the action research process educators can "reflect on their practice to improve it" and "develop a more energetic and dynamic environment for teaching and learning" (Bruce & Pine, 2010, p. 4).

A. Participants.

Participants were eighty graduate students enrolled in two sections of an introductory research methods course. The ages of these participants were not obtained as a part of this study. Reportedly, students enrolled in graduate classes fall within the 20-34 year old age group (National Center for Education Statistics, 2012b). Students in this study included those who were degree-seeking candidates, university staff seeking credits and those not officially enrolled. Degree seeking candidates represented the majors of special education, school counseling, education leadership, business and adult learning.

The research was conducted over four semesters: Spring, 2010; Fall, 2010; Spring, 2011; and Fall, 2011. This course is required of all education majors at this university and typically taken at the beginning of a core graduate program. The purpose of the course is to introduce students to the fundamental concepts and designs of quantitative and qualitative research. Students are introduced to research designs, statistical techniques, critical scholarly research reviews and the development of research questions. The course culminates in the development of a research proposal.

During the Spring 2011 semester, an additional instructor's class for the same course was included in the total sample. This instructor taught an additional section of the research methods course and applied the UDL approach in his class. The primary faculty of this study shared an orientation to the principles of UDL, her redesigned syllabus and access to her online course structure as a model to follow. The additional instructor's class was only involved in the survey data collection and no other parts of this action research. The total number of participants across each semester is provided in Table 1.

Table 1. Total study participants across 2010 and 2011 semesters.

Semester	Number of Students Enrolled (n=80)	Percentage of Total Sample
Spring 2010	18	22.5 %
Fall 2010	14	17.5 %
Spring 2011	33	41.3 %
Fall 2011	15	18.8 %

Note: Total number of participants in boldface includes 17 from the second instructor's research methods class.

B. Aims of the Research Study.

To guide this study, the research aims focused on (A) student perceptions of faculty use of UDL in their courses, (B) student engagement related to the infusion of these practices, and (C) the relationship between the use of UDL approaches and student engagement. The questions included the following: Which strategies are both implemented by the instructor and used by the students? What patterns, if any, exist? To what extent is instruction consistent with the principles of UDL? To what extent do students take advantage of options or participate in methods consistent with the principles of UDL? How engaged do students perceive themselves to be in class? What is the relationship between students reported use of UDL strategies and their level of interest and engagement?

B. Data Sources.

In this study, multiple sources of data were collected that were both quantitative and qualitative. As part of this study, students completed a survey of questions representing practices in the UDL approach. Multiple conversations were held between the researcher and instructor to discuss how instructional approaches were being considered, implemented and changed throughout the study. Reviews of the survey participant's responses were considered at the end of each semester to inform decisions of how the next semester's class would be adjusted or designed. The researcher was able to observe and participate in some of the class sessions. Informal conversations were held with students regarding their perceptions of the class and the researcher was able to read all student online blogs.

This course was taught using UDL as the overarching framework to guide teaching and learning. The author and faculty have extensive expertise in teaching and applying the UDL framework to college instruction and as UDL instructors; they have applied a UDL perspective to other courses that focus on the three principles of this framework. As educators, they understand the importance of how applying the UDL framework to a required graduate research course can result in instruction that both leads to positive student academic and affective outcomes. The UDL lens provides a coherent instructional model that addresses four key components for curriculum planning including (1) developing clear goals, (2) considering appropriate methods, (3) selecting a range of means to deliver the material and, (4) designing

assessment approaches that are formative in nature and provide ongoing opportunities to monitor progress.

IV. Research Measures and Approaches of the Study.

A. Course Instructional Planning.

This action research study involved the standard elements of reflective and responsive instructional practice, and was focused [in particular] on one instructor's actions in response to participant feedback. The instructor used a variety of curriculum approaches to design and develop her course. Detailed course lesson plans were designed prior to each course week and followed a planned schedule of activities, discussions, presentations, media, and assigned readings. The course was taught face-to-face and met on a standard day and time of the week. Course materials, lesson plans, course notes, digital presentations, assigned readings, audio and video were also posted to a companion online course website using the university course management system (CMS). Thus, students had digital access to all of the materials and resources used in each class.

Instructor Planning Checklists. This course is often perceived by students as one that is dry in content, disassociated from their academic program and, having little meaningful application to their work. Prior to this study, the course had typically been taught with a standard course syllabus and text. Course materials and activities had been structured around previous professor's interests and instructional preferences. From a UDL standpoint, this design offered few opportunities to engage and expand the neural learning networks that connect recognition (the what of learning), strategic (the how of learning), and affective (the why of learning).

From the start, there were opportunities in applying a UDL framework to this course. The course had established clear goals; the course required the use of common grading rubrics, and the course text included a number of graphic organizers. As part of the initial course planning for this course, the instructor organized a *Checklist for UDL Implementation* for each of the identified UDL principles across recognition, strategic and affective learning networks. These checklists provided a structure for her to align course goals with intended objectives and target specific UDL approaches that supported each. Checklists for each UDL strategy implementation are illustrated in Table 2.

B. Survey Measures.

During each semester of the course, an introductory overview to the UDL principles was provided to the class. At the end of the course, students were invited to complete the *Student Survey on Learning and Instruction* (Smith, 2008). An example of this survey is provided in the appendix (see Appendix 1: Student Survey on Learning and Instruction). This paper-based survey was presented to students during class time and took no more than twenty minutes to complete. Eighty graduate students completed this survey.

The survey used in this study was designed as part of the author's original dissertation research, "Perceptions of UDL in College Classrooms" (Smith, 2008). Survey items representing the features of UDL were adapted from the *Teaching Every Student in the Digital Age: Deriving UDL Solutions Template* (Rose & Meyer, 2002). Items specifically addressed UDL and were primarily representative of the three guiding principles of UDL across the three brain areas: (a)

Table 2. Checklists for Instructor’s UDL Strategy Implementation.

Recognition Learning Network	
<i>Specific Course Goal: Standards that ask students to learn specific content</i>	
<i>Intended Course Objectives</i>	<i>Instructor Targeted UDL Approaches</i>
<p><u>Objective 1:</u> Compare and contrast quantitative, qualitative and mixed-methods approaches to research</p> <p><u>Objective 2:</u> Explain what experimental, quasi-experimental and non-experimental research designs entail and describe their application to different research questions</p> <p><u>Objective 3:</u> Explain descriptive statistical techniques such as measures of central tendency, standard deviation and correlation</p>	<p>Provide multiple formats:</p> <ul style="list-style-type: none">• Provided in-class lecture on a course topic.• Provided a recorded lecture on a course topic.• Provided any other type of audio recording related to a course topic.• Provided a video that provided additional information on a course topic. <p>Highlighted critical features:</p> <ul style="list-style-type: none">• Provided lecture notes that summarized a topic.• Provided notes with color-coding or highlighting of key points.• Provided a graphic organizer that summarized a topic.• Provided other handouts that summarized a topic. <p>Provide multiple media and formats:</p> <ul style="list-style-type: none">• Provided digital course materials online (e.g., Blackboard).• Suggested/allowed for use of a magnifier on a computer screen to improve viewing.• Suggested/allowed for changing the background color of the computer screen to improve viewing.• Suggested/allowed for using a text-to-speech application to listen to course materials

Strategic Learning Network

Specific Course Goal: Standard that ask students to learn “how” to do something

Intended Course Objectives:

Instructor Targeted UDL Approaches

Objective 1: Select a research problem and formulate appropriate research hypothesis and/or questions.

Provide flexible models of skilled performance:

- Provided an example or model of an assignment.
- Provided an assignment rubric or template.

Objective 2: Conduct a review of educational literature from texts, journals and computer databases.

Provide opportunities to practice with supports:

- Facilitated a “hands-on” activity.
- Provided materials to read text alongside guiding questions.

Objective 3: Write a coherent synthesis of such literature as it relates to the research problem.

Provide ongoing, relevant feedback:

- Was available to students for feedback on an assignment or task.
- Suggested/allowed for students contacting other individuals to ask for feedback on an assignment or task.
- Provided constructive feedback on an assignment.

Objective 4: Prepare a viable research proposal.

Offer flexible opportunities for demonstrating skill:

- Suggested/allowed for an assignment that included images or video.
- Suggested/allowed for a spell checker to check written work.
- Suggested/allowed for a word processor or other digital writing tool to create an assignment.
- Suggested/allowed for a graphic organizer to plan an assignment.
- Suggested/allowed for creation of a web-based or other digital product for an assignment.
- Suggested/allowed for inclusion of Internet hyperlinks in an assignment.
- Suggested/allowed for maintaining a digital collection or portfolio of products created for the course.
- Suggested/allowed for use of a speech-to-text application to create a written assignment.

Affective Network Learning

Specific Course Goal: Standard that asks students to enjoy, appreciate and use a content or skill area

Intended Course Objectives

Objective 1: Demonstrate interest in educational research and its methods

Objective 2: Demonstrate desire to participate in appropriate methods of research

Instructor Targeted UDL Approaches

Offer choices of content and tools:

- Posted a website to learn more about a topic in the course.
- Allowed students to select their own topic when completing an assignment.
- Allowed students to select their own materials when completing an assignment.

Offer adjustable levels of challenge

- Allowed students to select their own topic when completing an assignment.
- Allowed students to select their own materials when completing an assignment.

Offer choices of learning content:

- Allowed students to decide between working alone or with partner(s) to discuss a topic or complete an assignment.

Offer choices of rewards:

- Provided feedback on an assignment.
-

recognition, (b) strategic, and (c) affective networks. During the development of the 2008 UDL study, the current UDL guidelines (National Center on UDL, 2011c) had not yet been developed and consequently were not a part of the survey design.

The survey developed for the present study included 33 items refined from the 2008 study. The thirteen interest and engagement survey items were adapted from the *Utrecht Work Engagement Scale for Students* (Schaufeli, Martinez, Pinto, Salanova, & Bakker, 2002). The Utrecht scale included specific dimensions related to engagement defined across vigor, absorption and dedication. The absorption and dedication items were adopted into the survey for this research. The survey used a Likert-type numerical rating scale, allowing participants to select from: 0-never, 1-sometimes, 2-often (a few times a month), and 3-very often (once a week).

V. Findings and Results.

A. Which Strategies were Implemented by the Instructor and used by the Students?

The central focus of this study followed the instructional practice of one faculty member over four semesters. From the beginning semester, a focus on designing the course through a UDL lens shaped her preparation. She developed course plans that aligned clear goals with attainable objectives.

Strategies to Support Recognition Learning. To address recognition learning, she incorporated a range of ways to represent course information by developing a companion course space in the online learning management system (LMS); this provided a digital location to post class notes, PowerPoint presentations, links to audio, video and text files, and course readings. The LMS also provided an organizing structure for each week of the course to list guiding questions, individual weekly goals, and graphics or pictures that represented the weekly course theme. She offered a number of multiple means to represent content and strengthen student understanding. Delineating research traditions was established early in the class by categorizing research traditions with specific names and pictures to help students recall the approaches; “ninja” for quantitative approaches and “pirates” for qualitative methods. Assigning this type of description provided a comedic comparison similarly used in Web memes to convey information in a humorous context (Shifman & Thelwall, 2009). To maximize student’s comprehension for how to conduct a research study, she developed a “lunch tray model” to illustrate and review the essential steps and components of the research process 1) menu informs the question, 2) drink represents the research questions, 3) mixed fruit cocktail is the literature review, 4) utensils are the methods and design, 5) entree includes the results and discussion, 6) the vegetables represent the research citations to backup your discussion, and 6) the dessert includes the conclusion. An illustration of the lunch tray model is provided in Figure 1.

Strategies to Support Strategic Learning. To support strategic learning, she provided timely feedback on course assignments; sometimes this was a comment on the student’s reflection blog, a written comment using track changes within a word-processed paper or, a weekly meeting with a student after class. Students were also provided a range of ways to represent and demonstrate their understanding of research through class activities. For example, at the beginning of each class, students engaged in hands-on sentence completion activities to recall research methodology steps. As a class, students were introduced to noted researchers through instructor made “trading cards” that included a picture of the researcher on the front and noted facts on the back of each card. To strengthen their grasp of reading, understanding and sharing research, students participated in mock “cocktail party reviews” by assuming the role of a favorite researcher and sharing research with others. As part of the final assignment, students participated in a research poster session by preparing posters that depicted the essential components of their “research study” and discussed these with class “conference participants”.

Strategies to Support Affective Learning. Finally to support affective learning, the instructor began the course by crafting readings and materials with the student’s interests in mind. She provided a brief multiple intelligences inventory (McKenzie, 2002) to better understand her student’s individual interests and preferences. She considered this information in her selection of instructional approaches and readings; making sure she connected to their backgrounds and interests. She provided frequent opportunities during the class for think-pair-share and small

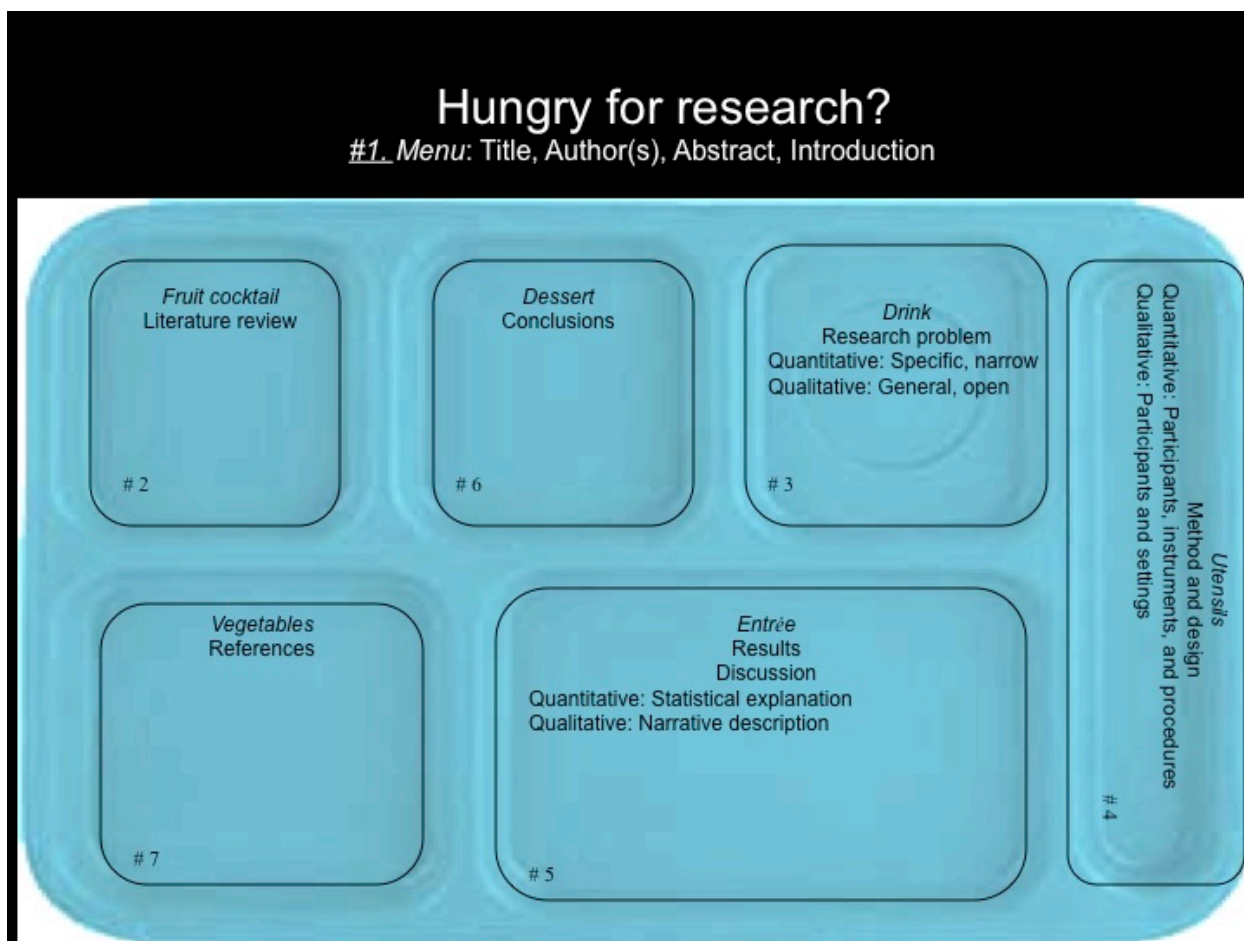


Figure 1. A Lunch Tray Model Representing the Research Process.

group discussions. An array of choices were offered in how students could consider the research topic that matched their interests, the researcher that they most wanted to explore, and the medium they chose to share this information. Students created personalized Wordpress® blogs to post their reflections to questions and prompts regarding course materials and readings. Throughout, the instructor included multiple sources of contemporary media to illustrate concepts from You Tube® videos.

B. How is Instruction Consistent with the Principles of UDL?

The student survey allowed the author and faculty to assess how consistent the instruction was with the principles of UDL from the students' perspectives. To contextualize student responses to survey questions, item descriptive statistics are provided in Table 3. A total of 33 items represented the principles of UDL for this survey. These are presented across each of the UDL learning networks and show both the instructor's planned strategies and those that were more preferred by students.

Recognition Network Learning. The UDL framework first addresses that educators need to consider providing multiple ways to present information that assist a learner in understanding the what of a learning situation. Providing instruction that addresses the recognition network of

learning includes offering multiple examples of the content, highlighting the critical features of the content, and providing a variety of media and formats. There were ten survey items addressing this area. These are illustrated in Table 3.

Table 3. Recognition Learning Network Results.

<i>UDL Recognition Learning Network</i>			
Instructor Strategy	Student Responses (n = 80)	M	SD
<i>Provided multiple examples</i>			
• Provided in-class lecture on a course topic.	Listened to an in-class lecture on a course topic	2.84	.43
• Provided a recorded lecture on a course topic.	Listened to a recorded lecture on a course topic	.65	.87
• Provided any other type of audio recording related to a course topic.	Listened to any other type of audio recording related to a course topic	1.24	.83
• Provided a video that provided additional information on a course topic.	Watched a video recording related to a course topic	1.64	.84
<i>Highlighted critical features</i>			
• Provided lecture notes that summarized a topic	Read lecture notes that summarized a topic	2.37	.83
• Provided notes with color-coding or highlighting of key points.	Read notes with color-coding or highlighting of key points	1.38	1.17
• Provided a graphic organizer that summarized a topic.	Viewed a graphic organizer that summarized a topic	2.33	.84
• Provided other handouts that summarized a topic.	Viewed other handouts that summarized a topic	2.87	.43
<i>Provide multiple media and formats</i>			
• Provided digital course materials online (e.g., Blackboard).	Accessed digital course materials online (e.g., Blackboard)	.61	1.04
• Suggested/allowed for using a text-to-speech application to listen to course material.	Used a text-to-speech application to listen to course material	.05	.22

Strategic Network Learning. The second network addressed in the UDL framework addresses the importance of providing multiple ways to present information that assists a learner in understanding the how of a learning situation. Providing instruction that addresses the strategic network of learning includes offering flexible models of skilled performance, offering opportunities for students to practice what they are learning with supports and scaffolds offering

ongoing, relevant feedback and, offering flexible opportunities for students to demonstrate their skill. There were fifteen survey items addressing this area. These are illustrated in Table 4.

Table 4. Strategic Learning Network Results.

<i>UDL Strategic Learning Network</i>			
Instructor Strategy	Student Responses	M	SD
<i>Provide flexible models of skilled performance</i>			
• Provided an example or model of an assignment.	Referred to an example or model of an assignment	2.39	.70
• Provided an assignment rubric or template.	Referred to an assignment rubric or template	2.25	.67
<i>Provide opportunities to practice with supports</i>			
• Facilitated a “hands-on” activity.	Engaged in a “hands-on” activity	2.81	.42
• Provided materials to read text alongside guiding questions.	Read text alongside guiding questions	1.48	1.07
<i>Provide ongoing, relevant feedback</i>			
• Was available to students for feedback on an assignment or task	Contacted the instructor to ask for feedback on an assignment or task	1.81	.75
• Suggested/allowed for students contacting other individuals to ask for feedback on an assignment or task.	Contacted another individual to ask for feedback on an assignment or task	1.24	.99
• Provided constructive feedback on an assignment.	Received constructive feedback from the instructor on an assignment	2.51	.55
<i>Offer flexible opportunities for demonstrating skill</i>			
• Suggested/allowed for an assignment that included images or video.	Created an assignment that included images or video	1.3	1.1
• Suggested/allowed for a spell checker to check written work.	Used a spell checker to check written work	2.67	.70
• Suggested/allowed for a	Used a word processor or other	2.76	.68

word processor or other digital writing tool to create an assignment.	digital writing tool to create an assignment		
• Suggested/allowed for a graphic organizer to plan an assignment.	Used a graphic organizer to plan an assignment	1.15	1.13
• Suggested/allowed for creation of a web-based or other digital product for an assignment.	Created a web-based or other digital product for an assignment	1.99	1.23
• Suggested/allowed for inclusion of internet hyperlinks in an assignment.	Included internet hyperlinks in an assignment	1.1	1.13
• Suggested/allowed for maintaining a digital collection or portfolio of products created for the course.	Maintained a digital collection or portfolio of products created for the course	2.11	1.06
• Suggested/allowed for use of a speech-to-text application to create a written assignment.	Used a speech-to-text application to create a written assignment	.04	1.91

Affective network learning. The third network addressed in the UDL framework addresses the importance of providing multiple ways to engage a learner in the why of a learning situation. Providing instructional strategies that addresses the affective network of learning includes offering choices of content and tools, offering adjustable levels of challenge, offering adjustable choices of learning content, and offering choices of rewards. There were five survey items addressing this area. These are illustrated in Table 5.

As noted across Tables 3, 4 and 4, there were mean scores greater than two (indicating use of the technique at least “often” or more frequently) across all of the UDL network areas. In particular, students indicated multiple representations that included in class lectures, handouts that summarized a topic, hands-on activities, multiple examples and rubrics were of most benefit to them. Opportunities that allowed them to express and represent their work through the use of digital supports and writing tools (e.g., spell checkers, word processors, digital portfolios) were encouraged. Finally, strategies and approaches that were engaging for their classroom learning included the opportunity to create digital portfolios (blogs), have the options to select their assignment topics and receive frequent feedback on their work.

C. How Engaged do Students Perceive Themselves to be in this Class?

Means and standard deviations for each of the 14 common items adapted from the Utrecht Work Engagement Scale for Students are provided in Table 6. All of these items were rated above the average by students in all of the classes with particular emphasis on tasks and graded assignments.

Table 5. Affective Learning Network Results.

Instructor Strategy	Student Perspectives	M	SD
<i>Offer choices of content and tools</i>			
• Posted a website to learn more about a topic in the course.	Visited a website to learn more about a topic in the course	2.12	.88
• Allowed students to select their own topic when completing an assignment.	Selected my own topic when completing an assignment	2.35	.69
<i>Offer adjustable levels of challenge</i>			
• Allowed students to select their own topic when completing an assignment.	Selected my own materials when completing an assignment	2.28	.73
<i>Offer choices of learning content</i>			
• Allowed students to decide between working alone or with partner(s) to discuss a topic or complete an assignment.	Decided between working alone or with partner(s) to discuss a topic or complete an assignment	1.51	.99
<i>Offered choices of rewards</i>			
• Provided feedback on an assignment.	Received feedback on an assignment	2.65	.51

Table 6. Total Interest & Engagement Scale Results.

Survey Questions Addressing Interest and Engagement	M	SD
When I am in class, I forget about everything else	2.02	1.07
I feel happy when working intensely on graded class assignments	2.23	.97
I get carried away when I'm working on class tasks and assignments	2.43	1.06
I feel happy when working intensely on tasks during class	2.45	.96
This class inspires me	2.65	1.04
Time flies when I'm in class	2.74	1.11
I am immersed in tasks during class	2.87	.97
I am enthusiastic about this class	2.90	.96
To me, graded class assignments are challenging	2.93	.91
I am immersed in the graded assignments I complete for this class	3.11	.99
I am proud of the graded assignments I complete for this class	3.36	.79
I find that graded class assignments are full of meaning and purpose	3.56	.67
I find that tasks are full of meaning and purpose	3.58	.65

The relationship of student's perception of faculty [use of] UDL use was determined by calculating a Pearson product-moment correlation to determine if a relationship existed between total student UDL and total student interest and engagement. The total student UDL composite

score (average of items attempted) represented 33 common items on *The Student Survey on Learning and Instruction in College Classrooms* for each participating student. The 14 items adapted from the *Utrecht Work Engagement Scale for Students* were grouped to form a total composite score (average score of items attempted) representing interest and engagement for each student. The total student interest and engagement composite represents the variables of interest and engagement. The Cronbach's alpha reliability coefficient for the total student UDL scale was .81 and .92 for the total interest and engagement scale. Reliability levels of .70 or greater are considered above average reliability (Schmitt, 1996).

The results of the Pearson product-moment correlation show that there was a moderate positive, statistically significant relationship between total student UDL and total interest and engagement ($r = .402, p < .01$.) In other words, when students perceive that the instructor is using more UDL strategies and technologies in their classes, they also report a higher level of their own interest and engagement.

B. Instructor Reflections and Perceptions.

In an interview with the instructor, she shared her perspectives on changes in course design, features she had implemented over the four semesters, and patterns that began to evolve.

Varied Representations of Content. Initially she began with a focus on developing the course to offer multiple representations across all three UDL areas: recognition, strategic and affective learning networks. This assured that a range of options were provided in how she presented course information, how students demonstrated their understanding, and how students became engaged with the content. She developed the categories for research traditions to help students recall specifics (e.g., ninja, pirate). She included a printable version of her course notes prior to each session. Students were encouraged to develop personalized blogs for their class reflections and encouraged to share content with multiple forms of media. Content material posted online was paired with pictures, cartoons, audio or video to provide multiple presentations. PowerPoint presentations used to supplement the lecture were posted prior to class. As previously noted, at the conclusion of each semester, the instructor would confer with this author to discuss practices and approaches and reflect upon what was and was not working in her practice.

Student Choice and Engagement. During the second semester of the study, she increased the emphasis on opportunities for student choice and engagement by bolstering instructional strategies that connected to student interests and backgrounds. Selected readings were purposefully chosen to match with student-expressed interests. Video and audio files were selected from contemporary artists or programs (e.g., the Glee© television program, songs by the artist Beyoncé©) to both enhance student comprehension and connect to the students' interests in mainstream pop media. In-class discussion groups were formed based on student interests and their designated multiple intelligence profiles. Increased opportunities to share feedback were provided through blog posts, email, in-class attention and after-class discussions. In addition, students were paired with peers to gain peer-feedback on research reviews.

Instructional Support and Scaffolds. By the third semester, a need for more strategic network items was identified by student evaluation comments. To adjust, the instructor provided more targeted feedback on course projects and quicker turnarounds on submitted paper drafts. Electronic comments, style and tracked changes were included in papers to scaffold student learning. She provided weekly meetings and conferences and offered more meaningful formative

assessment comments on reflective blogs. Students were provided multiple opportunities to demonstrate their understanding through in-class discussions, reflective blog posts, poster presentation displays and discussions of research as well as formal written papers.

UDL Resources. Finally, by the fourth semester, she included additional video clips and created “three minute reviews” to support noted and observed gaps in student prior knowledge. Students were required to access outside library resources to maximize their development of research products. What makes this course a success? As this instructor noted, using a UDL framework as a lens to plan her course helped her to focus on what was working in the instructional process. By focusing on the process she was able to target student learning versus the tool or application of technology. In fact, this instructor found that by the fourth semester, she observed that fewer of her students actually used the multiple resources in the course LMS but rather favored selected readings and their blogs. This suggested that they appreciated the opportunity to work with information that reflected their individual choice, interests and customized products. Developing clear goals was key to student understanding and helped her to separate the methods from the outcome. Employing the UDL framework helped to “delete a lot of the silliness” in the course and “helped to reduce course features that did not have an impact”. “If it wasn’t linked to the goal I took it out” (Instructor, personal communication, December 21, 2011).

VI. Discussion.

The purpose of this action-research study was to explore the reflective practice of one faculty member as she applied the UDL framework to her graduate course. This study also provided an opportunity to further analyze this college course using a survey measure designed with items representing the three UDL learning networks: recognition, strategic and affective learning. In the college classroom, student engagement is an important focus for college faculty and administrators as they seek to make the most of students’ experiences on campus and understand how college provides for these opportunities (National Survey on Student Engagement, 2011). “Educators must actively collaborate about the experience of their students, talk about what students know and can do, and design new approaches to engaging students at high levels” (p. 10). Strategies and technologies that are integral to the UDL framework, particularly those aligned with the affective neural network, can encourage student interest and engagement in a classroom setting.

While numerous articles and opinion papers have discussed the value of UDL in education, few studies have demonstrated if the applied use of UDL has a relationship to student interest and engagement. This study indicated that there was a statistically significant relationship in this area. Results of this study concurred with recent research about how people learn (Bransford, Brown, & Cocking, 2000, 2002; Bransford et al. 2006), how instructional approaches correspond with the learning brain (Rose, 2001, 2005; Rose & Meyer, 2002; Rose et al. 2005), and how multiple teaching strategies satisfy individual differences in learning. Furthermore, the study coincided with others’ contentions and findings that UDL approaches that make use of flexible opportunities inherent in digital formats can have a positive impact on student perceptions (Abell, Jung, & Taylor, 2011; Schelly, Davies, & Spooner, 2011)

This study contributes to the literature in this area in several ways. First, the results of this study provide a survey that begins to discern which UDL variables are perceived to have a relationship with student interest and engagement in college classrooms. Second, the results offer

data that support the relationship of UDL strategies and technologies on students' interest and engagement in college classrooms. Finally, the results provide a detailed look at how a faculty member redesigned a traditional graduate course through a UDL lens and the successes that evolved from this process.

The foundation of UDL is guided by three main principles: (1) provide multiple representations of content, (2) allow multiple opportunities for students to demonstrate mastery of content, and (3) offer multiple options to support learner engagement. Aligning the UDL framework to a college course provides a structure that includes multiple strategies and technologies to enrich an instructional situation. The UDL approach shifts reliance upon a single, printed text medium to one that is digital—increasing its transformable and malleable qualities (Rose & Meyer, 2002; Rose & Gravel, 2010). From an instructional design standpoint, UDL offers faculty a scientifically valid research framework (National UDL Center, 2011c; HEOA, 2008) from which they can plan and design effective courses that include attainable goals for all students. Many of the instructional strategies used by this faculty were identified as important to these students. Applying a UDL lens to this course had a positive relationship to student interest and engagement. Students indicated that specific strategies were important to their success in this course.

VII. Limitations.

The following limitations were factors that may have affected the results of this study. The design of the survey used in this study offered the opportunity to continue the construction of a new instrument to assess UDL in college classrooms. As with any evolving instrument, there are changes and adjustments that can be made. The subscales representing recognition, strategic and affective learning were based on the Rose and Meyer (2002) text *Teaching Every Student in the Digital Age: Universal Design for Learning*. These scales do not include the more expansive list of the nine detailed UDL Guidelines that have since been developed (National Center on UDL, 2011c).

In addition, there are only five items in the affective subscale of this survey and indicate the need for more representation in this area. Continued analysis and refinement of this survey is underway and an enhanced version of the measure is being tested with a more comprehensive representation of the nine UDL guidelines.

There was no sample technique used for this study, as this was an action-research design. The primary study focused on the reflective practice of one faculty member as she applied the UDL framework to her graduate class and results may not be appropriate to generalize to a larger population.

VIII. Conclusion and Implications.

Empirical studies demonstrating the value of UDL applied in college classrooms have been few. As faculties begin to see the benefits of UDL, there is a need to conduct research that can illustrate success of UDL and which aspects are most beneficial. Replication of this study is encouraged so that a broader sample of college classrooms might be considered as more faculties infuse UDL approaches and technologies in their teaching.

In view of the absence of a specific instrument for assessing effectiveness of UDL in postsecondary classrooms, the continued design, development and testing of such an instrument

is warranted. Since this survey addressed a relationship between the use of UDL technologies and strategies and student interest and engagement, it should continue to be refined and used. Ongoing efforts to update and enhance this survey are underway in hopes of providing a standard tool to guide postsecondary institutions and their faculties as they plan courses and consider specific UDL strategies that can be infused into instruction. This can ensure that UDL is being considered as a framework for the design of instruction, which can meet the needs of all students.

Continued efforts for enhancing faculty understanding of the benefits of instructional strategies and technologies, coordinated through a UDL lens, may encourage the broader understanding of why these approaches are beneficial to address learner variability in classrooms. Sharing research on best practices in the neurosciences, new learning theories, and the transformable qualities of emerging digital technologies may help expand the message on the value of UDL, for all learners, including those with disabilities. The changing knowledge base in the neurosciences will continue to inform how we consider and understand learning in the classroom. As a flexible educational model, UDL in turn can be an ever-changing model as we continue to learn and reflect more on the learning brain (Rose & Fischer, 2009; Rose & Strangman, 2007).

In addition, college students of the millennial generation often use the interactive features of digital media and mobile tools (Howe & Strauss, 2003; Dahlstrom et al. 2012). Expanding the instructional benefits of UDL approaches in college classrooms can capitalize upon contemporary tools such as electronic writing aids that automate organization, emerging Web 2.0 collaboration tools that facilitate discussion and sharing, and multiple formats that support and enhance their learning.

Because the UDL concept is relatively new, continued research is needed, particularly related to effectiveness of UDL for the myriad of diverse learners who are and will be attending college. Infusing UDL into campus environments underscores the focus of a campus community that provides effective teaching and learning for *all* students. The findings from this study suggest that when faculties consider the UDL framework to plan their course instruction and include UDL approaches and technologies in their classes, there is a positive relationship to student interest and engagement. Continuing to explore the effects of UDL approaches with students in college classrooms can provide additional perspectives on how these approaches can enhance their learning. The results from this study support the need for continued exploration in this area.

Appendix

Appendix 1. Student Survey on Learning and Instruction in College Classrooms.

Student Survey on Learning and Instruction in College Classrooms

College Name

Course Instructor Name and Email

You are being asked to complete this survey as part of an action research study being conducted by your instructor regarding student perceptions of how instructional strategies and technology are used in college courses. This study is a part of the reflective practice process and any information collected will be used in an effort to improve instructional practices. In addition, data collected as a part of this study will be shared with others who are interested in teaching and learning at the university level.

- You are not required to complete the survey, and there is no penalty for not completing it.
- If you do not wish to participate in the study, please do not complete the survey. Instead, return the blank survey to the envelope.
- By completing the survey, you are agreeing to participate in the study. Please return completed surveys to the envelope.
- The survey should take approximately 15 minutes to complete.
- If you have further questions about the survey, please ask your instructor at any time.
- This cover sheet includes contact information for the instructor; you may choose to tear off the cover page and keep it as a reminder of how to contact the instructor with questions about the study.

Thank you for your time.

To participate in the survey, please turn the page →

Directions: Respond to each item by placing a mark (X) in the box that indicates the frequency of your engagement in each activity.

	Very Often (Once a week)	Often (A few times a month)	Sometimes (Once a month or less)	Never
<i>How often have you done each of the following for this course?</i>				
Visited a website to learn more about a topic in the course				
Selected my own topic when completing an assignment				
Selected my own materials when completing an assignment				
Decided between working alone or with partner(s) to discuss a topic or complete an assignment				
Received feedback on an assignment				

Please turn the page →

	Very Often (Once a week)	Often (A few times a month)	Sometimes (Once a month or less)	Never
<i>How often have you done each of the following for this course?</i>				
Engaged in a “hands-on” activity				
Created an assignment that included images or video				
Used a spell checker to check written work				
Used a speech-to-text application to create a written assignment				
Used a word processor or other digital writing tool to create an assignment				
Referred to an example or model of an assignment				
Referred to an assignment rubric or template				
Used a graphic organizer to plan an assignment				
Created a web-based or other digital product for an assignment				
Included internet hyperlinks in an assignment				
Read text alongside guiding questions				
Contacted the instructor to ask for feedback on an assignment or task				
Contacted another individual to ask for feedback on an assignment or task				
Received constructive feedback from the instructor on an assignment				
Maintained a digital collection or portfolio of products created for the course				

Please turn the page →

	Very Often (Once a week)	Often (A few times a month)	Sometimes (Once a month or less)	Never
<i>How often have you done each of the following for this course?</i>				
Listened to an in-class lecture on a course topic				
Listened to a recorded lecture on a course topic				
Listened to any other type of audio recording related to a course topic				
Watched a video recording related to a course topic				
Watched a video that provided additional information related to a course topic				
Read lecture notes that summarized a topic				
Read notes with color-coding or highlighting of key points				
Viewed a graphic organizer that summarized a topic				
Viewed other handouts that summarized a topic				
Accessed digital course materials online (e.g., Blackboard)				
Used a magnifier on a computer screen to improve viewing				
Changed the background color of the computer screen to improve viewing				
Used a text-to-speech application to listen to course material				

Please turn the page →

*Directions: The following statements are about how you feel at work. Please read each statement carefully. Respond to each item by placing a mark (X) in the box that indicates how often you feel this way when thinking of **this class**.*

	Always (Every class)	Often (Once or twice each month)	Sometimes (Once a month or less)	Rarely (A few times or less)	Never
I find that tasks in class are full of meaning and purpose					
I find that graded class assignments are full of meaning and purpose					
Time flies when I'm in class					
I am enthusiastic about this class					
When I am in class, I forget about everything else					
This class inspires me					
I feel happy when working intensely on tasks during class					
I feel happy when working intensely on graded class assignments					
I am proud of the graded assignments I complete for this class					
I am immersed in the graded assignments I complete for this class					
I am immersed in tasks during class					
I get carried away when I'm working on class tasks and assignments					
To me, graded class assignments are challenging					

Please turn the page → Thank you for completing this survey. Please place it in the envelope with the other students' surveys.

References

- Abell, M.M., Jung, E., and Taylor, M. (2011). Students' perceptions of classroom instructional environments in the context of 'universal design for learning'. *Learning Environment Research*, 14, 171-185.
- Bransford, J.D., Brown, A.L., and Cocking, R.R. (1999). Technology to support learning. *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- Bransford, J.D., Brown, A.L., and Cocking, R.R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- Bransford, J.D., Vye, N., Stevens, R., Kuhl, P., Schwartz, D., Bell, P., Meltzoff, A. et al. (2006). Learning theories and education: Toward a decade of synergy. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology* (2nd ed.) (pp. 207-244). Mahwah, NJ: Erlbaum.
- Bruce, S.M. and Pine, G.J. (2010). *Action Research in Special Education: An Inquiry Approach for Effective Teaching and Learning*. New York: Teachers College Press.
- Bull, G. and Garofalo, J. (2009). Dynamic media. *Learning & Leading with Technology*, 40-41.
- Center for Universal Design. (2005). *What is universal design?* Retrieved from http://www.design.ncsu.edu/cud/univ_design/princ_overview.htm
- Cochran-Smith, M. and Lytle, S. (2010). *Inquiry as Stance: Practitioner Research for the Next Generation*. New York: Teachers College Press.
- Coyne, P., Ganley, P., Hall, T., Meo, G., Murray, E., and Gordon, D. (2006). In D. H. Rose & A. Meyer (Eds.). *A Practical Reader in Universal Design for Learning*, (1-13). Cambridge, MA: Harvard Education Press.
- Coyne, P., Pisha, P., Dalton, B., Zeph, L. A., and Cook-Smith, N. (2010, August). Literacy by design: A universal design for learning approach for students with significant intellectual disabilities. *Remedial and Special Education*, DOI: 10.1177/0741932510381651. Retrieved from <http://rse.sagepub.com/content/early/2010/08/30/0741932510381651>
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: HarperPerennial.
- Dahlstrom, E., deBoor, T., Grunwald, P., and Vockley, M. (2011). ECAR national study of undergraduate students and information technology. *Executive Summary*, Educause Center for Applied Research, 4-34.

Smith, F. G.

Dalton, B., Pisha, B., Eagleton, M., Coyne, P., and Deysher, S. (2002, January). *Engaging the text: Strategy instruction in a computer-supported reading environment for struggling readers. Executive Summary*. Center for Applied Special Technology. Retrieved from <http://www.cast.org/system/galleries/download/byCAST/EngagTextResearchRept1202.pdf>.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., and Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *The Journal of Technology, Learning, and Assessment*, 3(7), 1-33.

Eble, K.E. (1986). *The Craft of Teaching*. San Francisco: Jossey-Bass.

Fischer, K.W., and Bidell, T. R. (2006). Dynamic development of action, thought and emotion, W.Damon & R. M. Lerner (Eds.), *Theoretical Models of Human Development. Handbook of Child Psychology* (6th ed., Vol. 1, pp. 313-399. New York: Wiley.

Higher Education Opportunity Act. (2008). PL 110–315, 122 §3079.

Hitchcock, C., and Stahl, S. (2003). Assistive technology, universal design, and universal design for learning: Improved learning outcomes. *Journal of Special Education Technology*, 18(4), 45-52.

Howe, N., and Strauss, W. (2003). *Millennials go to College*. Washington, DC: The American Association of Collegiate Registrars and Administrative Officers.

Huang, H.B. (2010). What is good action research? *Action Research*, 8(1), 93-109.

Kortering, L.J., McClannon, T.W., and Braziel, P.M. (2008). Universal design for learning: A look at what algebra and biology students with and without high incidence conditions are saying. *Remedial And Special Education*, 29(6), pp.352-363.

Levin, D., and Arafeh, S. (2002). The digital disconnect: The widening gap between internet savvy students and their schools. *Pew Internet & American Life Project*. Washington, DC: American Institutes of Research.

McGuire, J.M., Scott, S.S., and Shaw, S.F. (2006). Universal design and its applications in educational environments. *Remedial and Special Education*, 27(3), 166-175.

McKenzie, W. (2002). *Multiple Intelligences and Instructional Technology: A Manual for Every Mind*. Eugene, OR: International Society for Technology in Education.

National Center for Education Statistics. (2010a). *Fast facts*. Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=60>

National Center for Education Statistics. (2012b). Enrollment trends by age. *The Condition of Education*. Retrieved from http://nces.ed.gov/programs/coe/indicator_ope.asp

Smith, F. G.

National Center on UDL. (2011a). Learner Variability and UDL. In *UDL Series*. Retrieved from http://udlseries.udlcenter.org/presentations/learner_variability.html?plist=explore

National Center on UDL. (2012b). *What is UDL?* Retrieved from <http://www.udlcenter.org/aboutudl/whatisudl>

National Center on UDL. (2012c). *UDL guidelines: Version 2.0*. Retrieved from <http://www.udlcenter.org/aboutudl/udlguidelines>

National Survey of Student Engagement. (2011). *Fostering student engagement campus wide annual results 2011*. Bloomington, IN: Indiana University Center for Postsecondary Research.

Oblinger, D.G., and Oblinger, J.L. (2005). *Educating the Net Generation*. Retrieved from <http://www.educause.edu/educatingthenetgen>

Prensky, M. (2010). *Teaching Digital Natives: Partnering for Real Learning*. Thousand Oaks, CA: Corwin.

Raz, A., and Buhle, J. (2006). Typologies of attentional networks. *Nature Reviews Neuroscience*, 7, 367-379.

Rose, D.H. (2001). Universal design for learning: Deriving guiding principles from networks that learn. *Journal of Special Education Technology*, 16(1), 66-70.

Rose, D.H. (2005). Cognition and learning: Meeting the challenge of individual differences. *ACM SIGACCESS Accessibility and Computing*, 83, 30-36.

Rose, D.H., and Gravel, J.W. (2010). Universal design for learning. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International Encyclopedia of Education* (119–124). Oxford: Elsevier. Retrieved from www.udlcenter.org/sites/udlcenter.org/files/TechnologyandLearning.pdf

Rose, D.H., and Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: ASCD.

Rose, D.H., and Meyer, A. (2006). *A Practical Reader in Universal Design for Learning*. Cambridge, MA: Harvard Education Press.

Rose, D.H., and Strangman, N. (2007). Universal design for learning: Meeting the challenge of individual differences through neurocognitive perspective. *Universal Access in the Information Society*, 5(4), 381-391.

Rose, L.T., and Fischer, K.W. (2009). Dynamic systems theory. In R.A. Shweder (Ed.), *The Child: An Encyclopedia Companion*. Chicago: University of Chicago Press.

Smith, F. G.

Russell, M. (2011). Personalizing assessment. In T. Gray & H. Silver-Pacuilla (Eds.), *Breakthrough Teaching and Learning* (pp. 111-126). New York: Springer Publishing.

Schaufeli, W.B., Martinez, I.M., Pinto A.M., Salanova, M., and Bakker, A.B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology*, 33(5), 464-481.

Schelly, C.L., Davies, P.L., and Spooner, C.L. (2011). Student perceptions of faculty implementation of universal design for learning. *Journal of Postsecondary Education and Disability*, 24(1), 17-30.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353.

Shifman, L., and Thelwall, M. (2009). Assessing global diffusion with web memetics: The spread and evolution of a popular joke. *Journal of the American Society for Information Science and Technology*, 60(12), 2567-2576.

Smith, F.G. (2007). *Perceptions of universal design for learning (UDL) in college classrooms*. Ed.D. dissertation, The George Washington University, United States -- District of Columbia. Retrieved September 21, 2008, from Dissertations & Theses @ George Washington University - WRLC database. (Publication No. AAT 3296852).

Spooner, F., Baker, J.N., Harris, A.A., Ahlgrim-Delzell, L., and Browder, D.M. (2007). Effects of training in universal design for learning on lesson plan development. *Remedial and Special Education*, 28(2), 106-116.

United States Department of Education (2010). *2010 national education technology plan*. Retrieved from <http://www.ed.gov/technology/netp-2010>

Wang, Y., Kretschmer, R.E., and Hartman, M.C. (2010). Teacher-as-researcher: Theory-into practice. *American Annals of the Deaf*, 155(2), 105-108.

Students' misconceptions in psychology: How you ask matters...sometimes

Annette Kujawski Taylor¹ and Patricia Kowalski¹

Abstract: Misconceptions about psychology are prevalent among introductory students. Just how prevalent and what can be done to change these misconceptions depends on valid methods of assessment. The most common method of assessment, the true/false questionnaire, is problematic. The present study compared true/false with forced choice formats to determine whether the formats give different estimates of student misconceptions. Introductory psychology students (N = 165) answered 39 misconceptions in both the true/false and forced choice formats. Students differed in accuracy when assessed with the different formats, with 33.05% accuracy for true/false and 41.29% accuracy for forced choice. In the analyses of individual items we observed that some items did not differ in level of accuracy across formats and other items did differ. We conclude that the true/false method of assessing misconceptions may overestimate students' level of misconception and recommend continued attention to how researchers assess misconceptions.

Keywords: assessment, psychology, student misconceptions

I. Introduction

Research in cognitive psychology has much to say about how people learn (Bransford, Brown, & Cocking, 2000). We know that meaningful learning is an active process. It involves the use of prior knowledge to help make sense out of new information. When prior knowledge supports or fits with new information, then the new information can be learned more easily. However, when prior knowledge contradicts new information—as when students have misconceptions—it actually makes learning the new information more difficult than if there were no prior knowledge at all (Lipson, 1982).

For over a decade we have been investigating the misconceptions that students bring with them to the introductory psychology classroom (see Taylor, Kowalski, & Laggren 2000 and Taylor & Kowalski, 2004). Consistently, we find low levels of accuracy in students' beliefs, ranging from 30% to 39% accuracy. In the time since our first studies, a number of other studies have reported similar findings (Amsel, et al., 2009; Kuhle, Barber & Bristol, 2009; Lilienfeld, 2010). Students come into the introductory psychology course with strongly held beliefs that are in contrast to the preponderance of the scientific evidence, and, unless there is some strong intervention to challenge these false beliefs, they blissfully leave the introductory psychology course with those beliefs still intact. Furthermore, examples of misconceptions that do not change based on simple instruction have been documented across a wide variety of academic disciplines such as earth science (Harackiewicz, 1999), physics (Baser, 2006; Dykstra, Boyle, & Monarch, 1992), biology (Boyes & Stanisstreet, 1991), political science (Lorenzo, 1999) and philosophy (Williamson, 2007).

¹ Department of Psychological Sciences, University of San Diego, San Diego, CA 92110, taylor@sandiego.edu

As scholar/teachers of psychological science we believe this low level of psychological science literacy is not an acceptable outcome. Reflecting on the tenacity of some students' misconceptions, our primary focus has shifted from enumerating beliefs and their sources to whether we can change students' false beliefs, how we can effect such change, why we should do so, and for how long or short a time such changes endure. In investigating how students might be encouraged to change their misconceptions, we have followed the lead of the science education and reading research literatures (Guzzetti, Snyder, Glass, & Gamas, 1993; Strike & Posner, 1992). Much of this literature suggests that to alter a misconception, students need to understand that what they think they know does not match the current scientific view and need to be presented with a plausible and useful alternative in a way they can understand. The strategy of presenting the misconception followed by a clear discussion of the scientifically correct view is referred to in the reading research as refutational text or refutational teaching (Guzzetti et al.). By using refutational texts and lectures in our introductory psychology course, we found that we significantly reduced student misconceptions and that students maintained much of this change over time (Kowalski & Taylor, 2009; Taylor & Kowalski, 2010).

Despite the desire to move forward in understanding how to change student misconceptions, we realize we may first need to step back. If we want to study factors that promote change in misconceptions we need a meaningful way to measure those beliefs. At a classroom level, Angelo & Cross (1993) recognized the importance of assessing student misconceptions as a way of informing instruction and suggested creating simple questionnaires, focusing on a number of troubling beliefs students are likely to hold. Within various disciplines, researchers have developed diagnostic instruments to assess such deficient understanding. The Force Misconceptions Test (Eryilmaz, 2002) and the Chemistry Concepts Inventory (Mulford & Robinson, 2002) are examples of such tools. Both instruments rely on multiple choice questions to sample the extent of alternative concepts in the discipline, and to study changes that occur following course instruction.

Within the field of psychology, the current methods of measuring misconceptions in psychology appear to be problematic. Psychological misconception assessments have been criticized for including outdated items, items that are not important in the science, and items that reflect claims that introductory texts do not cover (Griggs & Ransdell, 1987; Lamal, 1979). In addition, most studies have used a true/false approach in which students read a series of statements and designate each statement as being true or false. Although many criticisms of true/false testing may be misconceptions themselves (see Burton, 2005), it is generally agreed that true/false questions are difficult to write and that because of the role of guessing, true/false assessments make it difficult to know whether student responses reflect incorrect knowledge or simply a lack of knowledge. This distinction between incorrect knowledge and no knowledge is particularly important for misconception assessment because the intent is to identify student misinformation that may be tenacious and in need of being forcefully refuted.

There are additional aspects of the true/false assessments that make them particularly problematic in the study of misconceptions. Because of the nature of most misconceptions, these true/false assessments tend to phrase all items as "false" in order to be answered correctly—in other words, the misconception is a false statement students tend to believe. With all of the items needing to be correctly responded to as "false," it is difficult to know the degree to which responses reflect actual beliefs or reflect a demand characteristic, a response bias, or some other artifact of the test's format rather than the test's content. Sax (1989) noted that when students guess on true/false questions they tend to mark the statement "true". A "true" response set,

therefore, would overestimate student misconceptions. As an alternative to the true/false only questionnaire, Gardner & Dalsing (1986) found that by just adding an option for “I don’t know,” they reduced the level of reported misconceptions, presumably more accurately reflecting the students’ level of misconceptions.

Perhaps the strongest criticism of misconception assessments is that the questions include claims that are not clearly countered by the evidence, are poorly worded, or may hold a “grain of truth.” As McKeachie (1999) noted, it is difficult to write true/false questions that are clearly true or false. Brown (1984) and Ruble (1986) levied such a criticism of the true/false format for psychological misconception questionnaires. Both researchers provided evidence that the constraint of developing a single, brief statement about some complex behavior leaves most true/false statements of misconceptions ambiguous. Frequently such brief statements are only partly true and partly false. For example, many students accept as true what is known as the 10% myth, the statement that “Most people only use 10% of their brains.” Researchers in the psychological sciences clearly consider this statement false. However, outside of the research community people may have various interpretations of the word “use” in the sentence. It could be interpreted as meaning that at some point in time some people only use 10% of their available processing capacity, preferring to be lazy or in a deep stage of sleep, even if most of the brain itself is in constant use to maintain bodily functions and homeostasis. The ambiguity in interpreting true/false items becomes more obvious when one tries to write true versions of false statements. A straightforward “true” version of the 10% item could be, “Few people only use 10% of their brains.” or “Most people use 90% of their brains.” Neither of these options is a satisfactory “true” contrast to the “false” statement in that each could be read as suggesting there are some people who do use only 10% of their brains.

The problem with true/false assessments must be addressed before researchers can begin to evaluate the effectiveness of different pedagogies to effect change in misconceptions. We need an instrument that is reliable, valid, and sensitive enough to detect change without the ambiguity inherent in the true/false statements. In the present study we compared two formats which addressed the same content. A format which Bensley and colleagues have recently developed is a 2-item forced choice format (Bensley & Lilienfeld, 2010; Bensley, Lilienfeld, Ferree, Powell, & Southerly, 2011; Southerly, Bensley, Lilienfeld, Ferree, & Powell, 2011). The 2-item (A/B) forced choice format generally has a true/false statement as one option (usually the misconception) along with a complementary statement of the contrasting viewpoint based on the preponderance of the evidence. For the example above of the 10% myth, we adopted a version of the item taken from Bensley et al. (2010), “Which is most true about how much of the brain people typically use? A. People use all of their brains, but not all at once. B. Most people only use about 10% of their brains.” In this case, the two options do not exactly negate each other; however, the alternative provides a more accurate representation of what the evidence shows, in contrast to the incorrect option, the one which most students incorrectly mark as “true” in the true/false format (see Higbee & Clay, 1998; Kowalski & Taylor, 2009). The forced choice format, therefore, assesses whether the student both accepts the correct claim and rejects the incorrect claim when presented with the two alternatives simultaneously.

We therefore compared a true/false version with a force choice version of a test of misconceptions in psychology. We wanted to know whether there would be different estimates of introductory psychology students’ misconceptions across the two formats. We expected that items that show similar patterns of response across the two formats more closely represent what we might consider to be a “misconception.” In other words, if students pick the incorrect forced

choice response at the same rate as they designate the false statement as being “true,” then we believe that the item more clearly reflects a genuine level of misconception among the students. However, if students when presented with two options show a greater tendency to favor the more correct option, even if they incorrectly designate a false statement as being “true” in the true/false format, then that pattern of response suggests that the estimated level of misconception is tied to how we ask rather than to students’ underlying beliefs

For the present study we asked these research questions: Is there an overall tendency for students to respond differently to questions regarding psychological misconceptions when they evaluate true/false items compared with forced choice items that simultaneously include the incorrect and correct alternatives? When students agree with an item in the true/false format, is the agreement a result of the item format, or because they have underlying false beliefs? Evidence that students’ responses are related to item format would be suggested by a difference in the level of false belief between the true/false and force choice formats.

II. Method.

A. Participants.

The participants for this study included 164 introductory psychology students who completed the two questionnaires. Only 155 students completed both questions all the way to the end, so that the final analysis included only the 155. We did not collect demographic information; however, the 164 students comprised approximately 80% of the individuals in the introductory psychology subject pool. The individuals in the subject pool for that semester were primarily female, Caucasian, and traditional college-age freshmen. All participants received course credit for completing the questionnaire.

B. Materials.

We selected 39 items from previous questionnaires that we have used (Kowalski & Taylor, 2009; Taylor & Kowalski, 2004) as well as from the forced choice format developed by Bensley & Lilienfeld (2010). We worded all true/false (T/F) items so the correct response was “false,” except for the statement “The suicide rate is higher among the elderly than among adolescents,” which is true. For each item there was a (T/F) version and a forced choice (A/B) version. For example, the T/F version of the item regarding sugar and hyperactivity reads, “Too much sugar, such as from eating candy and sugary snacks, causes hyperactivity in children.” The forced choice version read, “Which statement about the effect of sugar on behavior is most true? A. Too much sugar, as from eating candy and sugary snacks, causes hyperactivity in children. B. Sugar has a limited effect on behavior, similar to any carbohydrate, such as potatoes or pretzels.” The preponderance of the scientific evidence favors response “B.” Furthermore, most legitimate websites that disseminate parenting information and which examine the origins of hyperactivity in school-aged children emphasize this finding rather than the “sugar = hyperactivity” explanation (see Comisarow, 1996; Huynh, 2010.).

C. Procedure.

The participants, all of whom were students in the Introductory Psychology Subject Pool for the fall 2011 semester, received notification of the study via an email announcement. The announcement provided instructions on how to access the survey online, where participants began with informed consent and then completed the questionnaire. The entire questionnaire took less than one hour to complete.

III. Results

For almost all of the items accuracy was greater when we tested participants with the A/B rather than the T/F version. Overall, a *t*-test comparing accuracy showed a significant difference with $t(38) = -4.539, p < .001$. On average, accuracy was 33.05% correct with the T/F format and 41.29% with the A/B format.

To determine whether students' responses differed depending on the format of the item, we calculated *t*-test differences between mean accuracies for each item based on either the T/F response or the A/B forced choice response. We then corrected the significance level for all comparisons by using Simes' (1986). Table 1 lists the items which met this standard of comparison. Fourteen of the 39 items met this criterion. For each of the 14 items, a greater percentage of students reported holding the misconception when answering the T/F item compared with the A/B item. Thus, for 36% of the items students' responses depended on how we asked the question.

We then subdivided the remaining items into two categories. There were those which did not meet this stringent Simes' correction procedure for significance but which independently met the .05 level of significance. Ten items met this criterion. On each of these 10 items, students also tended to favor misconceptions more frequently when responding to T/F items. Finally, the remaining 15 items met neither criterion for statistical significance. These were the items to which the participants responded at the same level of accuracy no matter how we asked.

We included items in this questionnaire based on the items' inclusion in previous questionnaires used by us and by others. There is, in fact, no absolute way to define what constitutes a misconception. Previous research has tended to differentiate items which truly represent wide-spread misconceptions from those which do not by using a criterion of 50% accuracy (Lamal, 1979; Vaughn, 1977). That is, if at least 50% of students hold a particular belief that contradicts the preponderance of scientific evidence, that is a sufficiently large percentage to consider that belief a popular misconception. Using this criterion, nine items would be considered misconceptions with the T/F format but not the A/B format. These nine items included seven of the 14 items that resulted in truly different responses across formats (T/F vs. A/B) and two additional items from the set showing significant differences without the Simes correction. We also noted that five items on the questionnaire failed to meet the 50% criterion of accuracy in both the T/F and the A/B formats and would not have been considered misconceptions with either format. Thus, of the 39 items we identified from the previous literature as reflecting frequent student misconceptions, 25 items appeared to be common misconceptions regardless of how we asked, nine depended on how we asked, and five did not appear to be misconceptions no matter how we asked the students.

Table 1. A comparison of accuracy rates for items stated in the true/false and forced choice formats and the difference between the two question formats.

Brief Item:	Accuracy T/F Percent	Accuracy A/B Percent	Accuracy A/B Accuracy T/F Percent
Items which provide different response accuracies based on question format.			
There's safety in numbers.	38.71	49.03	10.32
Opposites attract.	41.29	54.19	12.90
Women talk more than men.	45.16	58.06	12.90
Psychological profilers can describe a criminal.	17.42	30.97	13.55
Attachment is based on filling the need for food.	36.13	50.32	14.19
If unsure of a test answer, stick with the first hunch.	7.74	21.94	14.20
Right-brained and left-brained people are different.	18.06	33.55	15.49
Lie detectors are highly accurate.	57.42	73.55	16.13
It is better to vent your anger than to hold it in.	9.03	30.97	21.94
Most people experience a midlife crisis.	27.74	51.61	23.87
ESP is a well-established phenomenon.	49.68	75.48	25.80
Dyslexics see letters as their reverse.	43.23	16.77	-26.46
Autism has become an epidemic.	44.52	72.90	28.38
Dreams reflect symbolic wishes.	36.13	76.13	40.00
Items which provide different response accuracies based on looser statistical criterion.			
Too much sugar causes hyperactivity in children.	32.28	30.33	-1.95
Raising self-esteem improves achievement.	4.52	10.32	5.80
Playing Mozart increases infant intelligence.	38.71	32.36	-6.35
Subliminal messages affect buying behavior.	12.26	18.71	6.45
Elderly have higher suicide rate than adolescents.	15.48	22.58	7.10
People can learn while asleep.	49.68	57.42	7.74
People repress traumatic memories.	10.97	18.71	7.74
Crack babies have serious neurological deficits.	15.48	23.23	7.75
Immediate contact at birth is critical for bonding.	18.71	28.39	9.68
People with schizophrenia have a "split" personality.	44.52	54.84	10.32
Items which provide similar response accuracies regardless of how one asks.			
Most therapies are based on Freud.	46.45	46.45	0.00
Criminal acquitted based on insanity defense.	51.61	50.97	-0.64
Most people use only about 10% of their brains.	47.74	48.39	0.65
Electroconvulsive ("shock") therapy is dangerous.	26.45	27.10	0.65
Drug education programs are effective.	43.87	45.16	1.29
Full moon affects behavior.	77.42	78.71	1.29
Learning styles should match teaching styles.	10.97	9.68	-1.29
The first three years are critical to development.	3.87	7.74	3.87
Adolescence is a time of psychological turmoil.	19.35	15.48	-3.87
Older people are crankier and depressed.	54.19	49.78	-4.51
Inkblots reveal personalities.	39.35	43.87	4.52
Taste areas are defined on the tongue.	27.10	33.55	6.45
Therapy must examine childhood root causes.	29.68	36.77	7.09
Amnesiacs cannot remember their previous life.	30.97	39.35	8.38
Vision depends on light rays emitted from eyes.	64.52	85.16	20.64

IV. Discussion

The purpose of this study was to examine the different ways in which students can be asked about their misconceptions related to psychological science and whether these different assessment formats make a difference. One format which researchers previously used is a true/false format wherein students read a statement and designated whether they believe it reflects a true or a false statement. The second is a forced choice format in which researchers ask students to pick one of two options. We asked whether there is an overall difference in accuracy when we assess student misconceptions with true/false or forced choice formats. We also wanted to know whether student agreement with items in the true/false format reflects the question format or students' underlying false beliefs.

We found that for some items, how you ask matters, and for other items it does not matter very much. Overall, students reported fewer false beliefs when asked with the forced choice format. This suggests that our assessment of students' misconceptions can sometimes depend on how we ask. Although we do not know the exact thought processes, it is possible that for some claims, when students see only the more popular misconception (T/F version), they fail to think through the alternatives. However, when students see the correct response immediately next to the incorrect response (A/B version) they may stop to consider the correctness of the alternate choice. Further studies, perhaps with think-aloud procedures, may help determine these processes. Knowing what a student is thinking at the time of answering the item would help to determine if the student has an alternative understanding of the item rather than a misconception of the item.

We also noted that some claims on our instrument were not endorsed by the majority of students. We recommend that researchers pretest their instrument, which is likely to vary somewhat over time and over samples. For example, although other researchers have found support for the “extramission” misconception (Winer & Cottrell, 1996)—the idea that visual perception depends on light emitted from the eyes, rather than on the correct conception that visual perception depends on light entering the eyes—we failed to find support for this belief. Less than 35% of our participants thought this is an accurate statement when asked in the true/false format. (In visual perception, as a person sees something, tiny particles or light rays are emitted from the eyes.) The misconception rate fell to less than 15% when we asked about it with the forced choice format. (Which is most true about what happens in vision as a person sees something? A. Tiny particles or light rays emit from the eyes. B. Light is converted to nerve impulses in the back of the eyes.) Therefore, we conclude that how you ask matters but that cohort effects may affect beliefs; past cohorts endorsed extramission, current students do not. In addition, television programming may influence some of the misconceptions related to psychology. For several years we found that students strongly favored the misconception that the polygraph (“lie detector”) test is a highly accurate means of detecting dishonesty (see Taylor & Kowalski, 2004); however, in recent years the accuracy rate of recognizing this as being false has risen perhaps due to the influence of television shows based on fictionalized court cases which these days more correctly portray that information. Based on our informal conversations with students, the programs that do represent the polygraph as being highly accurate (“Maury”) tend to be perceived as having less credibility. Therefore, different cohorts of college students may express beliefs based on information they learn from various media outlets, and the content or quality of that information can change over time, leading to changing beliefs among the

general public, including students. Further investigations of the source of students' misconceptions will help clarify these types of cohort changes over time.

A. Conclusions

In conclusion, we found evidence that the use of the true/false format may over-estimate the prevalence of misconceptions that students endorse when they enter the introductory psychology classroom. Our overall data analyses showed that accuracy rates on average across all items, and across almost all individual items, were lower when we asked the true/false version of the item when compared to the forced choice version of the item. Although the misconceptions in this study reflect students' knowledge of psychological science, the findings fit well within the accumulating body of knowledge on student misconceptions across a wide range of disciplines (Bransford et al., 2000). Misconceptions exist in all disciplines. How you ask students about these preconceptions is likely to matter. Having the most accurate measure possible becomes important when instructors and researchers use the results of such tests in additional ways, such as examining the effectiveness of interventions to reduce these false beliefs. We recommend continued attention to methods of assessing misconceptions as researchers cautiously pursue factors which may reduce misconceptions in the classroom.

Appendix 1. Selected items, first in the True/False format and next as A/B forced choice format.

Drug education programs (i.e., DARE) are effective in deterring drug use among teenagers. Which of the following statements about drug education programs is most true?

- A. Drug education programs (i.e., DARE) are effective in deterring drug use among teenagers.
- B. Drug education programs (i.e., DARE) are ineffective in deterring drug use among teenagers.

Too much sugar, such as from eating candy and sugary snacks, causes hyperactivity in children. Which statement about the effect of sugar on behavior is most true?

- A. Too much sugar, as from eating candy and sugary snacks, causes hyperactivity in children.
- B. Sugar has a limited effect on behavior, similar to any carbohydrate, such as potatoes or pretzels.

Taste areas for sweet, sour, salty and bitter are well defined on the tongue.

Which statement is most true about the sense of taste?

- A. Taste areas for sweet, sour, salty and bitter are well defined on the tongue.
- B. People can perceive all taste qualities all over their tongue.

The suicide rate is higher among the elderly than among adolescents.

Which statement about the suicide rate is most true?

- A. The suicide rate is higher among the elderly than among adolescents.
- B. The suicide rate is higher among adolescents than among the elderly.

Most "crack babies" end up with serious neurological deficits.

Which is most true about "crack babies"?

- A. Most "crack babies" end up with serious neurological deficits.

B. Most “crack babies” develop normally in the long run.

Immediate contact between a mother and infant after birth is critical for bonding.

Which statement about mother-infant bonding is most true?

A. Immediate contact between a mother and infant after birth is critical for bonding.

B. Mothers and infants do not need immediate contact after birth to develop a bond.

A baby’s attachment for its mother is based on mom’s filling the physiological need for food.

Which statement is most true about the source of the attachment?

A. Baby’s attachment for its mother is based on filling the physiological need for food.

B. A baby’s attachment for its mother is based primarily on physical contact and not nourishment.

References

Amsel, E., Johnston, A., Alvarado, E., Kettering, J., Rankin, L., et al. (2009). The effect of perspective on misconceptions in psychology: A test of conceptual change theory. *Journal of Instructional Psychology*, 36(4), 289-295.

Baser, M. (2006). Effect of conceptual change oriented instruction on students' understanding of heat and temperature concepts. *Journal of Maltese Education Research*, 4, 64-79.

Bransford, J. D., Brown, A. L., Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, D. C.: National Academy Press.

Bensley, D. A., & Lilienfeld, S. O. (2010). The test of psychological knowledge and misconceptions. Unpublished manuscript.

Bensley, D. A., Lilienfeld, S. O., Ferree, S.K., Powell, L. A., Southerly, T. D. (2011, March). *Reliability and validity of the knowledge of psychology test*. Poster presented at the 82nd Annual Meeting of the Eastern Psychological Association, Cambridge, MA.

Boyes, E., & Stanisstreet, M. (1991) Misconceptions in first-year undergraduate science students about energy sources for living organisms. *Journal of Biological Education*, 25, 209-13. doi:10.1080/00219266.1991.9655208

Brown, L. T. (1984). Misconceptions about psychology aren’t always what they seem. *Teaching of Psychology*, 11(2), 75-78. doi:10.1207/s15328023top1102_3

Burton, R. F. (2005). Multiple-choice and true/false tests: Myths and misapprehensions. *Assessment & Evaluation in Higher Education*, 30(1), 65-72. doi:10.1080/0260293042003243904

Comisarow, J. (1996). Can sweet treats drive kids crazy? Sugar and hyperactivity in children. *Nutrition Bytes*, 2(1),

Dykstra, D. I., Boyle, C. F. & Monarch, I. A. (1992). Studying conceptual change in learning physics. *Science Education*, 76(6), 615-652. doi:10.1002/sce.3730760605

Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching*, 39(10), 1001-1015.

Gardner, R. M., & Dalsing, S. (1986). Misconceptions about psychology among college students. *Teaching of Psychology*, 13(1), 32-34. doi:10.1207/s15328023top1301_9

Griggs, R. A., & Ransdell, S. E (1987). Misconceptions tests of misconceived tests? *Teaching of Psychology*, 14(4), 210-214. doi:10.1207/s15328023top1404_4

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional intervention from reading education and science education, *Reading Research Quarterly*, 28(2), 116-159. doi:10.2307/747886

Hapkiewicz, A. (1999). Naïve ideas in earth science. *MSTA Journal*, 44(2), 26-30.

Higbee, K. L., & Clay, S. L. (1998). College students' beliefs in the ten-percent myth. *Journal of Psychology: Interdisciplinary and Applied*, 132(5), 469-476. doi:10.1080/00223989809599280

Huynh, N. (2010). Does sugar really make children hyper? *Yale Scientific*. Retrieved on 10-9-14 from <http://www.yalescientific.org/2010/09/mythbusters-does-sugar-really-make-children-hyper/>

Kowalski, P., & Taylor, A. (2009). The effect of refuting misconceptions in the psychology classroom. *Teaching of Psychology*, 36(3), 153-159. doi:10.1080/00986280902959986

Kuhle, B. X., Barber, J. M., & Bristol, A. S. (2009). Predicting students' performance in introductory psychology from their psychology misconceptions. *Journal of Instructional Psychology*, 36(2), 119-124.

Lamal, P. A. (1979). College students' common beliefs about psychology. *Teaching of Psychology*, 6(4), 155-158. doi:10.1207/s15328023top0603_8

Lilienfeld, S. O. (2010). Confronting psychological misconceptions in the classroom: Challenges and rewards. *Observer*, 23(7), retrieved on 11/11/11 from <http://www.psychologicalscience.org/index.php/publications/observer/2010/september-10/confronting-psychological-misconceptions-in-the-classroom.html>.

Lipson, M. Y. (1982). Learning new information from text: The role of prior knowledge and reading ability. *Journal of Literacy Research*, 14(3), 243-261. doi: 10.1080/10862968209547453

Lorenzo, D. (1999). Countering popular misconceptions of federal bureaucracies in American government classes. *PS: Political Science and Politics*, 32(4), 743-747. doi:10.2307/420166

Taylor, A.K. & Kowalski, P.

McKeachie, W. J. (1999). *Teaching tips: Strategies, research, and theory for the college and university teacher*, Boston: Houghton Mifflin Co.

Mulford, D. R., & Robinson, W. R. (2002). An inventory for alternative conceptions among first-semester general chemistry students. *Journal of Chemical Education*, 79(6), 739-744. doi:10.1021/ed079p739

Ruble, R. (1986). Ambiguous psychological misconceptions. *Teaching of Psychology*, 13(1), 34-36. doi:10.1207/s15328023top1301_10

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation* (3rd ed.), Belmont, CA: Wadsworth.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751-754. doi:10.1093/biomet/73.3.751

Southerly, T. D., Bensley, D. A., Lilienfeld, S. O., Ferree, S.K., Powell, L. A. (2011, March). *Psychology students' knowledge of psychology and psychological misconceptions*. Poster presented at the 82nd Annual Meeting of the Eastern Psychological Association, Cambridge, MA.

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice* (pp. 147-176). Albany, NY: SUNY.

Taylor, A., & Kowalski, P. (2004). Naïve psychological science: The prevalence, strength and sources of misconceptions. *Psychological Record*, 54, 15-25.

Taylor, A. K., & Kowalski, P. (April, 2010). *Goal orientation and learning strategies predict conceptual change over time*. Poster presented at the 90th Annual Western Psychological Association Conference, Cancun, Mexico.

Taylor, A., Kowalski, P., & Laggren, L. (April, 2000) *Myths & misconceptions about psychology: Strength of belief is not related to accuracy*. Poster presented at the 80th Annual Convention, Western Psychological Association, Portland, OR.

Vaughan, E. D. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology*, 4, 138-141. doi:10.1207/s15328023top0403_9

Williamson, T. (2007). *The philosophy of philosophy*. Oxford, UK: John Wiley & Sons.

Winer, G. A., & Cottrell, J. E. (1996). Does anything leave the eye when we see? Extramission beliefs of children and adults. *Current Directions in Psychological Science*, 5, 137-142. doi:10.1111/1467-8721.ep11512346

B. If you have an emergency you are more like to get help when a relatively small group observes that the person needs help.

People diagnosed with schizophrenia have a “split” personality.

Which is most true of people with schizophrenia?

A. People diagnosed with schizophrenia have a “split” personality.

B. Schizophrenics have just one personality but their perception of reality is shattered.

Opposites attract: people are most romantically attracted to people who differ from them.

Which is most true regarding interpersonal attraction?

A. Birds of a feather flock together: people are more romantically attracted to people who are similar to them.

B. Opposites attract: people are most romantically attracted to people who differ from them.

Women talk more than men (“Men are from Mars, women are from Venus”).

Which is most true about the ways men and women typically communicate?

A. Men and women talk about the same amount and make as many interruptions.

B. Women talk more than men (“Men are from Mars, women are from Venus”).

“Psychological profilers” have been shown to be effective at describing the personality of an unknown criminal.

Which is most true about criminal profiling by large investigative organizations and police departments?

A. Psychological profilers are not much more accurate than other people in describing the personality of an unknown criminal.

B. Profilers have been shown to be more accurate than other people at describing the personalities of unknown criminals.

A baby’s attachment for its mother is based on mom’s filling the physiological need for food.

Which statement is most true about the source of the attachment?

A. Baby’s attachment for its mother is based on filling the physiological need for food.

B. A baby’s attachment for its mother is based primarily on physical contact and not nourishment.

If you’re unsure of your answer while taking a test, you should stick with your initial hunch.

Which is most true about what you should do if you are unsure of your answers on a test?

A. You should change your answers when you think you should.

B. You should stick with your initial hunch about what is right.

Some people are right-brained and some are left-brained which can predict individual differences, such as who is more creative.

Which is most true about differences between the hemispheres of the brain?

A. Some people are R-brained and some are L-brained which can predict individual differences, such as who is more creative.

B. Which hemisphere is dominant in a person mostly does not matter because the brain works in an integrated fashion.

The polygraph ("lie detector") test is a highly accurate means of detecting dishonesty.

Which is most true about the standard polygraph or "lie detector" test?

- A. It is a highly accurate means of detecting dishonesty.
- B. It can misidentify people as liars, when they are telling the truth.

In visual perception, as a person sees something, tiny particles or light rays are emitted from the eyes.

Which is most true about what happens in vision as a person sees something?

- A. Tiny particles or light rays emit from the eyes.
- B. Light rays that enter they eye are converted to nerve impulses in the back of the eyes.

It is better to vent your anger or "blow off steam" than to hold it in.

Which is most true concerning the expression of anger?

- A. It is better to vent your anger or "blow off steam" than to hold it in.
- B. It is better to control the expression of your anger.

Most people in their "middle years," in their 40's or early 50's, experience a midlife crisis, often resulting in greater dissatisfaction and higher rates of divorce than other ages.

Which is most true about people in their "middle years," their 40's or early 50's?

- A. Some experience a need for a "mid-course correction," but most experience no greater dissatisfaction than other ages.
- B. Most experience a midlife crisis, often resulting in greater dissatisfaction and higher rates of divorce than other ages.

ESP (extrasensory perception) is a well-established psychological phenomenon.

Which is most true about ESP?

- A. There is only questionable evidence establishing ESP as a psychological phenomenon.
- B. ESP is a well-established psychological phenomenon.

The defining feature of dyslexia is that people see letters as their reverse.

Which is most true about dyslexia?

- A. The defining feature of dyslexia is that people see letters as their reverse.
- B. Dyslexics have more difficulty with sounding out letters than seeing them properly.

There has recently been an epidemic of cases of childhood autism.

Which is most true of autism?

- A. There has recently been an epidemic of cases of childhood autism.
- B. The way children are diagnosed has recently led to more reported cases of autism.

Dreams primarily reflect wishes communicated in symbolic language.

Which is most true about dreams?

- A. Dreams reflect everyday experiences and concerns.
- B. Dreams reflect wishes communicated in symbolic language.

References

- Amsel, E., Johnston, A., Alvarado, E., Kettering, J., Rankin, L., et al. (2009). The effect of perspective on misconceptions in psychology: A test of conceptual change theory. *Journal of Instructional Psychology*, 36(4), 289-295.
- Baser, M. (2006). Effect of conceptual change oriented instruction on students' understanding of heat and temperature concepts. *Journal of Maltese Education Research*, 4, 64-79.
- Bransford, J.D., Brown, A.L., and Cocking, R.R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, D. C.: National Academy Press.
- Bensley, D.A., Lilienfeld, S.O., Ferree, S.K., Powell, L.A., and Southerly, T.D. (2011, March). *Reliability and validity of the knowledge of psychology test*. Poster presented at the 82nd Annual Meeting of the Eastern Psychological Association, Cambridge, MA.
- Boyes, E., and Stanisstreet, M. (1991) Misconceptions in first-year undergraduate science students about energy sources for living organisms. *Journal of Biological Education*, 25, 209-13. doi:10.1080/00219266.1991.9655208
- Brown, L.T. (1984). Misconceptions about psychology aren't always what they seem. *Teaching of Psychology*, 11(2), 75-78. doi:10.1207/s15328023top1102_3
- Burton, R.F. (2005). Multiple-choice and true/false tests: Myths and misapprehensions. *Assessment & Evaluation in Higher Education*, 30(1), 65-72. doi:10.1080/0260293042003243904
- Developmental reading disorder: Dyslexia (n.d.). PubMed Health. Retrieved from: <http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0002379/>
- Dykstra, D.I., Boyle, C.F., and Monarch, I.A. (1992). Studying conceptual change in learning physics. *Science Education*, 76(6), 615-652. doi:10.1002/sci.3730760605
- Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching*, 39(10), 1001-1015.
- Gardner, R.M., and Dalsing, S. (1986). Misconceptions about psychology among college students. *Teaching of Psychology*, 13(1), 32-34. doi:10.1207/s15328023top1301_9
- Griggs, R.A., and Ransdell, S.E (1987). Misconceptions tests of misconceived tests? *Teaching of Psychology*, 14(4), 210-214. doi:10.1207/s15328023top1404_4
- Guzzetti, B.J., Snyder, T.E., Glass, G.V., and Gamas, W.S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional intervention from reading education and science education, *Reading Research Quarterly*, 28(2), 116-159. doi:10.2307/747886

Hapkiewicz, A. (1999). Naïve ideas in earth science. *MSTA Journal*, 44(2), 26-30.

Higbee, K.L., and Clay, S.L. (1998). College students' beliefs in the ten-percent myth. *Journal of Psychology: Interdisciplinary and Applied*, 132(5), 469-476.
doi:10.1080/00223989809599280

Kowalski, P., and Taylor, A. (2009). The effect of refuting misconceptions in the psychology classroom. *Teaching of Psychology*, 36(3), 153-159. doi:10.1080/00986280902959986

Kuhle, B.X., Barber, J.M., and Bristol, A.S. (2009). Predicting students' performance in introductory psychology from their psychology misconceptions. *Journal of Instructional Psychology*, 36(2), 119-124.

Lamal, P.A. (1979). College students' common beliefs about psychology. *Teaching of Psychology*, 6(4), 155-158. doi:10.1207/s15328023top0603_8

Lilienfeld, S.O. (2010). Confronting psychological misconceptions in the classroom: Challenges and rewards. *Observer*, 23(7), retrieved on 11/11/11 from
<http://www.psychologicalscience.org/index.php/publications/observer/2010/september-10/confronting-psychological-misconceptions-in-the-classroom.html>.

Lipson, M.Y. (1982). Learning new information from text: The role of prior knowledge and reading ability. *Journal of Literacy Research*, 14(3), 243-261. doi: 10.1080/10862968209547453

Lorenzo, D. (1999). Countering popular misconceptions of federal bureaucracies in American government classes. *PS: Political Science and Politics*, 32(4), 743-747. doi:10.2307/420166

McKeachie, W. J. (1999). *Teaching tips: Strategies, research, and theory for the college and university teacher*, Boston: Houghton Mifflin Co.

Mulford, D.R., and Robinson, W.R. (2002). An inventory for alternative conceptions among first-semester general chemistry students. *Journal of Chemical Education*, 79(6), 739-744.
doi:10.1021/ed079p739

Ruble, R. (1986). Ambiguous psychological misconceptions. *Teaching of Psychology*, 13(1), 34-36. doi:10.1207/s15328023top1301_10

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation* (3rd ed.), Belmont, CA: Wadsworth.

Simes, R.J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3), 751-754. doi:10.1093/biomet/73.3.751

Southerly, T.D., Bensley, D.A., Lilienfeld, S.O., Ferree, S.K., and Powell, L.A. (2011, March). *Psychology students' knowledge of psychology and psychological misconceptions*. Poster

Taylor, A. K. and Kowalski, P.

presented at the 82nd Annual Meeting of the Eastern Psychological Association, Cambridge, MA.

Stanovich, K.E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, 21(10), 590–604. doi:10.1177/002221948802101003

Strike, K.A., and Posner, G.J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice* (pp. 147-176). Albany, NY: SUNY.

Taylor, A., and Kowalski, P. (2004). Naïve psychological science: The prevalence, strength and sources of misconceptions. *Psychological Record*, 54, 15-25.

Taylor, A.K., and Kowalski, P. (April, 2010). *Goal orientation and learning strategies predict conceptual change over time*. Poster presented at the 90th Annual Western Psychological Association Conference, Cancun, Mexico.

Taylor, A., Kowalski, P., and Laggren, L. (April, 2000) *Myths & misconceptions about psychology: Strength of belief is not related to accuracy*. Poster presented at the 80th Annual Convention, Western Psychological Association, Portland, OR.

Vaughan, E.D. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology*, 4, 138-141. doi:10.1207/s15328023top0403_9

Williamson, T. (2007). *The philosophy of philosophy*. Oxford, UK: John Wiley & Sons.

Winer, G.A., and Cottrell, J.E. (1996). Does anything leave the eye when we see? Extramission beliefs of children and adults. *Current Directions in Psychological Science*, 5, 137-142. doi:10.1111/1467-8721.ep11512346

Examining exam reviews: A comparison of exam scores and attitudes

Jana Hackathorn¹, Kathryn Cornell², Amy M. Garczynski³, Erin D. Solomon³, Katheryn E. Blankmeyer³, and Rachel E. Tennial³

Abstract: Instructors commonly use exam reviews to help students prepare for exams and to increase student success. The current study compared the effects of traditional, trivia, and practice test- based exam reviews on actual exam scores, as well as students' attitudes toward each review. Findings suggested that students' exam scores were significantly higher for the exams following both a traditional review and a trivia review than for the exam following the practice test review. Immediately after the review, three attitude measures (i.e. confidence, feeling prepared, and helpfulness of review) were lowest after the traditional review. Finally, immediately after taking the exam, students reported higher ratings (i.e. confidence, feeling prepared, and helpfulness) after the trivia review. Implications of these results are discussed.

Keywords: Exam Reviews, Teaching Effectiveness, Learning

I. Introduction.

Using exam reviews to help students prepare for an exam leads to better test results in comparison to no review at all (King, 2010). However, a majority of the literature regarding exam reviews has been either anecdotal, or has mainly focused on measures of students' attitudes. For example, Middlecamp (2003) measured students' perceptions of a review game for an abnormal psychology course. The findings suggested that most of the students enjoyed the review game and felt it was challenging. However, Middlecamp did not measure the actual effectiveness of playing the game on exam scores. This is problematic, as student perceptions of the review's effectiveness may not align with the review's actual effectiveness. Furthermore, literature regarding direct comparisons of differing types of exam reviews is sparse. Arguably, any exam review, if done correctly, can garner positive effects for students' success. However, what remains unclear is whether certain exam reviews are more effective than others. Thus, the current study sought to directly compare the effectiveness of a few variations of exam reviews (i.e. traditional, trivia, and practice-test).

Traditional exam reviews, a common approach, ask students to compose questions that the instructor will answer during a scheduled, class-wide meeting. However, the success of a traditional exam review is contingent upon students coming to class prepared to ask questions. Moreover, the traditional exam review is often received by students as 'just another class lecture', passive and disengaging, and only including, or benefitting, the few students who prepared (King, 2010; Paul, Hollis, & Messina, 2006). In attempts to remedy this, many educators have created and experimented with new ways to implement reviews with the hopes of

¹ Jana Hackathorn, Murray State University, 209 Wells Hall, Murray KY 42071, jhackathorn@murraystate.edu

² Murray State University

³ Saint Louis University

improving exam scores, increasing attendance for review days, and also in influencing attitudes (e.g., satisfaction, feelings of efficacy) about the given subject.

One alternative is for instructors to implement trivia-style games, similar to the television show, *Jeopardy*, to review material. These are often utilized because students appear to enjoy them, and can be easily created for any level course regardless of material type or difficulty. For example, Keck (2000) created a trivia game, based on the game show *Jeopardy*, as a supplement to a traditional review before a final exam in an upper level chemistry course. Results indicated that students enjoy trivia-based review games, and that students who participated yielded higher exam scores than those who did not participate.

Similarly, Paul, Hollis, and Messina (2006) examined the effectiveness of a trivia game on exam scores. Their findings suggested that students' grades increased significantly across all four exams for students who played the trivia game, as opposed to students who chose not to play. Self-report measures indicated that the game experience taught students how and what to study. Perhaps most importantly, students reevaluated their study methods based on the inadequacies of their knowledge discovered during the game reviews.

A major limitation to Paul and colleagues' (2006) study, however, was that students chose whether or not to attend the class when the game was played. That is, students self-selected into the experimental or control group. This is problematic for a couple of reasons. First, it is unclear whether students who did not choose to play the game received a different type of review, or no review at all. Moreover, it is unclear whether this type of review is helpful for all students, or just the students who would choose to play a review game. That is, an individual difference variable, such as intrinsic motivation or interest in the subject material, could be producing higher exam grades. Perhaps those students who chose to play a trivia game would have scored higher on the exams regardless, because they possess greater motivation.

Finally, practice tests are an additional alternative to traditional reviews: however, past literature is inconclusive regarding their effectiveness. Some studies indicate that students prefer technology-based learning, particularly practice tests, because they appreciate the opportunity to check their knowledge level, and get a 'sneak peek' at the format and wording of exam questions (King, 2010; Pemberton, Borrego, & Cohen, 2006). For example, Balch (1998) assigned students to either take a practice test, or to review a completed practice test prior to an exam, and found that actively taking the practice test significantly improved subsequent exam scores. The author argued that the process of actually taking an exam, and subsequently getting answers incorrect, allows students to understand their weaknesses. Additionally, Balch argued that by taking the practice tests, the students are engaging in deeper cognitive elaboration, improving understanding and retention of material.

Additionally, King (2010) used a slightly different version of the practice test exam review, in which students used clickers to answer multiple choice questions. Attendance and exam grades were higher for clicker reviews relative to other reviews, however, similar to Paul and colleagues (2006), results are confounded by the fact that students self-selected into the exam review or no exam review conditions.

Conversely, other studies have shown little to no effects of the practice test review. Pemberton, Borrego, and Cohen (2006) found that using online computer-based practice tests in reviews did not improve exam grades compared to those who had traditional review. Additionally, Kalat (1983) showed that a practice test that mimicked the style of real exams had no effect on exam scores, as compared to a control group with no review.

In summary, past research is inconclusive regarding the effectiveness of different types of exam reviews. What does seem clear is that students report greater liking of non-traditional exam review types, but it is not clear whether this translates to higher exam scores. Thus, the current study compared the effectiveness of three types of exam reviews. Although there are a plethora of exam reviews available, the traditional, trivia, and practice test exam reviews were chosen as a focus for the current research because they are commonly used in college classrooms and there is inconclusive evidence as to their effectiveness. It is important to note that the purpose of the current study was not to discredit any type of exam review. Instead, as there are limitations in the past literature as well as contradictions in success rates and effectiveness, the current study sought to compare the effects of a traditional review, trivia game, and practice-tests on actual exam scores. In addition, the current study sought to measure students' attitudes toward each review, and used a within-groups design, utilizing only students who attended all three types of review as a means of comparison.

As this study is one of the first known to purposefully compare various types of review, we posed the following research question: which type of exam review will result in the highest exam scores? As little past literature has compared the effectiveness of these types of exam reviews, it was difficult to predict which would result in the highest exam scores. Therefore, no specific hypotheses were made as to which exam review type would be the most effective for actual exam scores. Additionally, we were interested in which type of review style would be rated most favorably by students. Based on prior literature (e.g., King, 2010), we expected that students' overall attitudes would be least favorable of the traditional exam review immediately following the review, and again after the actual exam was taken. Specifically, we expected students to rate the traditional exam review as the least helpful, report feeling the least confident and the least prepared relative to the practice-test or trivia review styles.

II. Methods.

A. Participants.

Participants in the current study were 78 undergraduates enrolled in two separate sections of a social psychology course, each with a different instructor, at a Midwestern university. Both courses met three days a week, for 50 minutes each, during a traditional 16-week semester. However, as it is consistent with the purpose of the current study, only those students who participated in all three of the exam reviews were included in the analysis of actual exam scores ($n = 44$). Those students' ages ranged from 18-22 years ($M = 19.75$, $SD = 1.54$, $Med = 20.00$). The classes contained Freshmen ($n = 7$), Sophomores ($n = 17$), Juniors ($n = 7$), and Seniors ($n = 5$). Eight students did not provide their college year. Additionally, there were 26 women, 10 men, and eight participants who chose not to answer. The ethnic background of students was predominantly Caucasian ($n = 27$), but also included African-American ($n = 3$), Indian ($n = 2$), Hispanic ($n = 1$), Bi-racial ($n = 1$), Asian ($n = 1$), and Bosnian ($n = 1$) ethnicities. Eight students' ethnic information was not reported. Finally, most of the students were psychology majors ($n = 14$), however there was also language majors ($n = 5$), biology ($n = 2$) and human resources ($n = 1$). The remaining 23 students did not answer the question.

B. Procedure.

Undergraduate psychology students in two upper level social psychology courses were instructed at the beginning of the semester that throughout the duration of the course an examination of various teaching techniques and exam reviews would occur. Students were also informed that any and all activities were voluntary, and they could choose not to complete any or all of the surveys. They did not receive any extra credit for their participation or completion of surveys. However, in an attempt to increase attendance in each of the exam reviews, extra credit was offered for participating in each exam review.

To control for variance across courses, the instructors worked together to cover the same material using the same notes and identical exam review procedures. Additionally, both courses used the same textbook, the same homework assignments, and assessed learning through the exact same multiple choice exams. On the class period prior to each of the three exams, students were presented with a different type of exam review. The reviews consisted of a trivia-style game review for the first exam, followed by a traditional review for the second exam, and finally a practice test prior to the third exam, which was also the final exam. To increase attendance and participation at each of the exam reviews, extra credit points were added to students' final grades. In order to maintain the integrity of the dependent variable, exam scores before the extra credits were added were used in the analyses.

1. Trivia Exam Review.

On the scheduled class day prior to the first exam, students participated in a trivia game in which they were randomly assigned to teams of five to six students. The trivia game was set up so that the instructors created a series of questions prior to the class. Questions ranged from giving definitions and having students recall or recognize (via a multiple-choice format) the correct concepts, to giving real-world examples and students correctly applying concepts. Students took turns representing their group for each question. Groups were asked questions pertinent to material on the upcoming exam, and a point was awarded to the first team with the correct answer.

2. Traditional Exam Review.

On the scheduled class day prior to the second exam, students participated in a traditional review. Students were instructed to come to class with questions regarding material that was unclear or needed further clarification. During the class period, students were called on to read a prepared question aloud and the instructor would answer the question for them, or provide clarification regarding subject matter that was unclear. Other students were not necessarily given an opportunity to answer the questions, but were allowed to ask follow-up questions or discuss the material until everyone in the class understood. There were no limits to how many questions each student could ask, and instructors used the full class time.

3. Practice Test Exam Review.

On the scheduled class day prior to the third exam, students participated in a computerized practice test on their own computers outside of the classroom. The computer practice test took

place during the same 50 minute time period that the scheduled class occurred. Students were informed that the amount of questions they answered correctly would not impact their course grade or extra credit. The 20 questions on the practice test mimicked the format, wording, and material of actual multiple choice questions of the upcoming exam, but were not identical to actual test items. Only 20 questions were given, as this provided students the opportunity to take notes, look up answers if necessary, or get further clarification from their notes or textbooks. After students finished the practice test, they were informed of which question(s) they got incorrect, the correct answer(s) and an explanation of the correct answer(s).

C. Measures.

1. Exam scores.

Each test consisted of a variety of multiple choice questions, short answer questions, and essay questions. However, only the multiple choice portion of each of the three exams was used to measure effectiveness of exam reviews. This was done to reduce instructor bias in the measurement of the dependent variable. Analyzing only the multiple choice portion reduces teacher bias in grading, providing a more objective measure. Students' answers were either correct or incorrect. No partial credit was awarded. The first exam contained 16 multiple choice questions, the second exam contained 23 multiple choice questions, and the third exam contained 21 multiple choice questions. Since each exam had a different number of multiple choice items, total scores were converted into percentages. Each exam was identical across sections of the course, and care was taken to make the three exams as comparable as possible. Each exam contained questions at the Knowledge, Comprehension, and Application levels of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956).

2. Attitudes Following Exam Reviews.

Immediately following each exam review, students were given three-single-item surveys assessing their exam confidence, feelings of preparedness, and the helpfulness of each exam review. Participants responded to three questions (i.e., "After this exam review, how confident do you feel about taking the exam?" "After this exam review, how prepared do you feel?" and "How helpful do you feel this exam review was?" on a 7-point Likert scale ranging from (1 = *not at all* to 7 = *very much*).

A slightly revised version of the attitude survey was administered again immediately after students completed the exam. This provided students an opportunity to reevaluate their perceptions of each review style, after having seen the items on the exam. Specifically, the three items were reworded to indicate tense (e.g., after taking this exam, how helpful do you feel this exam review was?).

III. Results.

As there were two instructors in the course, analyses were conducted to determine if there were any differences in mean scores across instructors. Results indicated that there were no significant differences in scores for any of the three exams. Thus, for the sake of brevity, the data was collapsed across the two classrooms.

To examine our main research question, which queried the effectiveness of each type of review on actual exam scores, a repeated measures ANOVA was conducted using only exam scores for the students who participated in all three exam reviews ($n = 44$). Results indicated there was an effect of review, $F(2, 86) = 30.93, p < .001, \eta_p^2 = .43$. Pairwise comparisons indicated that exam scores after traditional review ($M = .87, SD = .09$) and trivia reviews ($M = .86, SD = .10$) were significantly higher than after the practice-test-based review ($M = .74, SD = .10, ps < .001$). Traditional and trivia exam scores were not significantly different from one another ($p > .05$). See Figure 1 for an illustration of the differences between exam scores.

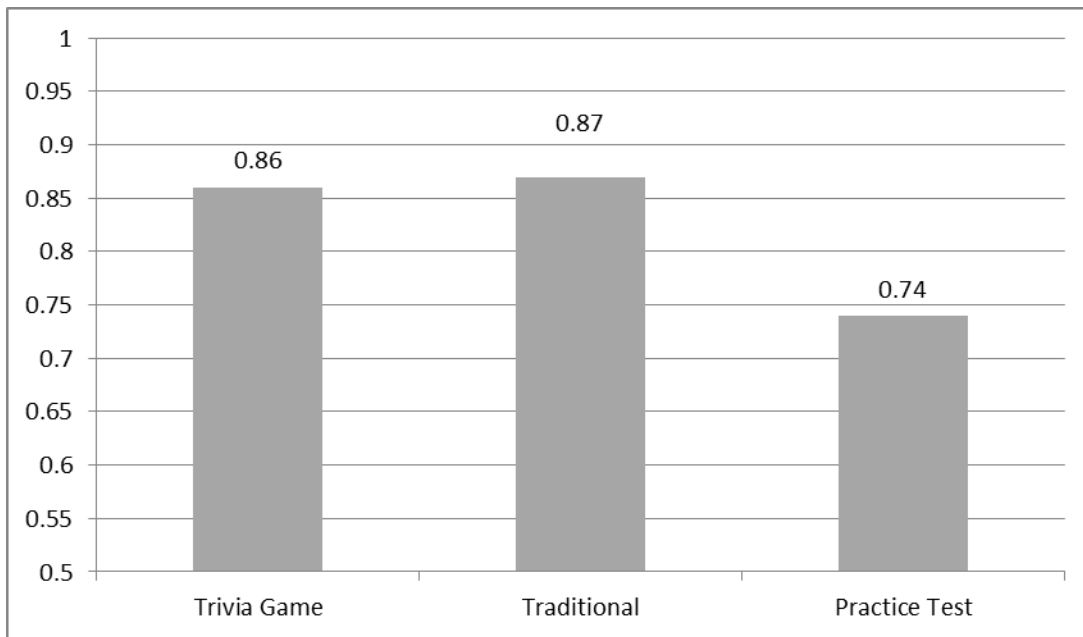


Figure 1. Percentage of correct answers between three exam review types.

In addition, it was hypothesized that attitudes toward the traditional review would be less favorable than the trivia or practice test reviews. Repeated measures ANOVAs were conducted on students' ratings of confidence, preparedness, and how helpful they felt the exam reviews were, immediately after the exam review. Findings indicated that, although the attitudes following the trivia review and the practice-test review were not significantly different from one another ($p > .05$), students reported more confidence after the trivia and practice-test review than after the traditional review ($p = .002$ and $p < .001$, respectively). Students also reported feeling more prepared after the practice-test review than both the trivia ($p = .019$) and the traditional ($p < .001$) review. Finally, students reported that both the trivia and practice-test review were more helpful than the traditional review ($p < .001$); however, student ratings of helpfulness for the trivia and practice-test reviews were not significantly different from one another ($p > .05$). See Table 1 for summary of analyses for each item.

Immediately following each exam, students rated their confidence again regarding each of the reviews. Results of repeated measure ANOVAs indicated students reported more post-exam confidence following the trivia review than after the practice-test review ($p = .003$), but ratings were not different from the traditional review ($p > .05$). Additionally, post-exam confidence was marginally higher after the traditional review than the practice-test review ($p = .075$). Students also reported feeling more prepared after the trivia and traditional review than

after the practice-test review ($p = .014$ and $p = .005$, respectively). However, the trivia and traditional review were not significantly different from one another in students' feelings of preparedness before the exam. Finally, after taking the exam, students reported the trivia game was more helpful than the traditional review ($p = .009$) but was not significantly different from the practice-test review ($p > .05$). Additionally, the traditional review and the practice-test review were not significantly different from one another ($p < .05$) in the students' reports of helpfulness. See Table 2 for summary of analyses for each individual attitude item.

Table 1. Analysis of Attitude Items Immediately after the Exam Review.

	Trivia M(SD)	Traditional M(SD)	Practice-test M(SD)	<i>F</i>	<i>p</i>	η^2_p
Confidence	4.64(1.11) ^b	3.98(1.03) ^{ac}	4.74(0.87) ^b	9.82	<.001	.18
Prepared	4.11(1.09) ^c	3.72(1.17) ^c	4.60(0.97) ^{ab}	8.86	<.001	.16
Helpful	5.39(1.11) ^b	3.89(1.40) ^{ac}	5.28(1.26) ^b	20.74	<.001	.32

Note: Superscript notations indicate significant difference from other review; ^a = difference from trivia, ^b = difference from traditional, ^c = difference from practice-test.

Table 2. Analysis of Attitude Items after the Exam.

	Trivia M(SD)	Traditional M(SD)	Practice-test M(SD)	<i>F</i>	<i>P</i>	η^2_p
Confidence	5.56(1.19) ^c	5.24(1.08)	4.79(0.98) ^a	5.28	.007	.14
Prepared	5.38(1.16) ^c	5.41(1.17) ^c	4.65(1.36) ^{ab}	5.52	.006	.13
Helpful	5.22(1.38) ^b	4.54(1.22) ^a	4.70(1.55)	2.97	.057	.08

Note: Superscript notations indicate significant difference from other review; ^a = difference from trivia, ^b = difference from traditional, ^c = difference from practice-test.

IV. Discussion.

The current study is the first known to directly compare the effectiveness of traditional, trivia, and practice-test reviews on actual exam scores. Findings indicate that exam scores were significantly higher following both a traditional and a trivia review than following the practice-test review. This suggests that compared to practice-test reviews, traditional and trivia exam reviews more effectively increase students' exam scores. These findings support and extend past literature which indicates that not only do students enjoy trivia games as an exam review, but that students who actively participate in trivia games tend to show increased exam scores (e.g., Keck, 2000; Paul, Hollis, & Messina, 2006), especially when compared to no review at all. Our findings also support the notion that practice tests have very little effect on exam scores (e.g., Kalat, 1982), although, it is important to note that practice tests reviews are better than no review at all (Balch, 1998).

Additionally, the current study examined students' attitudes toward each of the exam reviews. Based on findings in prior literature in which students' attitudes were measured, we expected that students' attitudes would be most favorable of the trivia and practice test reviews, and least favorable of the traditional review (King, 2010; Paul et al. 2006). That is, we expected students to rate the traditional exam review as least helpful, and would report they felt less confident, and less prepared than after the practice-test or trivia reviews. Our findings partially

supported this hypothesis. Results indicated that prior to taking the exam, all three attitude measures (confidence, feeling prepared, and helpfulness of review) were lowest after the traditional review, in spite of the fact that the traditional review was more helpful than the practice test. However, after taking the exam, the students' attitudes shifted in that they were least favorable of the practice-test-based review. Perhaps students are aware of which types of reviews are actually the most helpful to them, but only after taking the actual exam and not immediately after the review itself. Future studies should investigate if students are aware of their learning and studying styles, and whether this corresponds with the types of reviews they prefer.

At both points, immediately after the review and after taking the exam, students' ratings of the trivia game were the most favorable of any of the review styles. Importantly, students' scores on the actual exam were also highest after the trivia review. There are several possible reasons for this finding. Perhaps the trivia game reviews are more appreciated by students because they are similar to practice test reviews which give students a 'sneak-peek' at test wording while also explicitly pointing out areas where more studying is needed (King, 2010). Or, perhaps as students find the trivia games more fun they feel more at ease and less anxious (Middlecamp, 2003). This lack of anxiety may help them to study more effectively and increase their chances of success. However, measuring students' fun and anxiety toward the review was beyond the scope of this study.

Our findings also suggest that a small disconnect exists between the perceived effectiveness and the actual effectiveness of the traditional exam review. That is, students did not seem to realize the helpful nature of traditional exam reviews. Immediately following a review, students feel that the traditional reviews are the least helpful, but their actual scores were higher after the traditional review than the practice-test review. Moreover, even after taking the actual exam, students still did not perceive the effectiveness of the traditional review. It has been noted that the success of the traditional exam reviews are contingent upon students attending the session prepared with questions for the instructor, which suggests some amount of studying before the review session (King, 2010; Paul, Hollis, & Messina, 2006). It is possible that students perceive that the traditional review session is supposed to guide their studying or even replace their studying altogether. Thus, a lack of preparedness and student attitudes may lead to the perceived ineffectiveness of traditional reviews. Another possibility is that the students who feel the least prepared following the review session put additional effort into studying for the exam, thereby explaining the disconnect between feelings of preparedness and actual exam scores.

Past studies (e.g., King, 2010) have suggested that perhaps students find traditional exam reviews to be stuffy, awkward, or just plain boring. As part of this study, although not part of our explicit analysis, students were asked open-ended questions regarding what they liked and disliked about each exam review. After the traditional exam review, a theme emerged within students' comments pertaining to the dislikes of the traditional review stating that the traditional review was boring, an overall feeling that not all of the material was covered, and that the overall session was disorganized. As students generate their own questions to a traditional exam review, and ultimately take turns, the material mentioned in one question is not necessarily connected to the question that was asked before it. This may inhibit proper note taking, and even lead the material to become more confusing to them. Thus, they feel less prepared and confident, and ultimately underestimate the helpfulness of the review session. Perhaps if instructors structured the traditional exam reviews so that students' questions were grouped by chapter or concept, this might ease students' concerns.

V. Limitations.

Although this study adds to the literature regarding exam reviews, it is not without limitation. As would be important in any class, care was taken to ensure the exams were of similar difficulty. However, as each exam covers different topic material (e.g., self vs. influence), it is possible that there were inherent differences in difficulty. Thus, it is possible that the exam following the practice-test review was more difficult than the exam following the traditional and trivia reviews. A brief qualitative examination indicated that all three exams contained multiple choice questions at the knowledge, comprehension, and application levels of Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). However, it should be noted that the exam following the practice-test review contained four application level questions, while the traditional review exam contained only two application questions and the trivia exam contained only one. Although there is a slight discrepancy, it is our belief that this is not a noteworthy enough difference to sway scores in such an extreme direction.

VI. Recommendations for Future Research.

This study offers some important information for the scholarship of teaching and learning regarding the actual outcomes of exam review sessions, as well as some of students' perceptions of the effectiveness of each exam review style. However, it does not provide much qualitative information regarding what students may or may not appreciate about each type of review. Knowing the unique features that students like and dislike in an exam review, might help instructors to develop a review session that is not only effective, but also highly appreciated and enjoyed by students.

As instructors, we should use these findings, in addition to future findings, to create review sessions that students can simultaneously benefit from and enjoy. Perhaps, combining the different types of exam reviews to create one that is both effective and enjoyable is the most beneficial type of review. For example, each review could begin with a trivia review game that takes only half the allotted class time, and then cover the most 'missed' topics in a more traditional review fashion. Or, after each trivia question is answered, probing students for information regarding what part of that specific material is still muddy, or providing more information as to why the answer is the correct answer. This conglomeration of review styles could significantly ease students' anxiety, increase their enjoyment and ultimately their participation in exam reviews, and importantly increase their knowledge of the material which culminates in increased exam scores.

References

Balch, W.R. (1998). Practice versus review exams and final exam performance. *Teaching of Psychology, 25*, 181-184.

Bloom, B.S., Engelhart, M.D., Furst, F.J., Hill, W.H., and Krathwohl, D.R. (1956). *Taxonomy of educational objectives: Cognitive domain*. New York: McKay.

Drouin, M.A. (2010). Group-based formative summative assessment related to improved student performance and satisfaction. *Teaching of Psychology, 37*, 114-118.

Kalat, J.W. (1983). Effect of early practice test on later performance in introductory psychology. *Teaching of Psychology, 10*, 53.

Keck, M.V. (2000). A final exam review activity based on the Jeopardy format. *Journal of Chemical Education, 4*, 483.

King, D. (2010). Redesigning the preexam review session. *Journal of College Science Teaching, 40*, 88-96.

Middlecamp, M.K. (2003). Uncover the disorder: A review activity for abnormal psychology courses. *Teaching of Psychology, 30*, 62-63.

Paul, S.T., Hollis A.M., and Messina, J.A. (2006). A technology classroom review tool for general psychology. *Teaching of Psychology, 33*, 276-279.

Pemberton, J.R., Borrego, J., and Cohen, L.M. (2006). Using interactive computer technology to enhance learning. *Teaching of Psychology, 33*, 145-147.

Team-based learning in a subsection of a veterinary course as compared to standard lectures

Erin Malone and Amie Spieth

Abstract : Team-Based Learning (TBL) maximizes class time for student practice in complex problems using peer learning in an instructor-guided format. Generally entire courses are structured using the comprehensive guidelines of TBL. We used TBL in a subsection of a veterinary course to determine if it remained effective in this format. One section of the class was taught the material using PowerPoint based lectures. The other group was taught the same material by the same instructor using TBL. All students took the same examination on the material at the end of the course and again 18 months later. There were no differences in the course examination or course grades but grade distributions differed; TBL grades were more widely distributed and female TBL students outperformed male TBL students. TBL students scored significantly higher on the repeat examination. Objective student engagement was high and students were positive about the experience.

Keywords: veterinary education, urinary surgery, knowledge retention, course design, active learning, adult education

I. Introduction.

Team-Based Learning (TBL) is a constructivist course format designed to shift the focus of classroom time from instructor-led lectures and discussions to interactive small group sessions captained by students, all without decreasing class sizes, increasing instructor numbers or increasing class time (Michaelsen & Sweet, 2008a). With TBL, students review material prior to class in order to be prepared for deeper discussions and active learning during class. Units generally start with short tests (readiness assessment tests or RATs) covering basic material from the assigned readings. Students take the tests individually (IRAT) and then as a group (GRAT) and get immediate feedback on their responses. Misconceptions and errors are clarified with a short lecture. During the remaining sessions for the unit, students work with their teammates in carefully balanced groups of 6-7 members on complex real-world problems that require higher level thinking skills. The application exercises are designed to follow the “4S” guidelines: centered on a problem of *significance* to the students, all groups should be working on the *same* problem, students must make a *specific* choice, and groups should report their choices *simultaneously* (Michaelsen & Sweet, 2008b). The instructor guides the process to ensure the important components are “discovered” and applied appropriately. Students are held responsible for the preclass work with in-class assessments and via peer feedback and reviews (Sweet & Pelton-Sweet, 2008). Repeating the knowledge acquisition and application cycle through the readiness assessment tests and complex problems generally leads to increased understanding and retention (Pileggi & O’Neill, 2008; Goldberg & Dintzis, 2007), while group discussions encourage the development of teamwork, better communication and enhanced problem solving as well as engaging the learners and creating an active learning environment (Haidet, O’Malley, & Richards, 2002; Giuliodori, Lujan, & DiCarlo, 2006; Pickrell, Boyer,

Oehme, Clegg, & Sells, 2002). This format also means the instructor is available to assist with the more difficult problem solving and application skills rather than being involved primarily with the initial dissemination of knowledge. Unlike problem-based learning (PBL), the instructor creates problems and activities that guide the learning process rather than allowing students to determine their learning path. One instructor can manage a large number of students, as compared to PBL which relies on a facilitator for each group. Finally, TBL is more readily adapted to all levels of learning, including the clinical setting (Davis & Harden, 1999).

While group work, case-based discussions and PBL have been used and evaluated in veterinary curriculums, TBL has not been commonly used in veterinary colleges and we could find no peer-reviewed articles on the subject. TBL is widely used across many disciplines, including other health sciences, but is usually applied to an entire course (Michaelsen & Sweet, 2008b). In the veterinary curriculum at the University of Minnesota (UMN), many courses are taught by multiple instructors. The curriculum is very full and students have limited time to reflect on their learning. We were interested in finding a way to convert the ever expanding course material into a format that encouraged more active and deeper learning while not increasing student workload. Switching class sessions to TBL carried the potential for better understanding of the material due to more meaningful and thorough discussions with the instructor and with peers, more enjoyable class time, and better long term retention through application practice without altering the student: teacher ratio or increasing the time students spent on the material. TBL had not previously been used in our curriculum and coordinating efforts across multiple instructors with a new teaching modality could be highly time-consuming and possibly ineffective (Thompson, Schneider, Haidet, Perkowski, & Richards, 2007).

The UMN Urinary Systems Disorders course (CVM 6460) is taught as an introductory course to all second year veterinary students. The large animal subsection of CVM 6460 has been taught by one of the authors (EM) for several years. Historically, the classes were taught using the standard lecture/PowerPoint format, with many attempts made during class to make the discussions interactive using case examples and in-class questions. However, based upon student knowledge exhibited during clinical rotations, retention and understanding remained suboptimal with students who were taught using this format.

Our goal was to evaluate the use of TBL in a large animal urinary surgery as taught in a subsection of a urinary systems course. Our hypothesis was that TBL could be utilized in a subsection of a veterinary medicine course with equivalent or better test results and student and instructor satisfaction as compared to standard lectures. Furthermore, we anticipated TBL would lead to better long term retention of the material.

A. Course Description.

The Urinary Systems Disorders course is a required course. The large animal surgery portion of the course has routinely comprised three lectures of 50 minutes each. We chose this subsection to trial TBL due to the limited risk it posed to the students and interest of the instructor. Since the class is meant to serve as an introduction, the topics would be reintroduced further along in their education. If the TBL format failed to provide adequate learning, we could remedy deficiencies later. Also, this group of students was involved in a student leadership course that included discussion and practice of effective group interactions, including appreciating different strengths and styles, problem solving as a team, and giving peer feedback (Root Kustritz & Nault, 2010). We hoped this would counter the limited time available for group dynamics as

TBL generally utilizes peer feedback to improve member responsibility and group function but opportunities to adjust would be limited in a single session.

Institutional review board approval was obtained to alter the course during the spring of 2010. The CVM curriculum committee approved offering the subsection of the course twice during the semester to allow controlled evaluation. The course was required for all 92 students in the second year of the curriculum.

II. Methods.

A. Study Population.

Second year veterinary students were randomly divided based upon student identification numbers. One section of the class served as the control group, attending the standard set of three PowerPoint based lectures. The other students in the class were presented with the same material in a TBL environment. Each group of students was instructed not to attend the alternate class sections

B. Course administration.

Students in the TBL group attended a brief meeting to explain the process, the research questions, and obtain informed consent.. After explanation of TBL, portions of TBL were demonstrated during this session including diverse group formation, the concept of the individual (IRAT) and group (GRAT) readiness assessments, use of a classroom response system (i>Clickers), and advanced team based problem solving. The mock IRAT also served as a preclass survey to assess student study habits, use of social media during class, and frequency of group work. The explanation was performed individually for the few students unable to attend the introductory session.

A course website was developed on which *all* the students were able to access class objectives, lecture PowerPoints, class notes, study questions, and extra reading materials. The site also contained a single forum to allow students from either section to post questions about the material or about the course expectations. Students in the TBL portion of the course were asked to spend 1.5-2 hours reviewing the material prior to attending class and in lieu of a similar amount of class time. Students in the control group were also granted access to this material but were not required to use it.

C. TBL Session.

A single, separate class period was reserved for the TBL session. Groups of 6-7 students were formed transparently at the start of class to maximize group diversity (based upon interest in surgery, experience in large animal veterinary medicine, advanced degrees, and other factors). Each group was given a packet of material containing the IRAT questions in written form along with group member evaluation forms.

The 90 minute class started with students taking the IRAT on paper, followed by clicker responses to record answers. Groups then convened to discuss the questions and take the GRAT in a similar fashion. After the first student completed the IRAT, a two minute warning was called. Most students required 10-12 minutes for a 10 question IRAT, with only 5 minutes

needed for the corresponding GRAT. Upon completion, the instructor led the class in a discussion of the answers to the IRAT/GRAT questions.

Students were then presented with case studies in which they were required to reason through diagnosis and appropriate action, using their knowledge of the material and with small group discussion. The instructor was on hand to assist with questions. Following TBL “4S” guidelines, groups responded simultaneously with specific choice answers. Subsequently, the instructor led a class discussion of each of the cases, allowing the students to give explanations for their choices, while also giving input and suggestions. Two cases were covered in the remaining time period.

An additional 15 minutes were spent performing evaluations of group members and of the TBL format. For the peer evaluations, each student was given a total of 10 points multiplied by the number of other group members (not including themselves) to distribute (Davidson variation of the Michaelsen method; Levine, 2008).

D. Examinations.

At the end of the CVM 6460 course, students from both the control and TBL groups were given the same exam including 15 large animal urinary surgery questions (Appendix 1). Large animal urinary surgery questions were generally case related, application type problems and were pulled from a study guide that was given to both sections. Due to variations in class discussions, a deliberate attempt was made to ensure the questions were evenly distributed so that any question that reflected a topic more heavily emphasized in one section was balanced by a question that was better emphasized in the opposite section. Over half of the questions were either equally emphasized in both sections or were not emphasized in either section. Grades were assigned on a standard scale with >90% resulting in an A and <60% resulting in an F.

All students were asked to repeat the same 15 question examination during the first half of their 4th (final) year of veterinary school. Students were asked to repeat the examination without reviewing the material and without using any resources (online or print). Students further identified whether they were in the TBL or non-TBL groups, were pursuing training in large animal species, and which large animal rotations they had experienced at the time of the repeat examination. Participation was anonymous and voluntary.

E. Evaluation.

To ensure groups were similar, comparisons were made between TBL and control groups in terms of prior GPA and ultimate grades for the course using Student’s t-tests. Time spent on the website was evaluated both to determine if TBL students were accessing the materials prior to class and to obtain a baseline of website use by control group students.. To determine if the class format affected performance, Pearson’s chi square test was used to analyze test question accuracy and number of students both failing and excelling in the large animal urinary surgery subsection of the course. The impact of TBL group discussions was analyzed by comparing differences in IRAT and GRAT scores for group members. Calculations were performed to determine the correlation between IRAT/GRAT scores and examination scores. Student responses on pre- and post-course surveys and professor’s response to the format (based upon notes taken throughout the process) were also collected. Finally, repeat examination scores were compared using Student’s t-tests with individual questions analyzed by Pearson’s chi square

test. TBL and non-TBL students retaking the examination were compared using Pearson's chi square test for large animal interest and with Student's t-tests for number of large animal rotations.

III. Results.

A. Group comparison.

Of the 92 students in the second year class, 50 students were randomly selected to participate in the TBL section and 38 participated (76%). The remaining 12 students joined the non-TBL group. No attempt was made to determine why students declined participation. In the TBL group, 26% of the students were male (n=10), while in the non-TBL group, 20% were male (n=11). No significant difference was present between groups in terms of cumulative GPA (TBL : 3.38 ± 0.46 ; non-TBL : 3.29 ± 0.43 ; $p=0.3341$).

B. Preclass work.

All but two of the TBL students were on the course website at least 30 minutes in advance of the TBL class session (36/38; 95%) while 17/55 (69%) of the control group students were on by the same time frame, despite already having one standard class session. Most TBL students reported they followed the guidelines and studied 1-2 hours prior to class (22/31 students completing the survey), with one studying less and seven studying more than two hours. The majority of the students only reviewed the lecture PowerPoint and notes (27/31; 87%). The question forum on the course website was rarely used and only to ask what to study, not for questions on the material. In the preclass survey of the same students, 97% reported they never or rarely reviewed material prior to class and 37.5% reported they sometimes use extra resources.

C. Exam performance.

No students failed the course and there were no statistically significant differences in overall course grades between groups (TBL: 36.8% A's; non-TBL: 44.4% A's). There were no significant differences in urinary surgery test scores between TBL and non-TBL students (TBL 22.6 ± 4.0 ; non-TBL 21.9 ± 3.4 ; total 30 points; $p=0.3984$). However, non-TBL students were more compactly distributed for urinary surgery test scores, with TBL students having more A's and more F's on the subsection test (Figure 1). This pattern was not seen in course grades and student performance on the urinary surgery portion of the test did not vary by class rank quartile (Figure 2). Additionally, while scores for males and females were similar for the non-TBL students (21.5 ± 3.5 men; 22.1 ± 3.5 women; $p=0.5946$), women in the TBL group significantly outperformed the men (20.0 ± 2.8 men; 23.4 ± 4.1 women; $p=0.0116$). There were no significant performance differences between groups by sex.

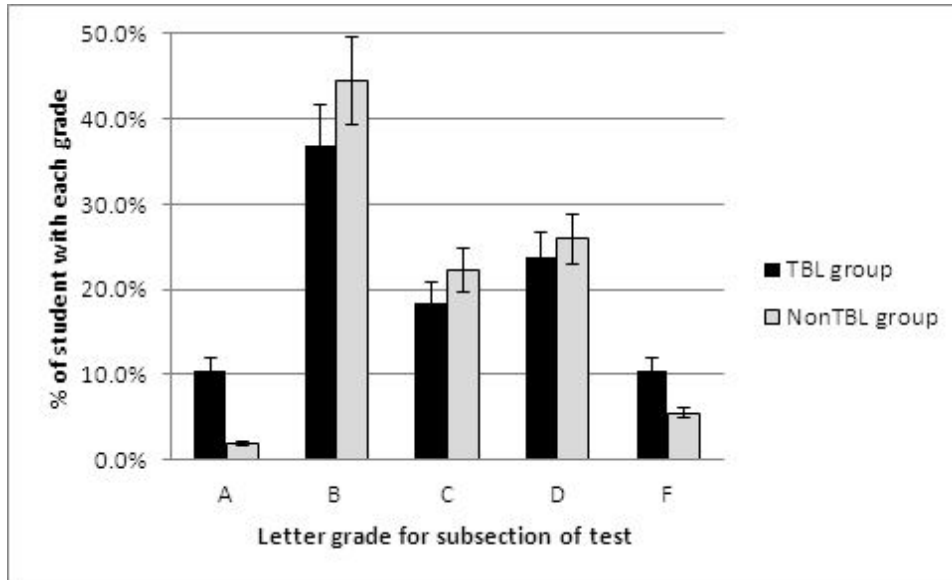


Figure 1. Performance on the large animal urinary surgery subsection of the test showing percent of class with grades of A, B, C, D or F. Chi square comparison between TBL (n= 38) and non-TBL groups (n=55): $p=.0713$; error bars depict standard deviation. The TBL students were overrepresented on both ends of the spectrum.

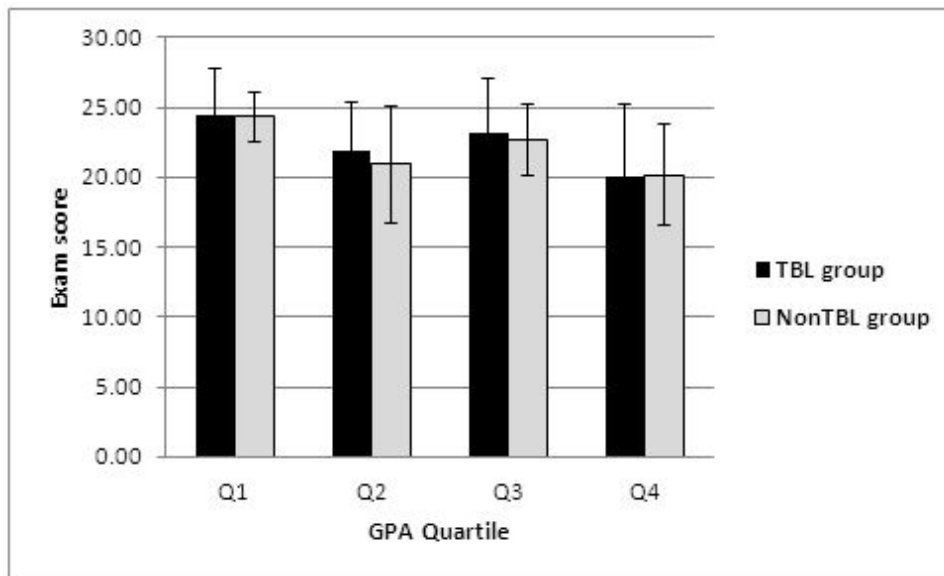


Figure 2. Exam scores by group and class rank quartile. Q1 = lowest, Q4= highest quartile; n=38 TBL, n=55 non-TBL. Exam scores are raw data (2 points per question, 100% =30 points); error bars depict standard deviation. No significant difference was seen by class quartile between groups.

D. Group performance.

Substantial improvement was noted between IRAT scores and GRAT scores for individuals (Figure 3). Improvement in scores was noted for 81.6% (31/38) students. Four students had identical scores and three had decreased scores (-1) on the GRAT compared to the IRAT. Two out of six groups had perfect scores on the GRAT (range 7-10) while no students had perfect scores on the IRAT (range 4-9). Two of the three students with lowered GRAT scores were in the same group; these students decreased scores from 8 points to 7 points. This group also had two students with the lowest IRAT scores (4 and 5). Twelve students improved their scores by 2 points (mode). However, the correlation between the IRAT score and the final exam score (34.9%) was better than that between the GRAT score and the final exam score (14.4%).

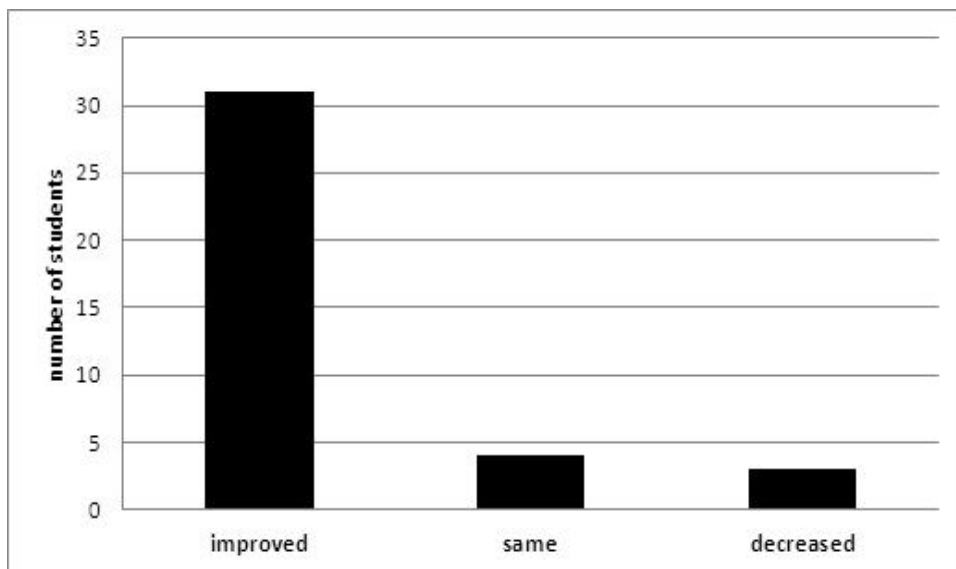


Figure 3. Change in scores between IRAT and GRAT by individual student. Total n=38. Improvement mode was 2/10 points. Three students had decreased scores; two of these students were in the same group.

As part of an associated study, we monitored group function using a previously validated survey (unpublished data; Appendix 2). The results of that work showed the biggest disparity was between one of the two groups with perfect GRAT scores and the group with the lowest GRAT scores. Many groups used consensus to reach an answer for the GRAT instead of majority votes.

E. Course evaluations.

No significant difference was noted between course evaluation scores from the previous year. Instructor evaluation scores were the same or higher than the previous two years. Due to the anonymous collection method, evaluation responses could not be stratified by TBL group. However, students included favorable comments about the new teaching methodology in midcourse and end of course evaluations.

The post-course survey solicited opinions on the level of learning and the TBL design (Appendix 3). Results showed students were generally cognizant of their level of knowledge (6/31 remarked they were surprised by their level of knowledge) and were comfortably challenged by the in class material (25/31; 81%). Six students reported being overwhelmed by the class material. Only one student felt it was not easy to ask questions about the material in class, five thought it was not very easy but was doable, and 26 felt it was easy to ask questions. In terms of applicable knowledge, 17/31 (55%) felt they could readily apply their knowledge to a case and an additional 11 thought they probably could but were not as confident.

Regarding team function, 27 students reported it was very easy to express their opinions within their team and the other four reporting it was somewhat easy. A large majority (28/31; 90%) felt their team accomplished a lot in terms of better understanding the material and no one reported that they became more confused after team discussions. When asked about the TBL format, 20 of the 31 that responded reported they definitely liked the format with an additional 10 reporting it was okay. One student would rather have had a different format. In the preclass survey given to the students during the informed consent meeting, one student reported regularly working with someone else and 25/32 (78%) said they sometimes study with others.

Prior to class, the TBL students admitted to significant off task work during class, with 41% using social media at least once during most classes, 31% frequently using social media during class, and 28% always accessing social media during class. Only 5/32 (16%) reported never using social media during class. During TBL, 24/31 (78%) reported using no social media during class.

Significantly more instructor time was required to perform administrative tasks required for TBL than for the standard course section, particularly as minimal work was required for the standard section of the course (updating of notes and PowerPoints only). This included setting up written and PowerPoint versions of the IRAT/GRAT questions, development of more advanced problems, developing peer evaluations, ensuring registration of clickers, and reassuring students of appropriate class times and responsibilities. Students frequently asked to verify the date of the classes and which class they were to attend. Instructor engagement and satisfaction with the actual class sessions was high for both groups; the higher student engagement in the TBL session was noticeable and rewarding.

F. Retention analysis.

Thirty-nine students participated in the repeat examination, 19 from the TBL group and 20 from the non-TBL group. Students in the TBL group significantly outperformed students in the non-TBL group (18.4 ± 4.0 vs 15.4 ± 4.6 ; total 30 points; $p= 0.0305$). This equated to a drop of 18.6% from the initial examination scores for the TBL group and a drop of 29.7% for the non-TBL group (Figure 4). Significantly higher scores for TBL students were noted for 3 questions, two of which reflected the topics covered during in-class case studies. No differences were noted between groups in terms of large animal interest or number of large animal rotations experienced.

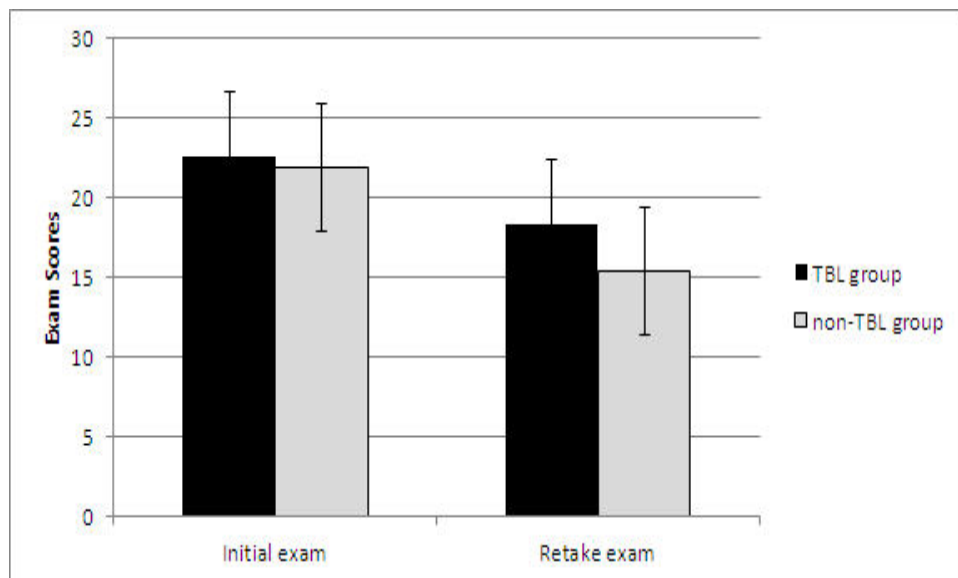


Figure 4. Comparison of scores between the course examination and the same examination taken 18 months later. Error bars denote standard deviation. No significant differences were noted between groups on the initial examination ($p=0.3984$; $n=38$ TBL, 55 non-TBL). However, scores for both groups significantly decreased over time (p values ≤ 0.002) and the scores for the non-TBL group were significantly lower on the repeat examination than for the TBL group ($p=0.0305$; $n=19$ TBL, 20 non-TBL).

IV. Discussion.

We were able use TBL successfully in a subsection of a veterinary course, with high levels of student and instructor satisfaction and with equivalent student test scores as compared to a control group taught by the same instructor. Knowledge retention appeared to be improved with the TBL format. We found similar strengths to TBL as others have: it allows highly interactive classes with large student numbers and with no increase in student work time and with at least similar student learning. The material that was standardly covered in three hours of lecture was covered in 1.5-2 hours of independent preclass work followed by 1.5 hours of in-class discussion. However, the amount of preparation time involved made it less appealing to convert already prepared instructional units of shorter duration.

After the TBL session (and before the exam), students reported feeling comfortable with the material and the majority thought they could apply it to an actual case. Overall test scores for the course were similar between groups. In contrast to other studies, we did not find either the lower performing or higher performing students did better with TBL (Letassy, Fugate, Medina, Stroup, & Britton, 2008; Koles, Stolfi, Borges, Nelson, & Parmelee, 2010; Wiener, Plass, & Marz, 2009). In fact, our TBL students tended to have a wider range of scores (higher and lower) regardless of class rank. The cause of this is unknown. Students may have felt more comfortable with the material and not reviewed as much for the test and/or students may not have learned the material as well that wasn't covered in the cases. Having more than one TBL session may enable students to estimate their level of self-directed learning more accurately. It is also possible that professional students may not follow this pattern as well as other groups, perhaps related to the higher level application-type problems. However, TBL is designed to be used in realistic,

complex situations and most TBL studies reflect this type of course design. Obviously test scores do not equate well to performance in clinics or in practice (Epstein & Hundert, 2002). Greater knowledge retention in the TBL group would suggest they may be better able to manage clinical cases related to the material, but it is impossible to assess true performance due to the high variability in student caseload exposure. Retention and usefulness is known to be enhanced if the learning context is similar to that required for use and with repeated test taking (Peile, 2006; Pyc & Rawson, 2010). The TBL format includes both and gives reason for optimism, particularly as students seemed to retain the material used most intensively in the classroom application exercises.

We found female TBL students outperformed male TBL students, but our overall numbers are small. Others have found no difference in small-group learning or even found females performing more poorly with TBL (Wiener, Plass, & Marz, 2009; Springer, Stanne, & Donovan, 1999). Comparing performances of men and women students in other professional programs is recommended to determine if sex differences truly exist.

Peer instruction and peer review is a key component of TBL and works to ensure students attend class and are prepared (Sweet & Pelton-Sweet, 2008). One of the disadvantages of using TBL in a subsection of a course is the lack of time for groups to mature (Sweet & Michaelsen, 2007). As all first year veterinary students take a year-long course in leadership and teamwork, we hoped those skills would translate into this course. We saw improvements in student work similar to that expected with TBL, with over 81% of students improving from the IRAT scores, most students improving by two points or more (out of 10), and two groups obtaining perfect GRAT scores. Students commented that their understanding was enhanced by group discussions. However, the final exam scores more closely correlated with IRAT scores than with GRAT scores, suggesting exam performance was not strongly linked to group performance in our course. On the other hand, this pattern has been seen with other TBL studies (Letassy, Fugate, Medina, Stroup, & Britton, 2008; Nieder, Parmelee, Stolfi, & Hudes, 2005). One team was obviously struggling with group function and their GRAT scores reflected this; we would hope that this group would have improved over time with more opportunities for practice and feedback.

Students reported higher engagement and less off-task work; groups were able to handle more challenging questions, and peer instruction was effective in answering many of the questions. Even more exciting was providing time for students to think and problem-solve without adding curricular hours or extra work. The format also allowed the instructor to spend limited class time on the more challenging topics as identified by responses to the IRAT and GRATs.

As reported previously for TBL, the instructor found the TBL session to be added work but more enjoyable than the standard classes. Since lecture time is short and related to student performance on the RATs, there is minimal set up for class discussions. Class sessions do require a small amount of set up to organize RAT grading and group materials. Grading is readily facilitated by classroom response systems or IF-ATs (Immediate Feedback Technique forms; Epstein Educational Enterprises, Cincinnati, Ohio). However, TBL requires careful design of RATs and complex problems. Instructors must think carefully about the desired course outcomes to work backwards to develop related course activities. This requires substantial advanced planning and cannot be successfully accomplished at the last minute.

As an added benefit, we found veterinary students open to the idea of being involved in new teaching methodologies and commented positively on the midcourse and end of course

evaluations. Many more students agreed to be in the study section than was anticipated. Students that committed to the TBL session were obviously nervous about their roles; however, all but two did the preclass work on the course website and all but one attended the class session. This was much different from their self-reported study habits. One major challenge was estimating how much material the students could cover on their own prior to class. The course website included links to many related articles that were not used. In retrospect, students did tend to review both the notes and the PowerPoint prior to the class, and this would likely take at least as long as covering the PowerPoint during lecture sessions. On the other hand, the notes did contain material that was not discussed in the standard lecture format and could have been formatted so that students could better see the high points and read those areas only if interested. The main questions students asked were related to what they should study. This is in congruence with other reports suggesting structure is necessary for proper learning, and providing this structure in terms of objectives or study questions is useful for independent learning (Sibley & Parmelee, 2008).

Limitations to this study include both small numbers and potential bias due to instruction by the investigator. Repeating the study over multiple classes would be recommended but impractical. Due to curriculum changes, repeating the study on the same course material is no longer possible. We believe the smaller numbers in our study are offset by the controlled study design that enabled us to evaluate learners within a single cohort. Investigator taught studies have been shown to be biased toward more significant findings (Springer, Stanne, & Donovan, 1999). It is possible we biased the control group in a negative direction; however, the classes have been taught in a similar manner for many years, and all attempts were made to minimize changes for this class. It was more difficult to avoid introducing a positive bias to the control group due to the need to focus learning objectives for the TBL classes; this focus carried over in the instruction for both groups. TBL exposure has been found to improve student performance overall (Wiener, Plass, & Marz, 2009). However, our section was included in the final test of the course and the TBL section was scheduled for the very end of the semester, minimizing the likelihood of any secondary effects. On the other hand, the leadership training given to all veterinary students may have prepared them better for effective group work and made it possible for them to work well together in a single session. Finally, the test questions were created for this year, and bias may have been introduced toward or against one study group. However, a set of study questions was provided to both groups and test questions drawn from that material in an attempt to minimize variation. In general, we recommend TBL for larger units of courses or for any new course unit. We found we were able to successfully run the class in a TBL format despite only having three hours of regular contact time (1.5 hours of TBL time). However, the unique format required advanced preparation for both students and instructors and created stress related to the differences from regular coursework. In the future, we will spend more effort focusing student reading requirements on the core material and will include more short videos instead of just providing the PowerPoint slides, as students felt overwhelmed and found it harder to review the slides on their own. With more courses using TBL or more sessions in a single course, students should likely be less worried about the course design and their roles. With larger instructional units, the general preparatory work (creating group folders, peer review forms, etc.) would be fewer hours in relation to contact time; repeated offerings of TBL are reportedly much less time-consuming, as well (Thompson et al., 2007; Mennenga & Smyer, 2010). Students may also be more successful with more practice with self-study and with enhanced group dynamics, particularly if they do not have previous training in effective group skills. In the future, we plan to convert course sections or courses to TBL if there are at least three classroom sessions

(equivalent to at least nine hours of regular contact time) to balance the set up demands and allow groups to mature. We have already started to use TBL for new instructional units no matter the length and have been working with colleagues to assist more widespread adoption of TBL in the veterinary curriculum.

Acknowledgments

Supportive funding was obtained through the University of Minnesota Undergraduate Research Opportunities Program. The authors wish to thank Ms. Deborah Wingert and Dr. Laura Molgaard for their critical review of the manuscript and for assistance with making the study possible. Previous presentation: Lilly Conferences on College and University Teaching, September, 2011

Appendices

Appendix 1. Questions from the Urinary Systems Disorders Examination.

1. You are presented with a 2 day old colt that you suspect has a ruptured bladder. Which of the following is consistent with your diagnosis?
 - a. the foal strains to urinate and has a pulsating urethra
 - b. peritoneal and pleural effusion can lead to poor oxygenation
 - c. hypokalemia, hypoglycemia, and hyponatremia
 - d. serum creatinine is 2x abdominal fluid creatinine
2. You recommend surgery as soon as anesthesia can get set up in order to minimize adhesions from the intraperitoneal urine
 - a. True
 - b. False
3. For the same colt, what is recommended in terms of fluid supplementation prior to surgery?
 - a. use fluids low in sodium to correct hypernatremia
 - b. add dextrose to minimize cardiac abnormalities
 - c. use LRS to provide proper levels of calcium
 - d. add potassium since he is off feed
4. You are presented with a calf with an umbilical swelling. The swelling is reducible and the calf seems sensitive to palpation. The calf has a temperature of 103F. With deep palpation you note the structure seems to extend cranially. On ultrasound you do identify a thickened structure extending from the umbilicus to the liver. You suggest treatment will likely involve:
 - a. Removal of the entire infected urachus
 - b. Removal of the entire infected umbilical vein
 - c. Removal of the entire infected umbilical artery
 - d. Marsupialization of the infected umbilical vein
 - e. Marsupialization of the infected umbilical artery
5. You did surgery to repair an umbilical hernia on a calf. What might be included in your postoperative instructions?
 - a. keep on broad spectrum antibiotics (penicillin/gentocin) for 5 days postop
 - b. monitor temperature until suture removal; call if >99F
 - c. stall rest until suture removal (10-14 days) then turnout with only 1-2 calves
 - d. monitor incision daily until suture removal: call if any heat, pain or discharge
6. You are called out on emergency after a rough calving on a nearby farm. They got the calf out before you arrived but it took a lot of work. You notice that the cow has a third degree rectovaginal tear (big open hole between the rectum and the vagina that extends about 6" deep). This is one of their best cows so they want to do what you think is best, Doc. You recommend:

- a. we should refer this cow to the UMN. She needs to have surgery done tonight to repair this tear or she will be contaminating her peritoneal cavity
 - b. we should refer this cow to the UMN. This type of tear needs to be repaired as soon as possible but it is tricky and best done in a sterile surgery room
 - c. we should fix this cow tonight; I can do it here if you can get me a good light. Luckily these tissues heal quickly so she has a good chance of coming out of this just fine
 - d. we should fix this cow if you want to breed her but we need to wait a few weeks. If I fix it tonight, it will just fall apart again. I am going to get her started on some antibiotics and flunixin meglumine and will check on her again in a couple of days
7. Treatment of pneumovagina in a mare would most likely include which of the following?
- a. Regular aspiration of air
 - b. Performance of a Caslicks
 - c. Ensuring ventral drainage
 - d. Creation of a cervicopexy
 - e. Insertion of a Buhner stitch
8. Which of the following steps is incorrect when performing a caslicks?
- a. secure the tail out of the way by attaching it to the horse's neck
 - b. inject lidocaine along the edges of the vulvar lips
 - c. trim a small margin of tissue from each side of the vulva to create a fresh edge
 - d. appose the fresh edge to the other side of the vulva with a simple continuous pattern
 - e. leave a small opening proximally to allow urination
9. The perineal body is
- a. benign growth that is frequently identified in maiden (unbred) alpacas
 - b. the reason you can't pass a urinary catheter into the bladder of a goat
 - c. a structure that provides support to the dorsal aspect of the vaginal vault
 - d. a remnant of a mesonephric duct, commonly found in XXY animals
10. You are a food animal vet and go out to check on your neighbor's herd of beef heifers. He is a gentleman farmer and isn't at home much and doesn't know much about cattle (they are mostly a tax write off). One has a large swelling protruding from her vulva. He reminds you he expects them to start calving in the next month. After getting the swelling back in where it belongs, you recommend which of the following :
- a. You recommend a C section now and then shipping the cow as she is likely to prolapse her uterus after calving
 - b. You recommend a C section now as the cow is going to die soon and you can try to salvage at least one of the pair
 - c. You will perform a Buhner stitch and everything should be good until after the calf is born
 - d. You will perform a Buhner stitch and take this one to your barn for close monitoring until after the calf is born

11. You are presented with a 6 month old intact (not castrated) bull calf, intended for breeding, with a 4 day history of stranguria. You palpate pulsations along the ischial urethra. Ultrasound reveals a grossly enlarged bladder. Which would provide the best chance of resolving the urethral obstruction while preserving the bulls breeding potential?
 - a. retrograde (eg from penis end) urethral catheterization and urethral lavage
 - b. ischial urethrostomy and placement of a catheter to keep the bladder empty
 - c. amputation of urethral process tonight with cystotomy tomorrow
 - d. celiotomy for placement of a Foley catheter within the bladder
 - e. placement of a chest drain to remove the urine and iv fluids to flush the bladder

12. You diagnose a bladder stone in a horse and successfully remove it via a perineal urethrotomy approach. Which of the following is most likely to be included in your discussions with the owner?
 - a. management of the associated postrenal azotemia
 - b. dietary suggestions about their current use of a grass-alfalfa mix hay
 - c. the risks of body wall herniation after this type of surgery
 - d. clinical signs such as stranguria that will alert them to a new stone
 - e. the technique of urethral dilatation if they catch the problem early enough

13. You suspect a ruptured urethra in a feedlot steer. Treatment would primarily require :
 - a. correcting electrolyte abnormalities
 - b. lancing the skin in the area to allow drainage
 - c. inserting a chest tube to drain the urine
 - d. no treatment; just send him directly to slaughter

14. Your client's favorite stallion gets kicked by a mare. You haven't seen it yet but they call wanting to know if it warrants an emergency call. You answer
 - a. it can wait. Stallions usually don't rupture the tunica albuginea as happens in bulls
 - b. it can wait. However, if he develops a persistent erection, they should call you right away
 - c. most definitely. Better to be safe than sorry. Sometimes they damage the urethra and that would lead to impaired fertility
 - d. most definitely. Trauma to that area leads to swelling that could eventually make him infertile

15. A beef breeding bull develops a large, firm, swelling surrounding the penis, just cranial to the scrotum. The bull is unable to extend the penis. What is the most likely cause of these findings?
 - a. hematoma involving rupture of the tunica albuginea
 - b. squamous cell carcinoma in the region of the fornix
 - c. urethral rupture and leakage of urine into tissues surrounding the penis
 - d. seroma associated with lymphatic blockage
 - e. granulomatous response to a foreign body in the fornix

Appendix 2. Classroom engagement survey questions.

Lecture	Group	
mins: 10 20 30 40 50 60 70 80 90 100 110 120		
Quadrant: 1 2 3 4		
		YES
general participation (raising hand, etc)		
listening to instructor (head nod, eye contact, etc)		
listening to group		
talking to group (on topic)		
talking to group (off topic)		
asking instructor questions		
answering questions		
working on other homework/studying for other classes/ busywork		
reading		
organizing ideas		
writing		
texting		
on computer (on task)		
on computer (off task)		
slumped back in chair		
leaning forward, engaged		
chairs moved so that everyone in group is included		
chairs moved so that most of group is involved		
1-2 outsiders in group		
no furniture moved		
hostile debate in group		
friendly debate in group		
even voice distribution		
1-2 people running the show		
group has divided into sub groups		
everyone in group is taking equal ownership for ideas		
eye contact		
paying attention to group		
paying attention to other groups		
attention has been lost		
asking peers questions		
using "majority rules" to make decisions		
making decisions based on best reasoning		

Appendix3. Student evaluations of TBL section.

	# responses
<i>I was surprised at how much I already knew going into class (before discussion)</i>	
Definitely surprised	6
Maybe a bit	16
Not at all; I knew what I knew	9
<i>I reviewed material prior to class</i>	
0hr	0
<1hr	2
1-2 hrs	22
2-3 hrs	6
>3 hrs	1
<i>I used the supplementary readings (not just notes/powerpoints)</i>	
Yes, I read more than one	2
Yes, I read one	2
No, stuck to notes/powerpoints	27
<i>During class, I felt</i>	
I was not challenged at all	0
Comfortably challenged	25
In way over my head	6
<i>I feel that I could apply the material to a real life case</i>	
Yes, definitely	17
Maybe	11
Not without more help	2
<i>I feel that it was easy to ask questions</i>	
Very easy	26
Not very easy, but doable	5
Not at all	1
<i>I felt that it was easy for me to express my ideas within my team</i>	
Very easy	27
Somewhat easy	4
Difficult	0
<i>I feel that my team accomplished</i>	
A lot in terms of better understanding material	28
Not a ton. I feel that I know about as much as before I met with my group	3
Nothing. The class was counterproductive. I am more confused.	0
<i>I felt that</i>	
My group was very focused during the entire class	25

Was focused, for the most part	5
Was unfocused	0

Overall, I liked the class format

Yes, definitely	20
It was okay	10
I would rather have had a different format	1

Other things I (probably) did during class (Check all that apply)

None	26
Email	4
Facebook	3
Study for other exams	
Played games/did crossword	

If I did others things during class, it was most likely for

< 1/4 of class	6
1/4-1/2 of class	
1/2-3/4 of class	
> 3/4 of class	

What I liked:

- It was more interactive than a typical lecture.
- Open discussions helped hold attention
- Being able to talk the cases through
- The ability to apply knowledge and further clarify topics
- Small groups
- Flexibility to study
- Allowed for multiple viewpoints to be considered
- Teamwork allowed for better understanding
- Easy to ask questions
- Fun
- Less pressure
- Real life scenarios

What I disliked:

- Combined with the out of class work, it was longer than we'd spend in a normal class
- The powerpoints/notes were sometimes hard to understand on their own.
- Not enough time to read the lecture beforehand/ lots of information to get through prior to class
- Some people were not on task and caused disruption
- The independent quizzes
- The suggested 90 min review of the material prior to class was not enough time
- The amount of material we had to get through
- Lots of material that we didn't touch on was in the notes we were told to review before class

References

- Davis, M.H., and Harden, R.M. (1999). AMEE Medical Education Guide No. 15: Problem-based learning: a practical guide. *Medical Teacher*, 21(2), 130-140.
- Epstein, R.M., and Hundert, E.M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, 287(2), 226-235.
- Goldberg, H.R., and Dintzis, R. (2007). The positive impact of team-based virtual microscopy on student learning in physiology and histology. *Advances in Physiology Education*, 31(3), 261-265.
- Giuliodori, M.J., Lujan, H.L., and DiCarlo, S.E. (2006). Peer instruction enhanced student performance on qualitative problem-solving questions. *Advances in Physiology Education*, 30(4), 168-173.
- Haidet, P., O'Malley, K.J., and Richards, B. (2002). An initial experience with "team learning" in medical education. *Academic Medicine*, 77(1), 40-44.
- Koles, P.G., Stolfi, A., Borges, N.J., Nelson, S., and Parmelee, D.X. (2010). The impact of team-based learning on medical students' academic performance. *Academic Medicine*, 85(11), 1739-1745.
- Letassy, N.A., Fugate, S.E., Medina, M.S., Stroup, J.S., and Britton, M.L. (2008). Using team-based learning in an endocrine module taught across two campuses. *American Journal of Pharmaceutical Education*, 72(5), 103.
- Levine, R.E. (2008). Peer evaluation in team-based learning. In L.K. Michaelsen, D.X. Parmelee, K.K. McMahon, & R. Levine (Eds.). *Team-Based Learning for Health Professions Education* (pp. 103-116). Sterling, VA: Stylus.
- Mennenga, H.A., and Smyer, T. (2010). A model for easily incorporating team-based learning into nursing education. *International Journal of Nursing Education Scholarship*, 7(1), Article 4.
- Michaelsen, L.K., and Sweet, M. (2008a). Fundamental principles and practices of team-based learning. In L.K. Michaelsen, D.X. Parmelee, K.K. McMahon, & R. Levine (Eds.). *Team-Based Learning for Health Professions Education* (pp. 9-31). Sterling, VA: Stylus.
- Michaelsen, L.K., & Sweet, M. (2008b). The essential elements of team-based learning. In L.K. Michaelsen, M. Sweet, & D.X. Parmelee (Eds.). *Team-Based Learning: Small Group Learning's Next Big Step: New Directions for Teaching and Learning* (No. 116, pp. 7-27). Hoboken, NJ: Wiley.
- Nieder, G.L., Parmelee, D.X., Stolfi, A., and Hudes, P.D. (2005). Team-based learning in a medical gross anatomy and embryology course. *Clinical Anatomy*, 18(1), 56-63.

Peile, E. (2006). Commentary: Integrated learning. *British Medical Journal*, 332(7536), 278.

Pickrell, J.A., Boyer, J., Oehme, F.W., Clegg, V.L., and Sells, N. (2002). Group learning improves case analysis in veterinary medicine. *Journal of Veterinary Medical Education*, 29(1), 43-49.

Pileggi, R., and O'Neill, P.N. (2008). Team-based learning using an audience response system: an innovative method of teaching diagnosis to undergraduate dental students. *Journal of Dental Education*, 72(10), 1182-1187.

Pyc, M.A., and Rawson, K.A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330(6002), 335.

Root Kustritz, M.V., and Nault, A.J. (2010). Professional development training through the veterinary curriculum at the University of Minnesota. *Journal of Veterinary Medical Education*, 37(3), 233-238.

Sibley, J., and Parmelee, D.X. (2008). Knowledge is no longer enough: Enhancing professional education with team-based learning. In L.K. Michaelsen, M. Sweet, & D.X. Parmelee (Eds.). *Team-Based Learning: Small Group Learning's Next Big Step: New Directions for Teaching and Learning* (No. 116, pp. 41-53). Hoboken, NJ: Wiley.

Springer, L., Stanne, M.E., and Donovan, S.S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, 69(1), 21-51.

Sweet, M., and Michaelsen, L.K. (2007). How group dynamics research can inform the theory and practice of postsecondary small group learning. *Educational Psychology Review*, 19(1), 31-47.

Sweet, M., and Pelton-Sweet, L.M. (2008). The social foundation of team-based learning: Students accountable to students. In L.K. Michaelsen, M. Sweet, & D.X. Parmelee (Eds.). *Team-Based Learning: Small Group Learning's Next Big Step: New Directions for Teaching and Learning* (No. 116, pp. 29-40). Hoboken, NJ: Wiley.

Thompson, B.M., Schneider, V.F., Haidet, P., Perkowski, L.C., and Richards, B.F. (2007). Factors influencing implementation of team-based learning in health sciences education. *Academic Medicine*, 82(10 suppl), S53-S56.

Thompson, B.M., Schneider, V.F., Haidet, P., Levine, R.E., McMahon, K.K., Perkowski, L.C., and Richards, B.F. (2007). Team-based learning at ten medical schools: Two years later. *Medical Education*, 41(3), 250-257.

Wiener, H., Plass, H., and Marz, R. (2009). Team-based learning in intensive course format for first-year medical students. *Croatian Medical Journal*, 50(1), 69-76.

Examining service-learning in a graduate physical education teacher education course

Karen S. Meaney¹, Jeff Housman¹, Arnoldo Cavazos¹, and Michelle L. Wilcox¹

Abstract: This study was designed to explore the impact of service-learning on graduate physical education teacher education students. Social-Cognitive Theory (Bandura, 1986; 1999) served as the framework to examine graduate student's experiences in a service-learning program. Participants were graduate students (N =16) enrolled in a curriculum and instruction in physical education course at a major university in the southwest United States. The course's service-learning component provided graduate students opportunities to teach physical activity to Hispanic-American and African-American children from low-socioeconomic backgrounds. Participant's described their experiences through weekly reflections and discussions. Content analysis of data sources indicated that participation in the service-learning program strengthened graduate student's efficacy for teaching, contributed to their acquisition of varied teaching strategies, and enhanced graduate students understanding of children living in low-income, minority households. Findings suggest service-learning can be a valuable pedagogy to infuse into graduate teacher education programs.

Keywords: experiential-learning, sport pedagogy, social-cognitive theory

Service-learning is a teaching technique that bridges academic study and civic engagement. Specifically, this dynamic pedagogy enables students to partake in meaningful community service directly related to their academic course content. This learning environment enables students to validate theoretical and often times complex principles discussed in class to authentic settings. These experiences are reinforced through critical reflection activities, which assist in connecting the service experience to academic subject matter (Cress, 2005).

John Dewey's work in the early 1900s examining the role of higher education in citizenship development initiated dialogue and the theoretical foundation for service-learning (Dewey, 1938); however, it wasn't until 1984, when David Kolb transformed Dewey's six-step inquiry process into a four-component learning cycle for experiential learning that service-learning curricula received considerable attention (Kolb, 1984).

The well-documented benefits of student participation in service-learning include favorable influences on student's personal outcomes such as heightened levels of self-efficacy, identity, and moral development; and social outcomes including reduced stereotyping and enhanced diversity appreciation (Eyler, Giles, Stenson, & Gray, 2003). Teacher education programs have utilized service-learning to prepare future teachers in numerous settings and academic disciplines (for a review see Anderson, Swick, & Yff, 2001). Early investigations examining the impact of service-learning on pre-service educators in a variety of content areas have revealed advantageous outcomes in regards to the pre-service educator's self-esteem and self-efficacy for teaching (Wade, 1995), dedication to the teaching profession (Green, Dalton, &

¹ Department of Health and Human Performance, Texas State University, San Marcos TX, 78666, km66@txstate.edu; housman@txstate.edu; ac1206@txstate.edu; lani.wilcox@gmail.com.

Wilson, 1994), increased diversity awareness (Tellez, Hlebowitsh, Cohen, & Norwood, 1995), and overall affirming learning experiences (Wade & Yarbrough, 1997). In regards to assessing the relationship of service-learning and academic outcomes, evidence supports the positive impact service learning has on critical thinking skills and academic engagement (Root & Swick, 2001).

Service-learning programs have also been investigated in the discipline of physical education teacher education (PETE) (Cutforth, 2000; Domangue & Carson, 2008; Kahan, 1998; LaMaster, 2001; Meaney, Bohler, Kopf, Hernandez, & Scott, 2008; Meaney, Griffin, & Bohler, 2009; Watson, Crandall, Hueglin, & Eisenman, 2002). Findings revealed that undergraduate students participation in physical education service-learning programs result in increasing perceived competence for teaching (LaMaster, 2001), developed moral reasoning (Cutforth, 2000) strengthened cultural competence (Domangue & Carson, 2008; Meaney et al., 2008), and enhanced pedagogical content knowledge (Meaney, Griffin, & Bohler, 2009).

Within the discipline of PETE service-learning research has been limited to examining undergraduate student experiences (Cutforth, 2000; Domangue & Carson, 2008; Kahan, 1998; LaMaster, 2001; Meaney, Griffin, & Bohler, 2009; Meaney et al., 2008; Watson et al., 2002). Little, if any research has examined the potential advantages and/or disadvantages of incorporating service-learning into graduate PETE programs.

Graduate and undergraduate students differ in a number of social and academic characteristics. In addition to graduate students possessing an undergraduate degree, graduate students are generally older; a significantly higher percentage of graduate students as compared to undergraduate students are married, and 75% of graduate students work full time (American Council on Education, 2005; United States Census Bureau, 2007). Furthermore, research examining personality and psychological comparisons of graduate and undergraduate education students revealed that differences in personality traits between graduate and undergraduate students may influence preferred ways of learning (Illovsky, 2010). Consequently, pedagogical strategies (i.e., service-learning) and teaching techniques that are effective with undergraduate students may not be as valuable with graduate students.

To our knowledge, incorporation of service-learning in graduate PETE courses has not been explored. However, investigations examining service-learning in graduate course work have been conducted in the disciplines of public administration, instructional technology, and early childhood (Brescia, Mullins, & Miller, 2009; Liu & Lambright, 2010; Quinn, 2006). Participation in a project based instructional technology service-learning course enhanced graduate students' problem solving skills while simultaneously nurturing their sense of caring for others (Brescia, Mullins, & Miller, 2009). Additionally, Liu and Lambright (2010) demonstrated that the hands-on experience provided through service-learning benefitted graduate students in the learning and acquisition of professional skills. The possibility exists that graduate students in PETE might also benefit from participation in a curriculum that incorporates service-learning.

I. Theoretical Framework.

Social-cognitive theory (Bandura, 1986; 1999) suggests that human learning occurs within a dynamic framework and initiates interaction between one's personal factors, environment, and behaviors. These dynamic relationships constitute an interactive model referred to as triadic reciprocity (see Figure 1).

Within the model of triadic reciprocity, personal factors may include one's motivation, self-efficacy, knowledge, fears, and expected outcomes. The environment is perceived in three stages: imposed, selected, and constructed (Bandura, 1999). One's imposed environment includes the way things are, that is situations an individual must interact with on a daily basis (i.e., school, work, family). While individuals may have minimal influence over imposed environmental factors, they do have choices in how they interpret and react to imposed factors. These choices regarding how one reacts to the imposed environment constitutes the selected environment. The resulting behaviors, the third aspect of triadic reciprocity, become one's constructed environment. Construction of one's environment demands actively engaging in one's surroundings and may often result in the acquisition of new knowledge, beliefs, and behaviors.

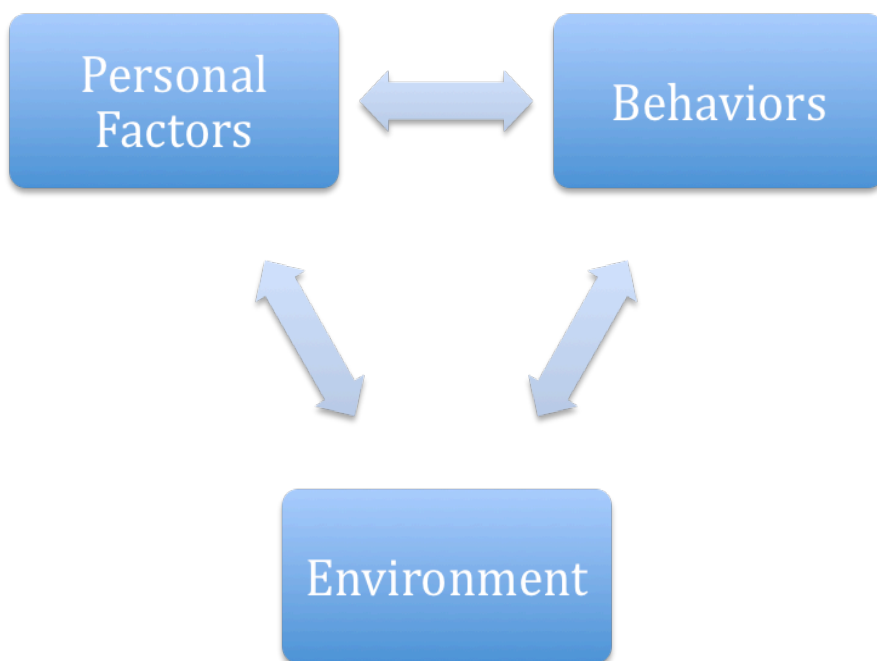


Figure 1. A model diagramming triadic reciprocity. (Adapted from Schunk [2012])

Social-cognitive theory provides researchers a paradigm to examine service-learning experiences. Previous research examining the impact service-learning has on undergraduate students in PETE has used social cognitive theory as a framework (Meaney, Griffin, & Bohler, 2009; Meaney et al., 2008). The purpose of this investigation was to extend the research encompassing service-learning in PETE programs through the examination of graduate student experiences in a service-learning program. Specifically, we explored the following research questions: a) what impact did teaching physical activity to low-income minority children have on graduate students' self-efficacy for teaching b) what effect did teaching physical activity to low-income minority children have on graduate students' cultural competence, and c) did participation in the service-learning program enhance graduate students' application of pedagogical content knowledge?

II. Methods.

A. Participants.

Graduate students enrolled in a graduate curriculum and instruction in physical education course in the Department of Health and Human Performance at a major university in the southwest United States volunteered to participate in this study. Participants included 16 graduate students (n = 11 females, n = 5 males; mean age = 25.6 years) who were enrolled in a Degree and Teaching Certification program culminating in a Masters of Education specializing in Teaching Physical Education. All of the participants held undergraduate degrees; eight were employed fulltime (i.e., 40 hours per week), 5 were employed part-time (20 - 30 hours per week), and three students were not employed.

B. Course.

The primary goals of the graduate curriculum and instruction course centered on enhancing graduate student understanding of the need for evidenced-based curriculum programs and instructional strategies in physical education. Furthermore, the course was intentionally designed to provide students multiple opportunities to design, implement, and assess physical activity instruction through participation in the service-learning component of the course. Examining pedagogical research and understanding learner variables (i.e., gender, socioeconomic status, cultural beliefs, values, geographic location) that influence participation in lifelong physical activity were also embedded throughout the course. Readings, discussions, and learning activities focused on increasing student's cultural competency for teaching. The graduate course met for three hours one night a week for a total of 15 weeks.

C. Service-Learning Program.

The service - learning component of the course enabled students to transfer theoretical concepts discussed in class into practice, while at the same time meeting local community needs. Working in teams, students planned, implemented, and assessed physical activity lessons for children between the ages of 5-12 years. The physical activity instruction was a component of the Family Fun & Fit (FFF) service-learning program. FFF is a collaborative project between the university and a local elementary school. Ford Elementary (pseudonym) is a Title I school; the majority of the children are Hispanic-American and participate in the free and reduced breakfast and lunch program; which is an indicator of the school's poverty level.

FFF was designed to be a win-win situation for everyone involved in the project. The graduate students enrolled in the curriculum and instruction class in physical education had multiple opportunities to instruct physical activity; the children enrolled in FFF could engage in developmentally appropriate physical activity in a safe, indoor environment at the school. In addition the parents and caregivers of the children engaged in health, wellness, and nutrition program delivered by students in the community health course offered by the same department via a companion service-learning project. FFF took place for two hours in the evening on a weeknight for three consecutive weeks.

D. Procedure.

Approval from the university's Institutional Review Board was obtained prior to the beginning of the academic semester. We described the FFF project during the first night of the graduate class and students were asked to volunteer to participate in the study. All students were informed that their participation was voluntary and no one would be penalized for not agreeing to take part in the investigation. We planned for the FFF service-learning project to occur between the 8th and 11th week of the semester. The graduate students collectively decided on three physical education curricula that had been presented, researched, and discussed throughout the semester to implement at FFF. The curricula were: Skill Themes & Movement Concepts (Graham, Holt/Hale, & Parker, 2009); Teaching Games for Understanding (Griffin & Butler, 2005) and Health Related Physical Fitness (Corbin & Lindsey, 2007). Working in three teams (i.e., two at five members each; one at six members) each team was responsible for delivering each curriculum one evening; all teams presented each night consequently three curricula were presented every session. In addition to creating the physical activity lesson, teams were also responsible for creating marketing materials, and registration forms.

The FFF program was held during weeks 12, 13, and 14 of the spring semester. Each night the graduate students arrived at least 30 minutes prior to the opening of FFF, prepared the registration materials, and organized the equipment and space where their physical activity lesson would occur. Three different teaching stations used throughout the duration of FFF included two stations in the elementary school gymnasium, as well as the outside play area.

Between 6:15 – 6:30 PM children and their families entered the cafeteria and registered and picked up their nametags. At 6:30 PM we welcomed FFF participants and re-emphasized the multiple purposes of FFF; 1) to provide children positive experiences in physical activity in a safe and welcoming environment, 2) to offer parents and caregivers opportunities to participate in culturally and economically relevant health and wellness education, and 3) to enable graduate student's real world experiences creating, delivering, and assessing physical activity instruction.

Following the welcome, the graduate students escorted their initial group of children to their teaching station and began the physical activity lesson. Each activity station lasted for 30 minutes and then the children rotated to the next station. When all of the children completed all three thirty - minute stations the children were re-united with their parents/caregivers in the school cafeteria for a closing session. During this time all FFF participants were offered a healthy snack (i.e., apple, orange, or banana) and also participated in a question and answer period led by different graduate student teams each night. The questions and answer period was designed to explore what the children and adults learned that night as well as what they enjoyed participating in or discussing. The closing session was also used to distribute door prizes to the participants. The graduate students had secured university t-shirts, water bottles, caps, and lanyards from various departments (i.e., athletics, physical education, bookstore) at the university to distribute to the FFF participants.

III. Data Collection and Analysis.

A. Weekly reflections and discussion group interviews.

Participants described their experiences with the children through weekly reflective narratives as well as verbal discussion groups. Group discussions were audio recorded and transcribed

verbatim. The week prior to the beginning FFF we asked graduate students to respond to specific open-ended questions via written reflective narratives. Questions were based on Social-Cognitive theory (Bandura, 1986; 1999) and designed to explore their personal factors (thoughts and feelings), their perceptions of the teaching environment (low-income Hispanic-American children), and the interaction of their personal factors and environment on their teaching behaviors (planning, instructional techniques, assessing). We conducted content analysis to identify categories from the reflections and discussion groups according to Patton's (2002) methodology.

We (two lead investigators) and two graduate students (X and Y) independently read copies of the reflections and discussion group notes several times to identify patterns and categories in the content. Following the preliminary analysis, the primary investigator and graduate student X and Y re-read—narratives and the discussion group notes. Individually, we all highlighted statements deemed salient and then independently identified patterns and formulated categories of the data sources.

Following the independent analysis, we collectively reviewed and discussed categories derived from the data. These discussions enabled us to identify and describe categories consistent among the four researchers. We then identified themes based on the following criteria: (a) Each researcher identified the category independently, and all three identified the same category; (b) the same category was identified in narratives, and discussion groups; and (c) the category was acknowledged in 60% of the responses to the directed reflective narratives.

Trustworthiness was established through confirmability audits, negative case analysis, (Lincoln & Guba, 1985) and multiple investigators (Denzin, 1970). Following the completion of the qualitative analysis, an external reviewer conducted a confirmability audit. The external reviewer was asked to place quotes from the narratives of the previously established themes. A total of 30 quotes were included in this review process and the external reviewer placed 26 quotes in the previously identified themes. Consequently, the inter-rater reliability for the external reviewer and researchers was 87%. Triangulation of the data occurred via the multiple data sources (i.e., reflections and discussion groups), as well as multiple reviewers (Pitney & Parker, 2009).

IV. Findings.

Content analysis of the data revealed three major themes: 1) the graduate students began the service-learning program with a variety of affective feelings, 2) participation in the program enabled the graduate students to design, implement, and assess instructional strategies which enhanced their teaching efficacy, and 3) getting to know and interacting with diverse children altered the graduate students pre-conceived stereotypes of low-income families and their interests in health and physical activity. Table 1 displays the identified themes and categories, identifies along with the sub-categories, the number of participants for each theme, and frequency of responses within each sub-category. The written reflective journals were used to calculate the number of responding participants and frequency of responses. We interpreted findings through the social-cognitive framework (Bandura, 1986; 1999). All names presented in this section are pseudonyms.

Table 1. Results of Content Analysis: Themes, Categories, Sub-Categories, Frequencies.

Theme	Category	Sub-Category	# of Participants	Frequency of Statements	
Affective	Positive	Excited	6	4	
		Purposeful		1	
		Comfortable		1	
		Interested		2	
	Negative	Nervous		10	8
		Scary		1	
		Anxiety		4	
		Unsure		1	
		Worried		3	
Teaching Efficacy	Instructional Strategies	Show & Tell	11	6	
		Brief Instructions		5	
		Improvised Lessons		5	
		Adapted Activities		7	
		Monitored Learning Environment		4	
	Knowing Children	Developmental Differences		9	4
		Personalities		6	
		Connected		5	
	Effectiveness	Confidence		13	7
		Competent		3	
		Comfortable		5	
	Pre-Conceived Stereotypes	Low Expectations		Less Interested	12
More Overweight			4		
Not Physically Active			6		
Free Program			3		
Little Knowledge			7		

A. Affective feelings.

Social-cognitive theory (Bandura, 1986; 1999) underscores the interaction of one's personal factors (i.e., motivation, anxiety, self efficacy) with the environment (i.e., service-learning, FFF) and the impact these interactions may have on an individual's behavior (i.e., teaching physical activity). Analysis of the weekly reflections and discussion groups revealed that ten graduate students had negative thoughts and were nervous and apprehensive at the start of FFF. Nick stated; "I am a little nervous about meeting the families for the first time, I am also a little worried about the language complications we may have with some families." Denise shared; "The elementary school is a new environment and by nature that is going to make you nervous, I love meeting new people but I feel really nervous to meet the families for the first time because it's going to be a lot of people at one time." Kimberly expressed the following:

To meet the families for the first time is nerve racking. It is like jumping into the deep end of a swimming pool for the first time, I do not really know what to expect. Is it going to be hard? Are the children going to enjoy the activities? Do I know what I am doing? I do not want to look like a failure.

On the other hand six graduate students shared positive thoughts about beginning the Family Fun & Fit program. Sean stated "I have no apprehensions about our initial meeting; I am excited and comfortable to meet the children and families and demonstrate how fun fitness can be." Michelle shared:

Meeting the families for the first time is very exciting, language may be the only significant barrier, I feel a little nervous but I believe our class is up to the challenge. I know the best way to deal with this nervousness is to get right in the activities and be enthusiastic; despite being a little nervous I am very excited to be in the gym and see our semester in action. I am excited about seeing what works and what does not work.

Luke expressed the following; "I feel that I have developed a strong sense of leadership and comfort when it comes to working with children. I also feel confident in my abilities to get them (the children) excited and motivated about physical activity."

Acknowledging that graduate students embrace a wide array of affective emotions prior to teaching is an important component of social-cognitive theory (Bandura, 1986; 1999). Personal factors such as nervousness, anxiety, excitement, and motivation have the potential to positively and/or negatively influence the graduate student's teaching experiences. Faculty instructing graduate curriculum and instruction classes need to identify strategies to harness and funnel these emotions in ways that benefit both the pre-service educators and their students. For example, formulating teaching groups with a combination of highly anxious and highly confident graduate students may assist in balancing the emotional make up of the instructional group. Engaging in discussions addressing and acknowledging pre-teaching emotions may be valuable to all students.

B. Teaching efficacy.

The second major theme revealed through the analysis was that despite the graduate student's initial nervousness and anxiety the apprehension dissipated. In turn, it appears that thirteen of the graduate student's self-efficacy and confidence for teaching grew stronger throughout their

participation in the program. Weekly interaction with the children provided nine graduate students a venue to become familiar with the children. Brian stated:

When you get to know the children, you get a better idea of how to handle them to get them to participate. The first week, I had a small child Santiago who cried the entire time he was in our group. Eventually I got him to participate by doing it first and showing him how much fun it would be, I had to try several different ways of displaying this in order to convince him, but eventually, I found a way to connect with him. The second week, I immediately went to my successful strategy. Getting to know the children was essential to our success as facilitators.

Previous research examining undergraduate students experience in service-learning supports the important role undergraduates place on getting to know their students (Meaney et al., 2008). Likewise, nine graduate students also perceived this process as valuable. Becoming familiar with the children participating in FFF enabled the pre-service educators to feel “connected” to the students. Connecting with the children aided the pre-service educators with motivational and instructional strategies.

In addition to getting to know the children, participation in the weekly program created a forum for eleven graduate students to explore various teaching techniques. Martha stated; “I realized that giving short and quick to - the - point instruction about the activity worked better than trying to go into detail about different aspects because when we started talking too much the kids didn’t pay attention.” Effective communication techniques are critical components of the teaching-learning process (Rink, 2010). Beginning physical education teachers often struggle to find the appropriate amount of verbal instruction. Excessive instruction often minimizes student’s practice and participation time that can result in students less skilled performance. It appears that participation in the service-learning program enhanced eleven of sixteen (68%) graduate students understanding of the importance of effective instruction and demonstration techniques.

The graduate students also highlighted the importance of embracing flexibility and adaptation when instructing physical activity. Josie stated:

At first I felt overwhelmed with the activity because it was not going the way we had planned. We made a few adjustments and with these adjustments I began to see as a teacher you have to be more apt to change things that do not seem to be working. I was gaining hands on experience showing the reasons not to be rigid in my teaching and it was refreshing to see that I was able to adapt and have a good lesson at the same time.

Kirk revealed; “We improvised and changed the lesson because no matter how much we plan out every detail, things will never turn out exactly as planned.” Monitoring the learning environment and adapting the tasks and environment to best meet the individual needs of students is a critical component of effective physical education instruction (Rink, 2010). Teachers need to appropriately increase and/or decrease the level of task difficulty based on student’s motor skill performance. The graduate students instructing physical activity at FFF had repeated opportunities to take the lead in adapting activities to enhance their students’ success.

Participation in the service-learning program also positively impacted 13 of the graduate students’ efficacy towards teaching. Representative statements include the following; “We gained a little confidence after the first night (Josephine),” “Everyone was so much more confident and relaxed (Mark),” “I feel more capable of promoting physical activity (Katie),” “I

feel fairly competent in my own ability to practice the promotion of physical activity (Audra),” and “I feel my ability to promote physical activity will continue to grow with practice (Mark).”

Self-efficacy is a principle tenet of Social-Cognitive theory, (Bandura, 1986; 1999). Bandura described self-efficacy as one’s perceived capabilities towards completion of a particular task. In regards to teachers, Bandura suggests that teachers with high levels of self-efficacy will demonstrate greater levels of tenacity, engage students in more problem solving challenges, and work harder than teachers who are less efficacious. Consequently, providing pre-service teachers with opportunities to increase their self-efficacy for teaching is critical.

C. Pre-conceived stereotypes.

Analysis of the weekly reflections and discussion groups also highlighted the significant influence participation in FFF had on twelve of the sixteen (i.e., 75%) graduate students’ previously held beliefs and attitudes encompassing children and families from low income backgrounds. Stacey reported:

My perception of the families did change. Before we started the program, I thought the families would be coming to FFF because the program was free. However, the families were excited to learn about health and wellness. The children and adults were eager to participate in our activities and homework assignments.

Mark’s comments included:

My experience at FFF was eye – opening. I learned people from low socioeconomic backgrounds do not have the knowledge and resources to make healthy and smart choices on a shoestring budget. When we gave the families fun activities for the family games, and homework the families were able to engage in physical activity which they spend little to nothing to participate, I feel this was the most surprising thing I learned, if you give these families ideas, games, and resources they will feel like being as active as possible.

Jackie revealed:

I have to admit before FFF I wasn’t expecting the participant’s to be this energetic and cooperative. I wasn’t expecting these kids would actually enjoy running and sweating; many of them didn’t give up and stayed to the end of every game. These perceptions all changed after this program. The families who wanted this to work for their families were there early and stayed late. I feel this is most surprising because I did not expect families to be so committed.

Allen stated:

My perceptions of individuals from low-socioeconomic backgrounds has changed because of this service-learning project. I guess I thought that because they (the FFF participants) typically have a lower household income and are generally more overweight, that they are less interested in exercise and living a healthy lifestyle. But in fact, every parent I spoke to was very passionate about their child relearning the proper diet and exercise habits now in order to combat the diseases that typically accompany sedentary lifestyle.

Providing graduate students avenues to alter pre-conceived stereotypes of diverse communities is essential in the process of developing culturally responsive teachers (Ladson-

Billings, 1994). Understanding the needs of diverse students is crucial given the broad range of race, ethnicities, culture and socioeconomic strata in today's schools.

V. Discussion.

This study was designed to explore the impact of participation in a service-learning program on graduate physical education teacher education students. We purposely attempted to answer the following research questions: a) what impact did teaching physical activity to low-income minority children have on graduate students' self-efficacy for teaching b) what effect did teaching physical activity to low-income minority children have on graduate students cultural competence, and c) did participation in the service-learning program enhance graduate students' application of pedagogical content knowledge?

Based on Social-cognitive theory (Bandura, 1986; 1999) we examined the interactions of the student's personal factors (self-efficacy, knowledge), the environment (teaching physical activity to low-income minority children) on graduate pre-service educators behaviors (teaching techniques and strategies). Findings suggest that graduate students reaped numerous benefits from participation in the FFF service-learning program. Specifically, the majority of graduate students reported that participation in FFF strengthened their efficacy for teaching physical activity (i.e., 82%), provided multiple opportunities to implement varied teaching techniques (i.e., 68%), and altered their pre-conceived stereotypes of children living in low-income families (i.e., 75%) (see Figure 2).

Bandura (1986; 1999) proposed that an individual's personal factors (i.e., knowledge, feelings, attitudes) interact with one's environment and consequently initiate a process that mandates the individual actively engage in the learning process. The graduate students reported a wide array of affective feelings at the start of the program. Emotions ranged from nervousness and anxiety to excitement and motivation. It appears that the FFF setting was one that enabled the graduate students to work with and through their emotions to create developmentally appropriate physical activity lessons for the children.

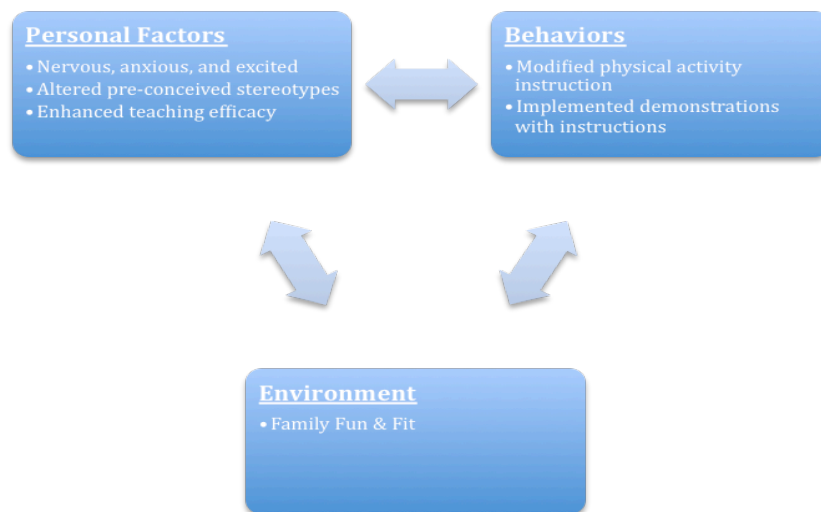


Figure 2. A model diagramming triadic reciprocity during graduate student participation in service-learning. (Adapted from Schunk [2012])

Specific to the first research question exploring the impact of participation in the service-learning program on teaching efficacy, findings support the graduate students' strengthened efficacy for teaching. Scholars in teacher education have defined teaching efficacy as "the extent to which the teacher believes he or she has the capacity to affect student performance" (Berman, McLaughlin, Bass, Pauly, & Zellman, 1977, p. 137). Research has demonstrated that a teacher's level of self-efficacy is correlated with student achievement (Ashton & Webb, 1986; Bolshakova, Johnson, & Czerniak, 2011; Lumpe, Czerniak, Haney, Beltyukova, 2012) commitment to teaching (Coladarci, 1992), teacher motivation (Midgley, Feldlaufer, & Eccles, 1989; Wood & Olivier, 2008), and novelty in instruction (Stein & Wang, 1988). Participation in the FFF service-learning program positively impacted the graduate students' levels of teaching efficacy. Particularly important to note is the reciprocal interaction of enhanced teaching efficacy and adaptation of instructional strategies. Providing the graduate students multiple opportunities to teach the children positively contributed to their heightened efficacy (i.e., personal factors), which transferred to promoting effective instructional strategies (i.e., behaviors). This finding supports Bandura's model of triadic reciprocity (Bandura, 1986) and previous research that demonstrated the positive impact participation in authentic field-based experiences had on undergraduate students' efficacy for teaching physical education (Gurvitch & Metzler, 2008).

Social-cognitive theory (Bandura, 1986; 1999) emphasizes the potential role one's environment may have in influencing learning and behaviors. The FFF program served as the graduate pre-service educators' imposed environment. Teaching physical activity to Hispanic-American and African-American children from low-income families was mandated through enrollment in the graduate curriculum and instruction course. However, the manner in which the pre-service educator chose to engage in the environment constituted their selected environment that resulted in the pre-service educators constructed environment.

Our second research question focused on examining the effect of participation in the FFF program on graduate students' cultural competence for teaching. Allowing graduate students to construct their environment provided a forum to alter their pre-conceived stereotypes encompassing minority children from low-income families. Shattering stereotypes is a critical factor necessary to nurture culturally competent and responsive teachers. Ladson-Billings (1994) underscores the importance of developing teachers that are culturally responsive and proposes that teachers that are culturally responsive hold high self-esteem as well as a profound respect for their students and their families. In addition, culturally responsive teachers believe that all students can achieve success.

Participation in the FFF service-learning program provided a forum for the graduate students to get to know and interact with diverse children. These interactions enabled the graduate students the opportunity to change their perceptions of low-income children and families. Because the majority of beginning teachers enter educational settings where approximately 75% of students are raised in low-income families, it is imperative that teacher education programs provide pre-service teachers multiple and varied opportunities to interact with children and adolescents from varied ethnic, racial, and economic backgrounds (National Collaborative on Diversity in the Teaching Force, 2004).

Our third research question explored the influence of participation in the FFF program on graduate students' application of pedagogical content knowledge. Sixty-eight percent of the graduate students reported specific examples of their increased ability to apply instructional techniques discussed in class to the authentic teaching environment. The necessity of providing "show and tell" demonstrations to benefit children's cognitive understanding of the verbal

instructions was perceived by the graduate students as a valuable result of participation in the program. Similarly, the graduate students also came to realize the need for brief instructions when working with elementary-aged children. The FFF constructed environment also presented multiple occasions for the graduate students to modify and alter their activities to better meet the needs of the children. Application of pedagogical content knowledge via PETE service learning programs has been documented with undergraduate students (Meaney et al., 2009); findings from this study document the positive impact participation in service learning also has on graduate students application of pedagogical content knowledge.

Taken collectively, findings from this study suggest that participation in a service-learning program positively impacted graduate students' efficacy for teaching, acquisition of instructional strategies, and cultural competence. Of particular importance is the unique contribution this investigation makes to the service-learning literature. While previous studies in PETE have explored the impact of service-learning with undergraduate student populations' experiences (Cutforth, 2000; Domangue & Carson, 2008; Kahan, 1998; LaMaster, 2001; Meaney et al., 2008, Meaney, Griffin, & Bohler, 2009; Watson, Crandall, Hueglin, & Eisneman, 2002), this was the first study to exam the effect of participation in service-learning with graduate PETE students. Even though graduate and undergraduate students differ in academic, social, and life experiences variables, participation in the FFF service-learning program provided valuable learning experiences for graduate PETE students. Consequently, both graduate and undergraduate teacher preparation programs should include service-learning experiences within the prescribed curriculum.

While findings from this study contribute to the knowledge base encompassing service-learning, it should be noted however, that this investigation reports and interprets the findings solely from the perspective of the graduate students. This is certainly a limitation of the study. Future research examining service-learning programs in teacher education should consider including assessing the benefits and/or disadvantages of participation in service-learning program by the children, families, and school personnel.

References

- American Council on Education (2005). *College Students Today: A National Portrait*. Washington, DC: American College of Education.
- Anderson, J.B., Swick, K.J., and Yff, J. (Eds.). (2001). *Service-Learning in Teacher Education: Enhancing the Growth of New Teachers, Their Students, and Communities*. Washington, DC: American Association of Colleges for Teacher Education.
- Ashton, P.T., and Webb, R.,B. (1986). *Making a Difference: Teachers' Sense of Efficacy and Student Achievement*. New York: Longman.
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology*, 2, 21-41.
- Berman, P., McLaughlin, M., Bass, G., Pauly, E., and Zellman, G. (1977). *Federal programs supporting educational changes: Vol. VII. Factors affecting implementation and continuation* (Rep. No. R-1589/7-HEW). Santa Monica, CA: RAND. (ERIC Document Reproduction Service No. 140 432.
- Bolshakova, V., Johnson, C., and Czerniak, C. (2011). "It depends on what teacher you got": Urban science self-efficacy from teacher and student voices. *Cultural Studies of Science Education*, 6(4), 961-997.
- Brescia, W., Mullins, C., and Miller, M. (2009). Project-based service-learning in an instructional technology graduate program. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-12. Retrieved from <http://www.georgiasouthern.edu/ijsotl>.
- Coladarci, T. (1992). Teacher's sense of efficacy and commitment to teaching. *Journal of Experiential Education*, 60(4), 323-337.
- Corbin, C.B., and Lindsey, R. (2007). *Fitness for Life*. Champaign, IL: Human Kinetics.
- Cress, C.M. (2005). What is service-learning? In C. M. Cress, P. J. Collier, V.L. Reitenauer (eds.) *Learning Through Serving: A Student Guidebook for Service-Learning Across Disciplines*. Sterling, VA: Stylus.
- Cutforth, N.J. (2000). Connecting school physical education to the community through service-learning. *Journal of Physical Education, Recreation and Dance*, 71, 39-45.
- Denzin, N.K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. Chicago: Aldine.

Dewey, J. (1938). *Experience and Education*. New York: MacMillan.

Domangue, E., and Carson, R.L. (2008). Preparing culturally competent teachers: Service-learning and physical education teacher education. *Journal of Teaching in Physical Education*, 27, 347-367.

Eyler, J.S., Giles, D.E., Stenson, C.M., and Gray, C.J. (2003). What we know about the effects of service-learning on college students, faculty, institutions and the community, 1993-2000, Third Edition. Introduction to Service-Learning Toolkit: Readings and Resources for Faculty, (2nd ed.), pp. 15-19. Providence, RI: Campus Compact.

Graham, G., Holt/Hale, S., and Parker, M.A. (2009). *Children Moving: A Reflective Approach to Teaching Physical Education (8th ed.)* New York: McGraw Hill.

Green, J., Dalton, R., and Wilson, B. (1994). *Implementation and evaluation of TEACH: A service-learning program for teacher education*. Paper presented at the Annual Meeting of the Association of Teacher Education, Atlanta GA.

Griffin, L.L. and Butler, J.I. (2005). *Teaching Games for Understanding: Theory, Research, and Practice (Eds.)* Chicago: Human Kinetics.

Gurvitch, R., and Metzler, M.W. (2008). The effects of laboratory – based and field-based practicum experience in pre-service-teachers self-efficacy. *Teaching and Teacher Education*, 25(3) 437-443.

Illovsky, M.E. (2010). Psychological comparison of undergraduate and graduate College of Education students. *International Journal of Teaching and Learning in Higher Education*, 23(30), 238-245). Retrieved from <http://www.isetl.org/ijtlhe>.

Kahan, D. (1998). When everyone gets what they want: A description of a physical education teacher education service-learning project. *Action in Teacher Education*, 19, 43-60.

Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall.

Ladson-Billings, G. (1994). *The Dreamkeepers: Successful Teachers of African-American Children*. San Francisco: Jossey-Bass.

LaMaster, K.J. (2001). Enhancing pre-service teachers field experiences through the addition of a service-learning component. *Journal of Experiential Education*, 24, 27-33.

Lincoln, Y.S., and Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Liu, Y., and Lambright, K.T. (2010). Looking beyond the undergraduate classroom: Factors influencing service-learning's effectiveness at improving graduate students' professional skills. *College Teaching*, 58, 118-126.

Lumpe, A., Czerniak, C., Haney, J., and Beltyukova, S. (2012). Beliefs about teaching science: The relationship between elementary teachers' participation in professional development and student achievement. *International Journal of Science Education*, 34(2), 153-166.

Meaney, K.S., Griffin, L.K., and Bohler, H.R. (2009). Impact of a service-learning program on pre-service educator's knowledge base for teaching. *International Journal for the Scholarship of Teaching and Learning*, 3(2), 1-17. <http://www.georgiasouthern.edu/ijstol>

Meaney, K.S., Bohler, H.R., Kopf, K., Hernandez, L., and Scott, L. (2008). Service-learning and pre-service educators' cultural competence for teaching: An exploratory study. *Journal of Experiential Education*, 31(2), 189-208.

Midgley, C., Feldlaufer, H., and Eccles, J. (1989). Changes in teacher efficacy and student self and task related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, 81, 247-258.

National Collaborative on Diversity in the Teaching Force (2004, October). *Assessment of diversity in America's teaching force: A call to action*. 1-14.

Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed). Thousand Oaks, CA: Sage Publications.

Pitney, W.A., and Parker, J.P. (2009). *Qualitative Research in Physical Activity and the Health Professions*. Champaign, IL: Human Kinetics.

Quinn, S.M.F. (2006). Facilitating service-learning for the first time with advanced graduate students: A mentoring perspective. *Mentoring & Tutoring*, 15(1), 97-110.

Rink, J.E. (2010). *Teaching Physical Education for Learning*. (6th ed). New York: McGraw Hill.

Root, S., and Swick, K.J. (2001). A framework for conceptualizing and doing research on service-learning in pre-service teacher education. In J. B. Anderson, K.J. Swick, & J. Yff (Eds.). *Service-Learning in Teacher Education: Enhancing the Growth of New Teachers, Their Students, and Communities* (pp. 141-152). Washington, DC: American Association of Colleges for Teacher Education.

Schunk, D.H. (2012). *Learning Theories: An Educational Perspective*. (6th ed). Boston: Pearson Education.

Stein, M.,K., and Wang, M.C. (1988). Teacher development and school improvement: The process of teacher change. *Teaching and Teacher Education*, 4, 171-197.

Tellez, K., Hlebowitsh, P.S., Cohen, M., and Norwood, P. (1995). Social service field experiences and teacher education. In J.M. Larkin & C.E. Sleeter (Eds.), *Developing Multicultural Teacher Education Curricula*. Albany, NY: State University of New York Press.

Meaney, K. S., Housman, J. M., Cavazos, A., and Wilcox, M. L.

United States Census Bureau (2007). American Community Survey. Retrieved from <http://factfinder.census.gov/servlet/>.

Wade, R.C. (1995). Developing active citizens: Community service-learning in social studies teacher education. *The Social Studies*, 86, 122-128.

Wade, R.C., and Yarbrough, D.B. (1997). Community service-learning in student teaching: Toward the development of active citizenry. *Michigan Journal of Community Service Learning*, 4, 42-55.

Watson, D.L., Crandall, J., Hueglin, S., and Eisenman, P. (2002). Incorporating service-learning into physical education teacher education programs. *Journal of Physical Education, Recreation, and Dance*, 73, 50-54.

Wood, L., and Olivier, T. (2008). Addressing the needs of teachers in disadvantaged environment through strategies to enhance self-efficacy. *Teacher Development*, 12(2), 151-164.

A multicultural personal growth group as a pedagogical strategy with graduate counseling students

Jennifer M. Johnson^{1,2} and Glenn W. Lambie¹

Abstract: The present study investigated a six-week multicultural personal growth group as a pedagogical strategy to support first-year graduate counseling students' (N = 20) levels of ethnic identity development (Phinney & Ong, 2007) and social-cognitive maturity (Hy & Loevinger, 1996). Students' levels of ethnic identity and social-cognitive development did not change; however, 90% (n = 18) of the students reported valuing the group experience and appreciated the group leaders facilitation of the pedagogical strategy. Implications for the scholarship of teaching and learning are discussed.

Keywords: counselor education and development, multicultural counseling and development, scholarship of teaching and learning, social-cognitive development

Multicultural personal growth groups may be an effective pedagogical strategy to promote graduate counseling students' ethnic identity development and social-cognitive maturity (e.g., Arthur & Achenbach, 2002; Rowell & Benschhoff, 2008; Villalba & Redmond, 2008). Specifically, personal growth groups are experiential pedagogical strategies that foster emotional learning experiences and insight into an individual's strengths and areas necessitating strengthening (Yalom & Leszcz, 2005). In addition, personal growth groups provide an opportunity for graduate students to engage in course material, participate in self-reflection, and promote feelings of hope and efficacy (e.g., Faith, Wong, & Carpenter, 1995; Lieberman, Yalom, & Miles, 1973). Within graduate counseling programs, personal growth groups support students' development of multicultural competencies (e.g., Arthur & Achenbach, 2002; Villalba & Redmond, 2008) and ethnic identity development (Rowell & Benschhoff, 2008). Furthermore, personal growth groups may promote graduate students' social-cognitive maturation, as group experiences encourage reflection, self-exploration, and disequilibrium (Johnson & Johnson, 2009; Lambie & Sias, 2009).

Experiential learning activities (e.g., multicultural personal growth groups) are beneficial to graduate students: (a) exposing students to multiculturalism, (b) increasing multicultural awareness, (c) aiding in graduate student cultural empathy, and (d) challenging students' beliefs about diversity (e.g., Pope-Davis, Breaux, & Liu, 1997; Ridley & Lingle, 1996; Villalba & Redmond, 2008). Specifically, Leonard (1996) suggested that multicultural-focused groups or consciousness-raising groups are an integral part of building multicultural self-awareness and social awareness of oppressive systems for counselors-in-training. Nevertheless, to determine the

¹ Jennifer M. Johnson, Department of Educational & Human Sciences, College of Education, University of Central Florida; Glenn W. Lambie, Associate Professor, Department of Educational & Human Sciences, College of Education, University of Central Florida

² Jennifer M. Johnson is now at the Department of Educational Leadership, Counseling, & Foundations at the University of New Orleans. Correspondence concerning this article should be addressed to Jennifer M. Johnson, Department of Educational Leadership, Counseling, & Foundations, Counselor Education Program, University of New Orleans, Bicentennial Education Bldg., 2000 Lakeshore Dr., New Orleans, LA 70148. Email: jennifermarie.jmj@gmail.com

effectiveness of multicultural pedagogical strategies, additional research is warranted (Hill, 2003; Pope-Davis & Ottavi, 1994; Ridley, Espelage, & Rubenstein, 1997; Vasquez, 1997). In particular, an investigation of the impact of a multicultural personal growth group (pedagogical intervention) on graduate counseling students' ethnic identity development (Phinney & Ong, 2007) and social-cognitive maturity (Hy & Loevinger, 1996) would contribute to the scholarship of teaching and learning literature (SoTL; Hutchings, Huber, & Ciccone, 2011).

I. Ethnic Identity Development and Social-Cognitive Maturity.

Scholars note the importance of ethnic identity development and social-cognitive maturity in the preparation of ethical and effective counseling professionals. We described these two constructs to set an accurate context for the SoTL investigation that follows.

A. Ethnic Identity Development.

Ethnic identity development (Phinney & Ong, 2007) emphasizes an ethnic sense of belonging, positive attitudes towards ones' ethnicity, and commitment to ethnic traditions and practices in an individuals' cultural background (Phinney, 1990; Phinney & Alipuria, 1996). Unlike racial identity that emphasizes physical characteristics (e.g., skin color); ethnic identity encompasses the acquisition or maintenance of cultural characteristics that define a particular ethnic group (Helms, 1996). Phinney's (1990) *Ethnic Identity Model* offers a conceptual framework of ethnic identity development as an awareness and commitment to ones' cultural background (e.g., engaging in cultural practices). In addition, the *Ethnic Identity Model* (Phinney, 1990) is organized into three distinct hierarchical levels: (a) unexamined ethnic identity, (b) ethnic identity search/moratorium, and (c) ethnic identity achievement according to individuals' identity formation surrounding their ethnicity. Phinney's model of ethnic identity development provides a comprehensive exploration of ethnicity while being inclusive in application for all ethnicities (Chavez & Guido-DiBrito, 1999).

Ethnic identity development has been empirically tested with diverse populations (e.g., Phinney, 1989; Phinney, 1992; Phinney & Alipuria, 1990; Phinney & Ambarsoom, 1987; Phinney & Tarver, 1988; Roberts et al., 1999; Sobansky et al., 2010; Syed & Azmitia, 2008; Utsey, Chae, Brown, & Kelly, 2002). In addition, higher levels of ethnic identity development are predictive of multicultural competence in counseling and counseling psychology graduate students (Chao, 2006). Specific to graduate counseling students, ethnic identity development correlates with cultural competence (e.g., Evans & Foster, 2000; Middleton et al., 2005; Neville et al., 1996; Vinson & Niemeyer, 2000), which is a desirable counselor quality. In addition, a small group experience in a multicultural counseling course promotes greater levels of students' self-identified scores of ethnic identity development as compared to non-participant students (Rowell & Benschhoff, 2005). Nevertheless, additional research is needed to examine the impact of a pedagogical intervention (multicultural personal growth group) on graduate counseling students' ethnic identity development.

B. Social-Cognitive Development.

Social-cognitive development (Loevinger, 1976; 1998), or *ego development*, describes the paradigm in which individuals perceive themselves and others through interpersonal and

intrapersonal experiences. In Loevinger's theory of social-cognitive development, ego maturation is central to an individual's personality formation, including domains of cognitive development, character development, interpersonal style, and conscious preoccupations (Loevinger, 1976). Within social-cognitive development, individuals mature through nine invariant hierarchical levels: (a) E2 - Impulsive, (b) E3 - Self-protective, (c) E4 - Conformist, (d) E5 - Self-aware, (e) E6 - Conscientious, (f) E7 - Individualistic, (g) E8 - Autonomous, and (h) E9- Integrated (Hy & Loevinger, 1996). Individuals' levels of ego functioning are based on their interaction with their environment, where assimilating new information into existing schema results in ego stability and adapting their existing schema to accommodate new information promotes developmental growth (Lambie & Sias, 2009). Furthermore, the levels of ego maturity are not age-specific, but are descriptive of hierarchical growth in a sequential manner based on individuals' interactions with their environment (Manners & Durkin, 2000).

Within the graduate counseling research, social-cognitive development (Loevinger, 1976; 1998) relates to desirable counselor qualities, such as resistance to burnout (Lambie, 2007), counselor effectiveness and perceptions of clients (Borders, Fong, & Neimeyer, 1986; Zinn, 1995), levels of multicultural competence (Cannon & Frank, 2009) and racial identity development (Watt, Robinson, & Lupton-Smith, 2002), acquisition of ethical and legal knowledge (Lambie, Hagedorn, & Ieva, 2010), the ability to develop a counseling theoretical orientation (Warren, 2008), and wellness (Lambie, Smith, & Ieva, 2009). In addition, counselors' social-cognitive development is important in their ability to be effective helping professionals working with clients and their diverse needs (e.g., case conceptualization; flexibility; Cannon & Frank, 2009; Lambie, 2007). Moreover, higher levels of ego maturity correlate with increased openness to individual differences and less stereotypical thinking (Watt et al., 2002). Furthermore, the fifth level of social-cognitive development (Self-aware, E5) was identified as the minimal level of function for counselors to be effective with their clients (Zinn, 1995). Nevertheless, research is warranted to examine whether a group experience may have the potential to promote student development (e.g., social-cognitive maturity) within graduate counselor preparation programs in graduate students.

II. Purpose of the Study.

Given the significance of ethnic identity development (Phinney & Ong, 2007) and social-cognitive maturity (Hy & Loevinger, 1996) in counselors' practice (e.g., counseling skills, counselor wellness, multicultural competencies; Evans & Foster, 2000; Lambie et al., 2009; Middleton et al., 2005; Vinson & Niemeyer, 2000); we examined the impact of a six-week multicultural personal growth group on graduate counseling students' levels of these two constructs in an effort to identify potential implications for supporting future counselor effectiveness and SoTL. The two research questions investigated were (a) What is the impact of a six-week multicultural personal growth group on graduate counseling students' levels of ethnic identity development and social-cognitive maturation? and (b) What are graduate students' perceptions of a multicultural personal growth group experience as a pedagogical strategy?

III. Methodology.

C. Procedures and Participants.

The sample included 20 masters-level counseling students at a large metropolitan research university in the southeastern United States. All the students were in their first-semester in a counselor education program and were enrolled in an *Introduction to the Counseling Profession* course. We attained approval to conduct the study from the university's Institutional Review Board. The data collection began during the first *Introduction to the Counseling Profession* class meeting and concluded following the completion of the six-week multicultural personal growth program. To protect the rights and confidentiality of participants, participation was voluntary, students' names and identifying information were not collected, and none of the data collected were reviewed or scored until after the completion of that academic semester. Of the 23 students participating in the group pedagogical intervention, 20 students completed all the data collection instruments (resulting in an 87.0% usable response rate).

D. Participant Characteristics.

The 20 graduate counseling students that participated were divided into two multicultural personal growth groups, where eleven students were in *Group One* and nine students were in *Group Two*. Overall, men were less represented than were women: there were five men (25%) and 15 females (75%). The ages of participants ranged from 22 - 51 years ($M = 26.80$, $SD = 7.28$). In addition, the racial/ethnic background identified reported by the participants was: Caucasian, 55% ($n = 11$); African American, 20% ($n = 4$); Hispanic/Latino, 5% ($n = 1$); Other, 15% ($n = 3$); and Asian, 5% ($n = 1$). Furthermore, participants were asked to indicate their level of cultural competence on a four-point Likert Scale (e.g., one being satisfied) and 10% ($n = 2$) identified as "very satisfied," 75% ($n = 15$) "somewhat satisfied," and 15% ($n = 3$) were "somewhat dissatisfied." Lastly, the participants represented three counseling specialties: (a) 55% ($n = 11$) mental health, (b) 30% ($n = 6$) marriage and family, and (c) 15% ($n = 3$) school counseling.

E. Pedagogical Intervention (Multicultural Personal Growth Group).

The multicultural personal growth group curriculum consisted of six closed group sessions that were one hour in duration. Each group session contained semi-structured activities that were co-facilitated by masters-level clinicians enrolled in counselor education doctoral programs who had previous group work experience. The activities focused on counselor self-awareness (a multicultural competency domain; Arredondo et al., 1996), which stipulates that it is imperative for counselors to be aware of their own cultural values and beliefs. Specifically, the group activities within the counselor self-awareness domain assert that: (a) culturally skilled counselors believe cultural self-awareness is essential; (b) counselors are aware of their cultural background have influenced values and biases about psychological processes; (c) counselors are able to recognize their limits of multicultural competency and expertise; and (d) counselors recognize their discomfort regarding differences between themselves and others related to race, ethnicity, and culture (Arredondo et al., 1996). Therefore, the groups pedagogical activities were designed to foster interpersonal and intrapersonal reflection, facilitate exploration in the students' self-

awareness of ethnic identity (e.g., commitment and exploration of their ethnic identity), and promote the students level of social-cognitive development.

Accordingly, the six sessions of the multicultural self-awareness groups facilitated graduate students' introspection of their biases and assumptions.

Group Session One consisted of introducing group members to the multicultural personal growth group, discussing group procedures (e.g., confidentiality), and an icebreaker activity to help group members become acquainted with one another. Pedagogical objectives included: (a) establishment of group rules, (b) introduction of group leaders and members (e.g., name, counseling specialization, reasons for wanting to be a counselor) through a think-pair-share activity, and (c) discussion of how group members' cultural background may influence their work as a future counselor. The think-pair-share activity encouraged group members to be reflective of their desire to be counselors and their ethnic background.

Group Session Two consisted of an activity where group members discussed the origins of their beliefs and attitudes regarding cultural diversity. Therefore, the pedagogical group objectives were to: (a) have group member define various types of diversity and (b) discuss how they learned about diversity. The activity was completed in pairs to facilitate more self-disclosure of group participants. Once the pairs completed their sharing, the group leader facilitated the group session with guiding questions (e.g., at what age do you remember learning about people who were different from you?). At the conclusion of group session two, the group facilitator asked each group member to bring in a tangible object that represented his or her cultural background for discussion in the subsequent group session.

Group Session Three involved group members sharing their tangible representation of their culture or ethnic background with their group members in a "round robin" fashion. The group leader facilitated the discussion with the following prompts: (a) discuss the significance of this cultural artifact to you; (b) what did you learn about yourself, your family, culture as a result of this activity?; and (c) how are your beliefs and attitudes shaped by your culture in terms of your worldview and how you treat others that are different from you? The purpose of this pedagogical strategy was to encourage group member participation, interpersonal learning, and appreciation of various cultural backgrounds.

Group Sessions Four and Five included a two-part activity where the objective was for the participants to reflect on spoken and unspoken messages from their upbringing and current beliefs and attitudes. An adaptation of an activity, called DOTS: Understanding your "hidden" biases from the principal researcher's institution was utilized where group leaders wrote a belief statement on the top of a poster board with "True" and "False" underneath with True on the far left side and False on the far right side. The statements included:

- (a) If you are poor, it is because you are lazy.
- (b) People with disabilities or mental disorders should be "hidden" from society.
- (c) Homosexuality is wrong.
- (d) Men are more competent than women are.
- (e) Belief in God according to Christianity is the only way to heaven.
- (f) If you work hard enough, you will be successful in America, regardless of race/ethnicity.

Group members were given color-coded dots that represented where group members received these messages. Red dots represented family of origin, blue dots were friends and peers, green dots represented K-12 school years, and orange dots were the media. Thus, group members were instructed to place their sticker on the True-False continuum as it pertained to the message they

received in their upbringing. Once students placed the stickers on each poster board, the group leader asked the group members if they noticed any patterns on the poster board. Additionally, the processing statements/questions helped guide the group discussion, for example, explain why you placed the dot where you did on the continuum of True-False.

Group Session Six, the final group, consisted of a closing activity where group members verbalized how they define themselves as cultural beings through a structured activity. The pedagogical objectives of the activity were to: (a) recognize moments when they were proud to be identified by one of the descriptors and (b) detect moments when it was painful to be identified with one of the descriptors. The group experience ended with the following processing questions for group members to discuss: (a) what have you learned from this experience and (b) how can you apply what you have learned to personal and professional life as a future counselor. For further information related to the pedagogical strategies integrated into the multicultural personal growth group, please contact us for a copy of the group curriculum.

F. Instrumentation.

The two constructs and the instruments used to measure these constructs were (a) ethnic identity development (*Multi-group Ethnic Identity Measure* [MEIM-R]; Phinney & Ong, 2007) and (b) social-cognitive development (*Washington University Sentence Completion Test* [WUSCT]; Loevinger & Hy, 1996). Study participants completed the instruments at two points during the semester: (a) first class of the semester and (b) last class of the semester. The primary variables used to examine the constructs were overall mean ethnicity identity development scores (pre- and posttest) and social-cognitive development level and total protocol scores (pre- and posttest) scores. In addition, the participants completed the *Students' Perception of Multicultural Personal Growth Group Experiences Questionnaire* (SPMPGGEQ), which is a six-item questionnaire regarding the students' overall perceptions of the pedagogical strategy.

Multi-group Ethnic Identity-Revised Measure. The MEIM-R (Phinney & Ong, 2007) measures the construct of ethnic identity development and is appropriate for use with all ethnicities (Ponterotto et al., 2003). Additionally, research findings support the administration of the MEIM-R with various age groups (e.g., adolescents and young adults; Phinney, 1989; 1990; 1992, Phinney & Alipuria, 1990; Phinney & Ong, 2007; Phinney & Tarver, 1988). Hence, the MEIM-R is applicable for college-age adults, the population for our investigation.

The MEIM-R is a ten-item assessment, revised from the original 14-item MEIM, and begins with an open-ended prompt for participants to specify their self-identified ethnic group. Following the open-ended prompt, participants rate themselves on a Likert scale (e.g., *5-strongly agree, 4-agree, 3-neutral, 2-disagree, and 1-strongly disagree*) relating to six statements that assess ethnic identity. For instance, "I have a strong sense of belonging to my own ethnic group." Moreover, these six questions consist of two subscales: (a) exploration (items 1, 4, and 5) and commitment (items 2, 3, and 6). The exploration subscale assesses whether individuals are engaged in activities related to their ethnicity, while, the commitment subscale describes the level of attachment to ones' ethnicity. The eighth question asks participants to categorize their ethnicity based on the ethnic groups provided. Lastly, the final two questions (questions nine and ten); request identification of the ethnicities of the participants' mother and father using the provided ethnic categories.

Scoring of the MEIM-R consists of either averaging individual subscale scores (e.g., exploration and commitment), or averaging combined subscale scores. Nonetheless, the

instrument authors' suggested the mean calculation of the total score for "studies concerned only with the overall strength of ethnic identity or the degree to which ethnic identity is achieved" (Phinney & Ong, 2006, p. 278). Therefore, an averaged total score of 3.33, for example, indicates moderate ethnic identity development on a scale of one to five, where five represents high levels of ethnic identity. Thus, high MEIM-R scores indicate higher levels of ethnic identity development.

The MEIM is a psychometrically sound instrument with adolescent and adult populations (Phinney & Baldelomar, 2006; Phinney & Ong, 2006; Roberts et al., 1999). Specifically, an exploratory factor analysis of the MEIM-R calculated two subscales (e.g., exploration and commitment) which measure ethnic identity development (Phinney & Ong, 2006). In a sample of 192 university students, moderate reliability was found in the subscales of exploration and commitment, respectively, with Cronbach's alphas of .83 and .89 (Phinney & Ong, 2006). In addition, Phinney and Ong (2007) conducted a confirmatory analysis with a sample of 241 university students and concluded the two-factor model (e.g., exploration and commitment to an individuals' ethnic identity) to be a good fit for the construct of ethnic identity (e.g., $\chi^2/df = 1.91, p < .001$; AGFI = .96, CFI = .98, RMSEA = .04).

Washington University Sentence Completion Test. The WUSCT (Hy & Loevinger, 1996) measures social-cognitive development of individuals and is suitable for individuals in pre-adolescence through adulthood and in numerous settings (e.g., K-12 schools, universities, and hospitals; Manners & Durkin, 2001). Our participants completed the short form of the WUSCT (Form 81-1; 18 sentence stems), which is a semi-projective inventory with open-ended sentence stems. For example, one sentence stem is, "The thing I like about myself is."

We scored the 18-item WUSCT using the technical foundations manual (Hy & Loevinger, 1998), which includes practice exercises reviewed by experts in the field in a three-step process. The first step consisted of scoring each individual sentence stem in the assessment and matching it to an ego maturity level. Secondly, we summed the ego levels for each individual assessment; this number represented the total protocol rating (TPR). Lastly, we matched the TPR to ego maturity levels in the technical foundations manual. The TPR included a range of scores: (a) Impulsive (E2) TPR = 36 - 67, (b) Self-Protective (E3) TPR = 68 - 75, (c) Conformist (E3) TPR = 76 - 81, (d) Self-Aware (E5) TPR = 82 - 90, (e) Conscientious (E6) TPR = 91 - 100, (f) Individualistic (E7) TPR = 101 - 108, (g) Autonomous (E8) TPR = 109 - 118, and (h) Integrated (E9) TPR = 119 and above. Therefore, higher WUSCT TPR scores and ego levels correlate with increased social-cognitive development.

The psychometric properties of the WUSCT (Hy & Loevinger, 1996) are strong (e.g., Manners & Durkin, 2001; Novy & Francis, 1992). Specifically, Novy and Francis (1992) sampled 265 adults and found a high and significant reliability for half of the WUSCT with the first half coefficient $\alpha = .84$ and the second half coefficient $\alpha = .81$. In addition, the construct validity (e.g., Westenberg & Block, 1993), predictive validity (Hart & Hilton, 1988), and discriminant validity (e.g., Hauser, 1976; Loevinger, 1979) of the WUSCT have been supported.

Students' Perception of Multicultural Personal Growth Group Experiences Questionnaire. We administered the SPMPGGEQ, which is an anonymous six-item survey at the end of the last multicultural personal growth group to gain insight into the graduate students' perceptions of the experience with this pedagogical strategy. Group leaders explained that completing the survey was voluntary and left the room for students to complete the questionnaire. Twenty of the graduate students completed the SPMPGGEQ that included the following question prompts:

- (a) Describe the experience that you had with the facilitators of the multicultural personal growth group.
- (b) Was there anything that you would change about how the group was facilitated?
- (c) What did you like about the personal growth group experience?
- (d) What did you dislike about the personal growth group experience?
- (e) How did you like the activities that were part of the personal growth group?
- (f) Do you feel that participating in the PGG was a valuable learning experience to learn about ethnic / cultural identity?

G. Data Analysis.

A time-series research design was selected for our investigation as the primary variables (pedagogical intervention, multicultural personal growth group) were manipulated without random assignment. Tabachnick and Fidell (2007) suggested that time series designs are utilized to assess the effect of an intervention before and after the intervention. Nevertheless, threats to validity may influence findings from a time-series research design, such as history (factors other than the intervention influencing change) and testing (due to a practice effect; Fraenkel & Wallen, 2011).

After the data collection process, the data were scored and entered into a database and analyzed by Statistical Package for the Social Sciences (Version 19.0) using paired-sample *t*-tests. Prior to beginning the data analysis process, data was screened and preliminary analyses were conducted (e.g., assessing normality and checking for outliers). A sample size of 20 was acceptable for identifying a large effect size (power = .80) at the .10 level (Cohen, 1992).

IV. Results.

H. Ethnic Identity Development.

The MEIM-R (Phinney & Ong, 2007) was used to obtain participants ethnic identity development scores ($N = 20$) on a five-point Likert Scale (e.g., 1 = *strongly disagree*). The pre-test and post-test MEIM-R scores, with descriptive data are presented in Table 1. A paired-sample *t*-test indicated that the graduate students' MEIM-R mean scores did not change from the initial assessment to the posttest, $t(19) = .86, p > .05$. Therefore, the graduate students did not experience a significant change in their ethnic identity development per the pedagogical intervention.

I. Social-Cognitive Development.

The WUSCT-Form 81 (Hy & Loewinger, 1996) was used to obtain participants' ($N = 20$) social-cognitive development scores. The pre-test and post-test WUSCT scores with descriptive data are presented in Table 1. The pretest WUSCT level scores were Self-Protective (E3; $n = 1$; 5%), Conformist (E4; $n = 3$; 15%), Self-Aware (E5; $n = 11$; 55%), and Conscientious (E6; $n = 5$; 25%). The mean pretest social-cognitive maturity level was Self-Aware (E5; $M = 5.00, SD = .79$, range = E3 - E6; TPR score, $M = 87.10, SD = 6.04$, range = 75 - 98). The students' posttest level scores were Self-Protective (E3; $n = 1$; 5%), Conformist (E4; $n = 3$; 15%), Self-Aware (E5; $n = 11$; 55%), Conscientious (E6; $n = 4$; 20%), and Individualistic (E7; $n = 1$; 5%). The average

posttest social-cognitive maturity level was Self-Aware (E5; $M = 5.05$, $SD = .89$, range = E3 - E7; TPR score, $M = 86.55$, $SD = 7.25$, range = 71 - 101).

A paired-sample t -test indicated that the graduate students' WUSCT mean level scores did not change from the initial assessment to the posttest, $t(19) = .24$, $p > .05$. Similarly, there was no change from pretest WUSCT TPR and posttest WUSCT TPR scores, $t(19) = .44$, $p > .05$. Thus, graduate students did not experience a change in social-cognitive maturity because of the pedagogical intervention.

J. Students' Perceptions of the Multicultural Personal Growth Group Experience.

Overall, the graduate students' perception of the multicultural personal growth group questionnaire results indicated that 90% ($n = 18$) of the students valued the group experience and appreciated the group leaders facilitation of the groups. The first SPMPGGEQ question asked participants to describe their experience with group facilitators of the multicultural personal growth groups. We divided the feedback into two themes (a) group facilitators' behaviors and (b) group facilitators' demeanor. Overall, group members reported that the group leaders set a positive foundation for the groups, as evidenced by, facilitating the discussion of group rules and creating a safe place to discuss intimate details of their life. Moreover, the students reported valuing the facilitators' ability to engage students in self-exploration, balance "airtime" between group members, and challenge members when necessary. Additionally, the students described the group facilitators' demeanor, to be positive, supportive, encouraging, helpful, honest, and non-judgmental.

The second SPMPGGEQ question asked for suggested changes to how the group leaders facilitated the groups. Approximately, 75% ($n = 15$) of the students reported that they enjoyed the groups the way they were. There was some conflicting feedback where one participant felt the group was too long (because the groups were after class), while another student wanted more time to process during the group. In addition, a student noted that the group activities could have been more challenging and the inclusion of experiential activities outside of the group would have been beneficial.

The third SPMPGGEQ question asked the students what they enjoyed about the personal growth group experience. The main themes students identified were: (a) ability to share and help one another in the group setting, (b) cohesion within the group, (c) personal growth, (d) enjoyed the weekly topics, and (e) feeling hopeful about the future. The fourth SPMPGGEQ question, asked the students what they disliked about the group experience. The main themes derived from this question were students felt there was not enough time to process group content, and the duration of the group could have been longer (e.g., a semester). In addition, one student noted there was a lack of conflict in the groups and wanted to delve into deeper issues. Finally, one other student reported an interest in receiving more feedback from his or her peers regarding a topic raised in the group.

The fifth SPMPGGEQ question asked specifically about the group activities and if the students liked or disliked them. Overall, the students reported enjoying the group activities because they *promoted self-reflection* and were *insightful*. However, some students felt the activities were too structured, impeding group interaction. Lastly, the final SPMPGGEQ question asked the students to reflect on the group experience as it related to learning about their ethnic and cultural identity. The main themes the students expressed were: (a) learning about personal biases, (b) learning about own cultural background and group members' cultural group, and (c)

realizing there are more similarities than differences between cultures. One student stated, *it's not often you get the chance to have such personal encounters with so many ethnicities*, while another group member stated, *I learned a lot about other races and more importantly how people feel about their identity*. Finally, one graduate student disclosed, *I have come to terms with things I haven't been able to in years and its due only to the fact that this group made me comfortable to speak*.

V. Discussion.

The present study facilitates intentional scholarly inquiry of faculty members into their pedagogical practices (e.g., integrating multicultural personal growth groups in the *Introduction to Counseling Profession* course) to examine graduate counseling student development (e.g., ethnic identity development and social cognitive maturity; Hutchings et al., 2011). The literature and research indicated a need to investigate the potential impact of a multicultural personal growth group on graduate students' ethnic identity development and social-cognitive maturity scores. The results of the statistical analysis identified that graduate students in these multicultural personal growth groups levels of ethnic identity development and social-cognitive maturity did not change.

Table 1. Ethnic Identity Development and Social-Cognitive Maturity Descriptive Statistics.

Variables	Ethnic Identity Development (MEIM-R)		Social – Cognitive Maturity (WUSCT-Form 81) Level (TPR)	
	Pretest	Posttest	Pretest	Posttest
Graduate Counseling Students (N=20)				
<i>M</i>	3.61	3.71	5.00 (87.10)	5.05 (86.55)
<i>SD</i>	.89	.79	.79 (6.04)	.89 (7.25)
Range	1.7 – 5.0	2.0 – 5.0	E3 – E6 (75 – 98)	E3 – E7 (71 – 101)

Note. MEIM-R = Multi-group Ethnic Identity-Revised Measure; WUSCT—Form 81 = Washington University Sentence Completion Test, short form; TPR = Total protocol rating; E = ego development scheme level.

The findings presented here are inconsistent with previous empirical research investigating the influence of multicultural group experiences with ethnic identity development (Rowell & Benschhoff, 2008). Specifically, Rowell and Benschhoff (2008) employed a quasi-experimental research design to investigate the impact of a multicultural personal growth group on participants' (N = 183) ethnic identity scores and found that participants' in the group experience increased their ethnic identity development scores more than individuals not participating in the groups. The inconsistency between our findings and Rowell and Benschhoff's results may be attributed to differences in sample sizes and length of the pedagogical intervention (group experience). Moreover, the activities in Rowell and Benschhoff's personal growth groups could have initiated more introspection into their ethnic identity and the groups occurred throughout the semester.

The findings that the graduate counseling students' social-cognitive maturity scores did not change following the completion of the six week multicultural personal growth group

experience was also consistent with previous research. Specifically, Lambie and colleagues (2010) found that participating in a 13-week course in ethical and legal issues in counseling did not change counselor education students' levels of social-cognitive maturity. Therefore, promoting developmental growth in the area of social-cognitive maturity may be more difficult in a short period of time (six-weeks). Specifically, Sias, Lambie, and Foster (2006) suggested that pedagogical interventions designed to promote developmental growth in adults should be a minimum of six months to one-year in length.

Nevertheless, change in social-cognitive maturity has occurred in models such as, the Deliberate Psychological Education (DPE; Sprinthall & Mosher, 1978). The DPE curricula include a combination of course work and action-oriented experiential activities (e.g., field experiences). Cannon and Frank (2009) utilized the DPE with graduate counseling interns and found a statistically significant difference in social-cognitive maturity scores between treatment group participants in the DPE program ($n = 20$) and control group participants who did not experience the DPE model ($n = 39$). Thus, the incorporation of experiential activities (e.g., service learning opportunities) that promote cognitive dissonance may be necessary to increase graduate students' levels of social-cognitive development.

The graduate counseling students' SPMPGGEQ data supported that the group experience was helpful to the participants' personal development, understanding of personal biases, and an opportunity to have an intimate encounter with peers of different ethnicities. Therefore, the students' levels of ethnic identity and social-cognitive development may not have changed per the pedagogical strategy; however, the participants appeared to grow and benefit from the experiential group intervention.

K. Implications for the Scholarship of Teaching and Learning.

Although our findings were mixed, no changes in the graduate counseling students' ethnic identity development and social-cognitive maturity scores, yet, students' indicated personal growth and awareness, which are relevant implications for SoTL. However, interpretation of the study findings should be done with caution as self-report data can be biased (Fraenkel et al., 2011). In other words, students may have responded on the SPMPGGEQ in a socially desirable way and/or felt they changed when change may not have occurred. Therefore, graduate counseling faculty members can integrate multicultural personal growth groups to assess students' ethnic identity and social cognitive development with measures that do not solely rely on student perceptions (e.g., observation checklists) in a time-series research design. Graduate students may experience development due to the introspective nature of the group activities. Moreover, as group experiences are required in graduate counseling curricula for accreditation (Council for the Accreditation of Counseling and Related Educational Programs; CACREP, 2009), programs may want to incorporate these multicultural group experiences to provide cathartic opportunities for graduate students. In addition, a multicultural-focused group experience may be used as part of group counseling courses to encourage self-reflection of cultural identity in a group setting.

To investigate the efficacy of multicultural personal growth groups in the future, faculty can strengthen the pedagogical intervention by increasing the duration of the groups in order to support graduate students' developmental growth. The present study utilized a six-week group curriculum; however, a longer group experience (e.g., 15 weeks) may result in a change in graduate counseling students' ethnic identity and social cognitive development. Likewise, the

intensity of the pedagogical group experience may be increased by modifying activities within the group curriculum to create more cognitive dissonance in students, which may result in additional intrapersonal growth.

L. Limitations of the Study.

Limitations are present in all studies and need to be appreciated when interpreting results. A limitation of our study was convenience sampling (participants from one university) and the sample size ($N = 20$), restricting the generalizability of the findings and potentially influencing the statistical analysis results. Second, as ethnic identity development and social cognitive maturity may occur over longer periods of time, the pedagogical intervention may have been too short to influence change in students' development.

M. Recommendations for Future Research.

In spite of the stated limitations, the findings of this preliminary study offer suggestions for effective SOTL research, which include: (a) having a long-term intervention (e.g., semester-long) and (b) using data collection methods that are not solely self-report. Additionally, future research can focus on obtaining a larger sample size in order to increase generalizability of the study's findings. More study participants engaged in the pedagogical intervention from various graduate counseling programs may yield statistically significant findings. Furthermore, the inclusion of a control group that is not receiving the pedagogical invention may strengthen the research design.

Multicultural competence is supported as a necessary skill of counselors-in-training and is positively correlated with ethnic identity and social cognitive development (Evans & Foster, 2000; Middleton et al., 2005; Neville et al., 1996; Vinson & Niemeyer, 2000; Watt et al., 2002). We investigated the impact of a pedagogical intervention (multicultural personal growth group) graduate counseling students' levels of ethnic identity development and social-cognitive maturity, and their perceptions of their experience with this experiential group intervention. The results were limited; nevertheless, the graduate student survey data indicated the inclusion of a group experience was beneficial to group participants' growth. Therefore, our study offers a sound initial investigation of the instructors' pedagogy and offers areas for future SoTL research.

References

- American Counseling Association (2005). *ACA Code of Ethics*. Alexandria, VA.: Author.
- Arredondo, P., Toporek, M. S., Brown, S., Jones, J., Locke, D. C., Sanchez, J. & Stadler, H. (1996) Operationalization of the Multicultural Counseling Competencies. Alexandria, VA: AMCD.
- Arthur, N., and Achenbach, K. (2002). Developing multicultural competencies through experiential learning. *Counselor Education and Supervision, 42*, 2-4.
- Borders, L.D., Fong, M.L., and Neimeyer, G.J. (1986). Counseling students' level of ego development and perception of clients. *Counselor Education and Supervision, 26*, 36-49.
- Cannon, E.P., and Frank, D. (2009). Promoting ego development and multicultural competence during internship. *International Journal for the Advancement of Counseling, 31*(3), 199-211.
- Chao, R. (2006). Counselors' multicultural competencies: Race, training, ethnic identity, and color-blind racial attitudes. In G. Walz, J. C. Bleuer, & R. K. Yep (Eds.), *Vistas: Compelling perspectives on counseling* (73-76). Alexandria, VA: American Counseling Association.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1*(3), 98-101.
- Council for Accreditation of Counseling and Related Educational Programs. (2009). *2009 Standards for Accreditation*. Alexandria, VA: Author
- Evans, K.M., and Foster, V.A. (2000). Relationships among multicultural training, moral development and racial identity development of White counseling students. *Counseling and Values, 45*, 39-48.
- Faith, M.S., Wong, F.Y., and Carpenter, K.M. (1995). Group sensitivity training: Update, meta-analysis and recommendations. *Journal of Counseling Psychology, 42*(3), 390-399.
- Fraenkel, J., Wallen, N., and Hyun, H. (2011). *How to design and evaluate research in education*. (8th ed.). New York: McGraw-Hill Companies, Inc.
- Hart, B., and Hilton, I. (1988). Dimensions of personality organization as predictors of teenage pregnancy risk. *Journal of Personality Assessment, 52*, 116-132.
- Hauser, S.T. (1976). Loevinger's model and measure of ego development: A critical review. *Psychological Bulletin, 83*, 928-955.
- Helms, J. E. (1996). Toward a methodology for measuring and assessing racial as distinguished from ethnic identity. In G. R. Sodowsky & J. C. Impara (Eds). *Multicultural Assessment in Counseling and Clinical Psychology*. Lincoln, NE: Buros Institute of Mental Measurement.

Hill, N. (2003). Promoting and celebrating multicultural competence in counselor trainees. *Counselor Education & Supervision, 43*, 39-51.

Hutchings, P., Huber, M.T., and Ciccone, A. (2011). *The Scholarship of Teaching and Learning Reconsidered: Institutional Integration and Impact*. San Francisco, CA: Jossey-Bass.

Hy, L., and Loevinger, J. (1996). *Measuring Ego Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Joyce, A., MacNair-Semands, R., Tasca, G., & Ogrodniczuk, J. (2011). Factor structure and validity of the Therapeutic Factors Inventory-Short Form. *Group Dynamics: Theory, Research, and Practice, 15*(3), 201-219.

Lambie, G.W. (2007). The contribution of ego development level to burnout in school counselors: Implications for professional school counseling. *Journal of Counseling & Development, 85*(1), 82-88.

Lambie, G.W., Hagedorn, W.B., and Ieva, K.P. (2010). Social-cognitive development, ethical and legal knowledge, and ethical decision making of counselor education students. *Counselor Education and Supervision, 49*(4), 228-246.

Lambie, G.W., Smith, H.L., and Ieva, K.P. (2009). Graduate counseling students' levels of ego development, wellness, and psychological disturbance: An exploratory investigation. *Adultspan Journal, 8*(2), 114-127.

Leonard, P.J. (1996). Consciousness-raising groups as a multicultural awareness approach: An experience with counselor trainees. *Cultural Diversity and Mental Health, 2*(2), 89-98.

Lieberman, M.A., Yalom, I.D., and Miles, M.B. (1973). *Encounter groups: First facts*. New York: Basic Books.

Loevinger, J. (1976). *Ego Development*. San Francisco, CA: Jossey-Bass.

Loevinger, J. (1979). Construct validity of the Sentence Completion Test of ego development. *Applied Psychological Measurement, 3*, 281-311.

Loevinger, J. (1998). *Technical Foundations for Measuring Ego Development: The Washington University Sentence Completion Test*. Mahwah, NJ: Lawrence-Erlbaum.

Manners, J., and Durkin, K. (2000). Processes involved in adult ego development: A conceptual framework. *Developmental Review, 20*(4), 475-513.

Middleton, R., Stadler, H., Simpson, C., Guo, Y., Brown, M., Crow, G., and Lazarte, A. (2005). Mental health practitioners: The relationship between white racial identity attitudes and self-reported multicultural counseling competencies. *Journal of Counseling and Development, 83*, 444-456.

Neville, H.A., Heppner, M.J., Louie, C.E., Thompson, C.E., Brooks, L., and Baker, C.E. (1996). The impact of multicultural training on White racial identity attitudes and therapy competencies. *Professional Psychology: Research and Practice*, 27(1), 83-89.

Novy, D.M., and Francis, D.J. (1992). Psychometric properties of the Washington University Sentence Completion Test. *Educational and Psychological Measurement*, 52(4), 491-494.

Phinney, J. (1989). Stages of ethnic identity development in minority group adolescents. *Journal of Early Adolescent*, 9, 34-49.

Phinney, J. (1990). Ethnic identity in adolescent and adulthood: A review of research. *Psychological Bulletin*, 108, 499-514.

Phinney, J. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7, 156-176.

Phinney, J., and Alipuria, L. (1990). Ethnic identity in college students from four ethnic groups. *Journal of Adolescence*, 13, 171-183.

Phinney, J., and Ambarsoom, H. (1987). Ethnic identity in young adults from three ethnic groups. ERIC Document Reproduction Service No. ED 283 057.

Phinney, J., and Baldelomar, O. (2006). *The conceptualization and measurement of ethnic identity development: Exploration, commitment, and statuses*. Unpublished manuscript.

Phinney, J., and Ong, A. (2006). *Testing competing models of the structure of ethnic identity with the Multigroup Ethnic Identity Measure*. Unpublished manuscript.

Phinney, J., and Ong, A. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 3(54), 271-281.

Phinney, J., and Tarver, S. (1988). Ethnic identity search and commitment in Black and White eighth graders. *Journal of Early Adolescence*, 8, 265-277.

Ponterotto, J.G., and Potere, J. C. (2003). The Multicultural Counseling Knowledge and Awareness Scale (MCKAS). In D. Pope-Davis, H. Coleman, W. Liu, & R. Toporek (Eds). *Handbook of Multicultural Competencies in Counseling & Psychology*. Thousand Oaks, CA: Sage Publications, Inc.

Pope-Davis, D., Breaux, C., and Liu, W.M. (1997). A multicultural immersion experience: Filling a void in multicultural training. In D. B. Pope-Davis & H. L. K. Coleman (Eds.), *Multicultural Counseling Competencies: Assessment, Education and Training, and Supervision* (pp. 237-241). Thousand Oaks, CA: Sage.

- Pope-Davis, D.B., and Ottavi, T.M. (1994). The relationship between racism and racial identity among White Americans: A replication and extension. *Journal of Counseling & Development, 72*(3), 293-297.
- Ridley, C.R., and Lingle, D.W. (1996). Cultural empathy in multicultural counseling: A multidimensional process model. In P. Pedersen, W. Looner, & J. Draguns (Eds.), *Counseling Across Cultures* (pp. 21-45). Thousand Oaks, CA: Sage.
- Ridley, C.R., Espelage, D.L., and Rubenstein, K. (1997). Course development in multicultural counseling. In D. B. Pope-Davis & H. L. K. Coleman (Eds.), *Multicultural Counseling Competencies: Assessment, Education and Training, and Supervision* (pp. 131-158). Thousand Oaks, CA: Sage.
- Rowell, P.C., and Benshoff, J. (2008). Using personal growth groups in multicultural counseling courses to foster students' ethnic identity development. *Counselor Education & Supervision, 48*, 2-15.
- Sias, S.M., Lambie, G.W., and Foster, V.A. (2006). Conceptual and moral developmental of substance abuse counselors: Implications for training. *Journal of Addictions and Offenders Counseling, 26*, 99-110.
- Sobansky, R., Gutkin, T., Galloway, A., Saunders, A., Yetter, G., and Song, S. (2010). Moving toward a culturally invariant measure of ethnic identity among diverse adolescents: Redefining the MEIM. *International Journal of Psychology: A Biopsychosocial Approach, 7*, 85-110.
- Statistical Package for the Social Sciences* (Version 19.0) [Computer software]
- Syed, M., and Azmitiz, M. (2008). A narrative approach to ethnic identity in emerging adulthood: Bringing life to the identity status model. *Developmental Psychology, 44*(4), 1012-1027.
- Tabachnick, B., and Fidell, L. (2007). *Using multivariate statistics*. (5th ed.). Boston: Pearson Education, Inc.
- Utsey, S.O., Chae, M.H., Brown, C.F., and Kelly, D. (2002). Effect of ethnic group membership on ethnic identity, race-related, stress and quality of life. *Cultural Diversity and Ethnic Minority Psychology, 8*(4), 366-277.
- Vasquez, L.A. (1997). A systemic multicultural curriculum model: The pedagogical process. In D. B. Pope-Davis & H. L. K. Coleman (Eds.), *Multicultural Counseling Competencies: Assessment, Education and Training, and Supervision* (pp. 131-158). Thousand Oaks, CA: Sage.
- Villalba, J.A., and Redmond, R.E. (2008). "Crash": Using a popular film as an experiential learning activity in a multicultural counseling course. *Counselor Education and Supervision, 47*(4), 264-276.

Johnson, J. M. and Lambie, G. W.

Vinson, T., and Neimeyer, G. (2000). The relationship between racial identity development and multicultural counseling competency. *Journal of Multicultural Counseling and Development, 28*(3), 177-193

Warren, E.S. (2008). *Ego development and theoretical orientation among counseling students*. (Doctoral dissertation). Retrieved from Dissertations & Theses.

Watt, S.K., Robinson, T.L., and Lupton-Smith, H. (2002). Building ego and racial identity: Preliminary perspectives on counselors-in-training. *Journal of Counseling and Development, 80*(1), 94-100.

Westenberg, P.M., and Block, J. (1993). Ego development and individual differences in personality. *Journal of Personality and Social Psychology, 65*, 792-800.

Yalom, I.D., and Leszcz, M. (2005). *Theory and Practice of Group Psychotherapy* (5th ed.). New York: Basic Books.

Zinn, B. (1995). The relationship of ego development and the counseling effectiveness of counselor trainees. (Doctoral dissertation, The University of Missouri at St. Louis, 1995). *Dissertation Abstracts International, 56*, 7082.

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

Robert Lake and Kent Rittschof¹

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 as Banks (2006), Sleeter and Grant (2009), and Nieto 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 Rasch 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.

	<u>Z Infit</u>	<u>Z Outfit</u>	Pre	Post	Change
<u>Course Inconsistent Items on <i>Homosexuality</i></u>					
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
<u>Course Consistent Items on <i>Homosexuality</i></u>					
22. mental illness	-.3	.0	0.58	-0.25	-0.83
34. many contributions	.2	.6	0.44	-0.44	-0.88
<u>Course Inconsistent Items on <i>Race</i></u>					
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
<u>Course Consistent Items On <i>Race</i></u>					
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
<u>Course Inconsistent Items on <i>Social Class</i></u>					
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
<u>Course Consistent Items on <i>Social Class</i></u>					
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
<u>Course Inconsistent Items on <i>Women's Equality</i></u>					
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 <i>Women's Equality</i>					
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.

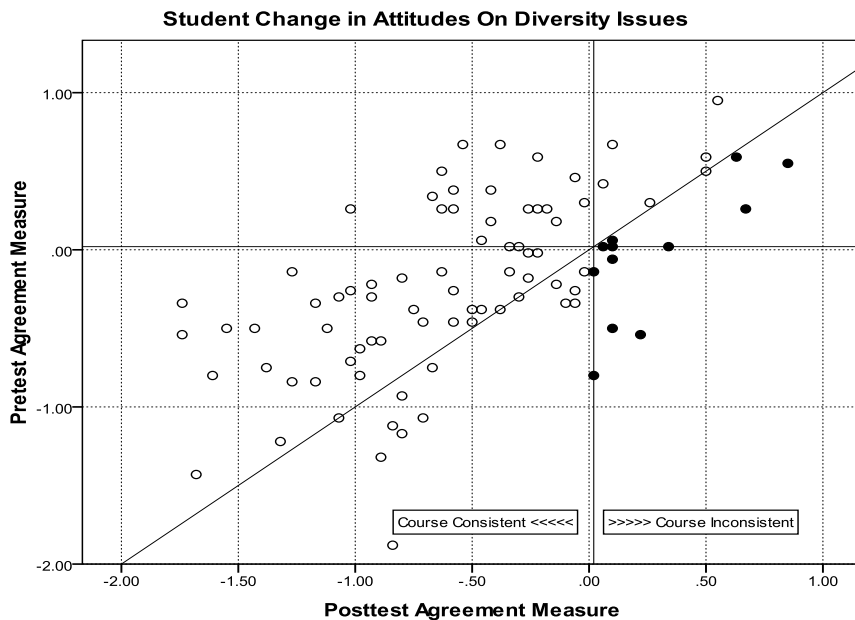


Figure 1. Pretest to posttest change in attitudes on course inconsistent diversity statements. Lower measures represents course consistent attitudes. Students located to the left of the diagonal line showed measured change toward increasingly course consistent diversity attitudes. Students located to 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 in fit 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.

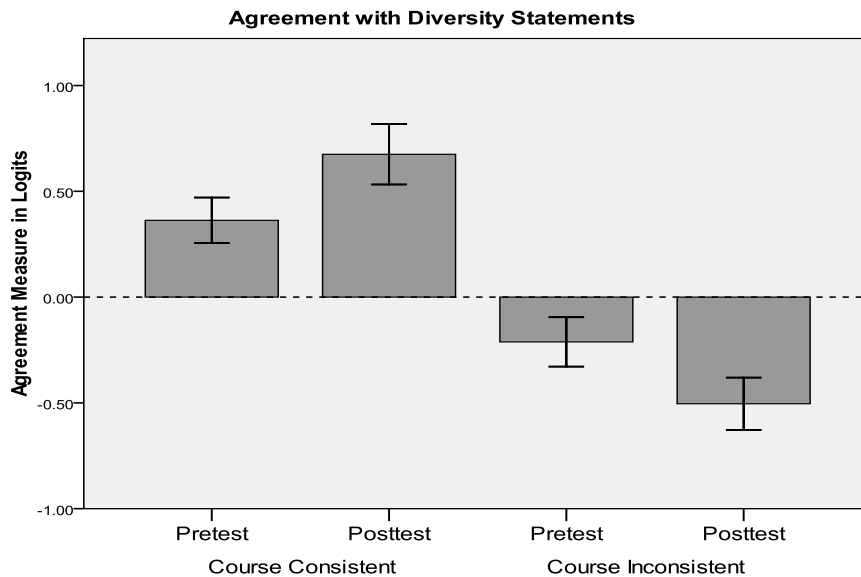


Figure 2. Comparisons of pretest and posttest measures of agreement (and 95% confidence intervals) with diversity item statements that are course consistent and course inconsistent. The mean measure for each comparison is set at zero logits. Effect sizes for changes illustrated were $d = .52$ standard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

Lake, R. and Rittschof, K.

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*, 17 (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.

Lake, R. and Rittschof, K.

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.

Book Review

Team Teaching: Across the Disciplines, Across the Academy

Richard L. Mehrenberg¹

Citation: Plank, K. M. (Ed.). (2011). *Team teaching: across the disciplines, across the academy*. Sterling, Va.: Stylus Pub.

Publisher Description: For those considering adopting team teaching, or interested in reviewing their own practice, this book offers an over-view of this pedagogy, its challenges and rewards, and a rich range of examples in which teachers present and reflect upon their approaches.

The interaction of two teachers—both the intellectual interaction involved in the design of the course, and the pedagogical interaction in the teaching of the course—creates a dynamic environment that reflects the way scholars make meaning of the world. The process naturally breaks down the teacher-centered classroom by creating a scholarly community in which teachers and students work together to understand important ideas, and where students don't just learn content, but begin to understand how knowledge is constructed, grasp the connections between disciplines as well as their different perspectives, see greater coherence in the curriculum, and appreciate how having more than one teacher in the classroom leads naturally to dialogue and active learning.

Regardless of discipline, college instructors seek out new ways to help their students to master content, make new connections, and better appreciate the complexities of their field. In the book, *Team Teaching: Across the Disciplines, Across the Academy*, editor, Kathryn M. Plank attempts to address these goals through the pedagogical strategy of team teaching.

Based upon the philosophy that “two heads are better than one”, Plank describes many of the perceived benefits of team teaching. Some of the most important reasons given to encourage instructors to team teach are that it promotes deeper cognitive and ethical thinking among students, it allows cross-disciplinary studies to occur within the same classroom, and that it allows instructors to improve their own craft through ongoing observation and reflection of a colleague's teaching style.

Plank acknowledges that there are also several obstacles that may prevent quality team-teaching to occur. Perceived inconsistencies in grading and possible increased costs are two challenging, but surmountable problems. Plank asserts that although there may be some initial struggles to overcome, the potential rewards for teachers and students are worth it.

The remainder of the book consists of five case studies about team teaching at the college level. *Chapter One* shares the story of a paleontologist and a molecular biologist who decided to team teach a science class around the theme of “origins”. The authors stressed the importance of having students examine the scientific method through two diverse, but equally valid lenses. Challenges included figuring out how to successfully balance instructional responsibilities, and a perceived inequity of instructor authority based on gender .

Chapter Two describes the diverse course offering between a philosopher and a biologist.

¹ Assistant Professor of Special Education at Millersville University

The instructors discussed how they incorporated lecture, writing, and hands-on activities to help first year students explore diseases and their impact on society. The biggest struggle for the pair was ensuring that assessments were viewed as valid and fair by students. They plan to address this issue in future sections through the incorporation of student generated rubrics.

Shifting from the sciences to the arts, *Chapter Three* concerns the combined efforts of a musician and an artist. Their class explored the various ways that art impacts the community. The authors include their actual course syllabus so that the reader can get a more detailed understanding of the class, student expectations, and logistics. Field trips, guest speakers, and service requirements were all techniques used to immerse students in authentic learning opportunities. The authors mentioned that the ongoing support by their administration and colleagues greatly enhanced their experience.

The team teachers in *Chapter Four* highlighted how their personal differences that enhanced their experience in teaching for a course on racism. A black female social work professor and her white male colleague taught a course titled, *Racism in the United States: Implications for Social Work Practice*. The authors stated that the course gave students an open forum to ask questions and discuss topics with instructors from significantly different life experiences and backgrounds. Although it was perceived that a few students pushed boundaries, or took course topics personally, the instructors felt that the team teaching experience modeled a professional example of shared power and mutual respect.

Chapter Five concludes the book with the recollections of an instructor on his team teaching experiences from twenty years ago. The author took part in a scientific literacy course for non-science majors. A revolving panel of faculty from different fields took part in this basic course to give students a hands-on approach to the scientific method. The author explains how the course was successful, but was eventually abandoned as being deemed too cost-restrictive to run.

This text is recommended for college instructors with little to no familiarity with team teaching. The case studies, syllabi, and course outlines provides the reader with tangible examples of how this particular instructional technique could be utilized in any number of different environments. The reader would also learn to how to avoid, or at least be aware of, some of the more common challenges faced by teaching teams.

There are also two main criticisms of the book. In the introduction, Plank notes that the average college instructor lags behind his K-12 counterpart when it comes to the implementation of innovative pedagogy. This factor is especially true when it comes to co-teaching. Beyond team teaching, there exists a robust body of K-12 literature of effective ways for two or more teachers to share teaching responsibilities within classroom. Station, parallel and alternative teaching are but a few of the other common co-teaching models. Case studies that employed these methodologies would be a helpful inclusion to demonstrate the wide variety of co-teaching choices.

A second criticism deals with the lack of empirical evidence regarding the effectiveness of team teaching at the college level. The introductory chapter focuses mainly on descriptive, rather than empirical data. Likewise, few of the case studies included student feedback, final grades, or peer observations that could be used as an objective measure of the methodology's success. Additional inclusion of this type of data would give readers additional confidence when attempting to implement team teaching on their own.

Mission

Founded in 2001, the Journal of the Scholarship of Teaching and Learning (JoSoTL) is a forum for the dissemination of the Scholarship of Teaching and Learning in higher education for the community of teacher-scholars. Our peer reviewed Journal promotes SoTL investigations that are theory-based and supported by evidence. JoSoTL's objective is to publish articles that promote effective practices in teaching and learning and add to the knowledge base.

The themes of the Journal reflect the breadth of interest in the pedagogy forum. The themes of articles include:

1. Data-driven studies: formal research projects with appropriate statistical analysis, formal hypotheses and their testing, etc. These studies are either with a quantitative or qualitative emphasis and authors should indicate the appropriate domain. Acceptable articles establish a research rigor that leads to significant new understanding in pedagogy.
2. Reflective essays: integrative evaluations of other work, essays that challenge current practice and encourage experimentation, novel conclusions or perspectives derived from prior work
3. Reviews: Literature reviews illuminating new relationships and understanding, meta-analysis, analytical and integrated reviews, etc.
4. Case studies: These studies illustrate SOTL and its applications, usually generalizable to a wide and multidisciplinary audience.
5. Comments and communications: Primarily, these are comments based on previously published JoSOTL articles, but can also include book reviews, critiques and evaluations of other published results in new contexts or dimensions

Submissions

Authors are encouraged to submit work in one of the following categories:

- **Traditional Research Reports: data driven studies with either a quantitative or qualitative emphasis**
- **Reflective Essays on SoTL**
- **Reviews of current themes in SoTL research including meta-analysis**
- **Case studies illustrating SoTL and its applications**
- **Comments and Communications on previous Journal articles, or book or software reviews**

All submissions for JoSoTL should be submitted using the online submission process on our website beginning on July 1, 2012.

Please follow this link to get more detailed information on the submission process for the Journal of the Scholarship of Teaching and Learning:

<http://josotl.indiana.edu/about/submissions#onlineSubmissions>

If you have additional questions or run into problems please contact the journal at josotl@iu.edu.

Editorial Board

All members of the **JoSoTL Editorial Board** are affiliated with **FACET**, the Faculty Colloquium on Excellence in Teaching, at **Indiana University**.

<p>Don Coffin Associate Professor of Economics</p>	<p>dcoffin@iun.edu Division of Business and Economics Indiana University Northwest 219.980.6913</p>
<p>David J. Malik Chancellor's Professor of Chemistry Executive Vice Chancellor of Academic Affairs</p>	<p>dmalik@iupui.edu or dmalik@iun.edu Indiana University Purdue University Indianapolis Administration IU Northwest 219.980.6707</p>
<p>Eugenia Fernandez Associate Chair of Computer & Information Technology</p>	<p>efernand@iupui.edu Purdue School of Engineering and Technology Indiana University Purdue University Indianapolis 317.274.6794</p>
<p>Joan E. Lafuze Professor of Biology</p>	<p>jlafuze@indiana.edu Instructional Programs Indiana University East, Richmond, IN 765.973.8246</p>
<p>Julie Saam Associate Professor of Secondary Science Education</p>	<p>jsaam@iuk.edu Division of Education Indiana University Kokomo, IN 765.455.9302</p>
<p>Ellen A. Sigler Professor and Department Head</p>	<p>elsigler@wcu.edu Educational Leadership Western Carolina University 828.227.7415</p>
<p>Beth Kern Associate Director of FACET for The Mack Center for Inquiry on Teaching and Learning Associate Professor of Accounting</p>	<p>bkern@iusb.edu School of Business & Economics Indiana University South Bend 574.520.4352</p>
<p>Robin K. Morgan Director, FACET Professor of Psychology</p>	<p>rmorgan@ius.edu Indiana University Southeast, New Albany, IN 812.941.2298</p>

Style Sheet for the *Journal of the Scholarship of Teaching and Learning*

John Dewey¹ and Marie Curie²

Abstract: This paper provides the style sheet for the Journal of the Scholarship of Teaching and Learning. Manuscripts submitted for publication should adhere to these guidelines.

Keywords: radiation, metacognition, identity theory, constructivism, educational philosophy.

I. General Guidelines for the Manuscript.

The final manuscript should be prepared in 12-point, Times New Roman, and single-spaced. Submissions should be double-spaced. All margins should be 1 inch. The text should be fully left- and right-justified. The title (in 16 point bold) and author's name (in 12 pt. bold) should be at the top of the first page. The author's name should be followed by a footnote reference that provides the author's institutional affiliation and address. The abstract should be indented 0.5" left and right from the margins, and should be in italics.

Except the first paragraph in a section subsequent paragraphs should have a 0.5" first line indent. Use only one space after the period of a sentence (word processors automatically adjust for the additional character spacing between sentences). The keywords should be formatted identically to the abstract with one line space between the abstract and the keywords. Authors should use keywords that are helpful in the description of their articles. Common words found in the journal name or their title article are not helpful.

Pages should be unnumbered since they will be entered by the Journal editorial staff. We will also insert a header on the first page of the article, as above.

References should be incorporated in the text as authors name and date of publication (Coffin, 1993), with a reference section at the end of the manuscript (see below for the desired format for the references). Titles of articles should be included in the references in sentence case. Unless instructed otherwise in this Style Sheet, please use APA style formatting. Footnotes should incorporate material that is relevant, but not in the main text.

A. Plagiarism.

It is essential that authors refrain from plagiarism. Plagiarism is a violation of ethics and, in serious cases, will lead to a manuscript being rejected by this journal. No future manuscripts will be accepted from authors who have submitted a plagiarized manuscript.

¹Department of Educational Philosophy, Indiana University Northwest, 3400 Broadway, Gary, IN 46408, jdewey@iun.edu.

²Institut Pasteur, University of Paris, 75015 Paris, France.

B. Unique work.

This journal does not accept previously published work. We also do not accept work that is being considered for publication by another journal. If your manuscript is accepted, you will be required to sign a form stating that your manuscript has not been previously published.

II. Section and Sub-Section Headings.

A. Major Sections.

Major section headings should be flush-left, bold-faced, and Roman numeral numbered. Major section headings should have one-line space before and after. The first paragraph(s) of the article do not require a major heading.

B. Sub-Sections.

Sub-section headings should also be flush-left, in italics, and alphabetically numbered. Sub-section headings should have a one-line space before and after. Sub-sub-sections should appear at the beginning of a paragraph (i.e., with an 0.5" indent, followed immediately by the text of the sub-sub-section), with the heading also in italics.

III. Tables and Figures.

Tables and figures should be inserted in the text where the author believes they best fit. They may be moved around a little to better correspond to the space requirements of the Journal. If necessary, tables and figures may occupy an entire page to ensure readability and may be in either portrait or landscape orientation. Insofar as possible, tables should fit onto a single page. All tables and figures should be germane to the paper. Tables should be labeled as follows with the title at the beginning (in bold), with data entries single-spaced, and numbered. Column labels should be half-line spacing above data.

Table 1. The title of the table.

Unit	Length, inches
Point	1/12
Pica	1/6

Figures should have their captions follow the image. Captions should be single-spaced, with title in bold. Additional text should not be in bold. The Editorial staff may adjust layout to allow optimal use of space.

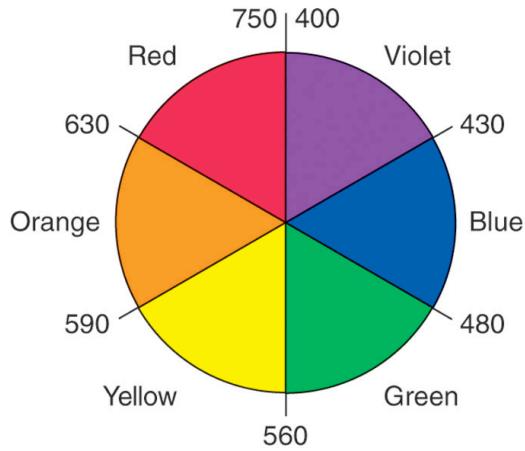


Figure 1. Color wheel with wavelengths indicated in millimicrons. Opposite colors are complementary.

Acknowledgements

Acknowledgements should identify grants or other financial support for this research by agency (source) and number (if appropriate). You may also acknowledge colleagues that have played a significant role in this research.


Appendix

Please insert any appendices after the acknowledgments. They should be labeled as follows:

Appendix 1. The Title of the Appendix.

References

- Coffin, D.A. (1993). Using the competitive edge. *Journal of Economic Education*, 24 (1), 62-69.
- Garcia, J. and Rodriguez, P. (2002). The determinants of football match attendance revisited: Empirical evidence from the Spanish football league. *Journal of Sports Economics*, 3 (1), 18-38.
- Hamilton, S. J. (1995). *My name's not Susie: A life transformed by literacy*. Portsmouth, NH: Boynton/Cook Publishers.
- Pappas, D. (2004). Fixing the fan cost index: A more realistic view. Retrieved April 21, 2004, from <http://www.baseballprospectus.com/article.php?articleid=2790>.



Contact Info for the Journal

JoSoTL Editorial Office

Indiana University Purdue University Indianapolis
755 W. Michigan St, UL 1180D
Indianapolis, IN 46202

josotl@iupui.edu

ISSN: 1527-9316