ITEM RESPONSE THEORY SCALING OF AN ACADEMIC MEDICAL CENTER STUDENT SATISFACTION SURVEY

BY

Christopher S. Meiers

Submitted to the graduate degree program in the
Department of Psychology and Research in Education
and the Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

_____
Chairperson

Committee Members _____

_____

_____

_____

Date defended: _____

The Dissertation Committee for Christopher S. Meiers certifies
that this is the approved version of the following dissertation

ITEM RESPONSE THEORY SCALING OF AN ACADEMIC MEDICAL CENTER
STUDENT SATISFACTION SURVEY

_____
Chairperson

Date approved: _____

ABSTRACT

Item response theory scaling has been well established in achievement testing, but the practice has seen limited use in student development research. The purpose of this study was to explore the reliability and validity of a graded response model scaling of an existing college student satisfaction survey. The existing survey was found to measure three aspects of student satisfaction: community, diversity, and substance abuse. The three scales were found to have acceptable internal consistency. Parameter estimates using the grade response model were able to be obtained for the community and diversity but not the substance abuse subscale. A reliable community satisfaction score scale was created for the community subscale.

Validity evidence was accumulated based on established research about student academic program, student involvement, and mode of survey. Differences in community satisfaction scores between involved and not-involved students were present, differences in community satisfaction scores between online- and paper-based surveys were present, and differences between community satisfaction scores and student academic program were not present.

ACKNOWLEDGMENTS

I want to thank my dissertation committee for guiding me through this experience: Drs. Rita Clifford, Bruce Frey, Dongbin Kim, Vicki Peyton, and William Skoruspki. I also want to acknowledge Dr. Skorupski's role as my advisor. His passion for the field is an inspiration and encouraged me to pursue this research. I am grateful for everything that I have learned from him and look forward to future collaborations with Billy as a colleague.

I have worked with an amazing and talented group of professionals at the University of Kansas and in the profession. I do want to personally thank Dr. Cindy Derritt, Dr. Lisa Erwin, Anne Flaherty, Ryan Gove, Dr. Joey Linn, Dr. Sheri Blessing Phillips, Jeremy Ring, Dr. Marlesa Roney, Dr. Herb Songer, Robert Turvey, and my staff for their support, impact on my professional life, and friendship. Thank you to Dr. Jeff Radel for volunteering as a dissertation reader. Special acknowledgment goes to Fonda Briles. If it wasn't for her, I would be lost on a daily basis.

I cannot adequately express how Dr. Dorothy Knoll, Dean of Students at KUMC, has impacted my life. This research simply would not have been possible without her and I would not be the professional or person that I am today. This research is dedicated to her career as a professional, student advocate, and mentor to many in the profession.

And finally, absolutely none of this would be possible without the love and support of my family and friends. My parents, Tom and Mary Meiers have always been there for me and I cannot thank them enough. My son, Jackson, is my inspiration. The support of my sister, Jenny Lutters, and her husband, Corey, as well as a great group of friends mean the most to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF APPENDICES

CHAPTER ONE – INTRODUCTION

The practice of student affairs in higher education is primarily focused on the non-curricular contexts of a student's experience. The prominence of assessment in student affairs started in the 1990s and has become an essential activity at higher education institutions. Assessment in student affairs has been determined necessary for accountability purposes, strategic planning, policy development and decision-making, political reasons, and accreditation purposes (Upcraft & Schuh, 2000a). Student satisfaction assessment is a process to determine how students evaluate the quality of their experience (Upcraft & Schuh, 2000b) and is a vital assessment activity in order to determine the effictiveness of student development programming and services. Measuring student satisfaction plays a primary role in the practice of student affairs due to the research that has identified what students gain from being satisfied with their educational experience (Astin, 1993).

Measuring student satisfaction in the context of environmental assessment is by far the most difficult and challenging type of assessment in the practice of student affairs (Astin, 1991). What confounds attempting to measure student satisfaction is identifying what is to be measured, that most student environments have the same level of self-production, and that external factors often cannot be controlled.

The most prominent national measure of student development is the National Survey of Student Engagement (NSSE). The NSSE collects information from hundreds of colleges on an annual basis about student participation in programs and services the institutions provide for student learning and personal development. In 2009, 643 colleges and universities participated in the survey (Quick Facts about NSSE). NSSE provides an estimate of how undergraduate students spend their time and what they gain from college. Due to the popularity of NSSE, it has

been adopted for specific groups of students such as community college and law school students (About the National Survey of Student Engagement).

NSSE has been demonstrated to have psychometric properties that measure a student's opinion about the campus climate (Kuh, 1991). Surveys like NSSE primarily focus on developmental issues and what students gain from college and don't necessarily focus on satisfaction. Student affairs professionals are interested in understanding what students gain from college but are also equally interested if students are satisfied with the services since participation in services is related to satisfaction (Wiers-Jenssen, Stensaker, & Grøgarrd, 2002). In addition, the NSSE survey has recently been criticized of having limited validity for its intended purposes and it has been suggested that new approaches to surveying college students are warranted (Schmidt, 2009).

A student's experience at an academic medical center can be remarkably different from an undergraduate student's experience at a traditional college campus. Prior to enrollment at an academic medical center, health professional students have a well-defined focus on a career choice and most of their curricular and non-curricular activities are related to their academic preparation. The demographics of students at an academic medical center are remarkably different than an undergraduate campus. The overwhelming majority of the health professional students are at the graduate or professional level (as opposed to the undergraduate level), and on average, students at an academic medical center are older, more likely to have a significant spouse or partner and have children. Most of the related literature and research findings about student satisfaction have been based on samples of traditional undergraduate students in traditional academic programs (Aitken, 1983; Aldridge & Rowley, 1998; Athiyaman, 1997; Elliot & Shin, 2002; & Gatfield, 1999). In addition, accreditation bodies of health professional

programs have increasingly become interested in exploring data about student satisfaction due to the relationship between a student's satisfaction, the quality of their educational experience, and patient care (National Academy of Sciences, 2003).

For the reasons mentioned above, the Division of Student Services at the University of Kansas Medical Center (KUMC) started to conduct the KUMC Student Services in 1993. The KUMC Student Services Survey contains questions related to the KUMC environment, student satisfaction, diversity issues, and student wellness. The survey is sent to all KUMC students and responses are completely confidential (Erwin & Meiers, 2006). The results of the survey primarily are used to identify areas of concerns to students at KUMC. Examples of previous areas of concern that have been identified through the survey previously include issues related to sexual harassment and diversity (Knoll & Erwin, 2001).

Like most self reported student surveys that are created at higher education institutions, the KUMC Student Services Survey was designed to quantify student beliefs about the educational and non-educational experience into aggregate groups such as school affiliation, gender, ethnicity, and student involvement. Differences between groups were determined by statistical tests such as analysis of variance. If the statistical tests were found to be significant, it was concluded that student beliefs were different. A major limitation of this approach is the findings are dependent on the group of students taking the survey, and the overall validity of the instrument in measuring the constructs of interest is not addressed.

The most predominant way of scaling student development surveys is based on classical test theory approaches as opposed to item response theory (IRT) approaches due to the complexity of modeling polytomous data in IRT models and the underlying assumption of unideminsionality that is needed for IRT models. This can be demonstrated by simply

conducting a search for "IRT" in the two major student development publications, *Journal of College Student Development* and *Journal of Student Affairs Research and Practice*. Only one article has been published in the *Journal of College Student Development* that used an item response theory approach and it was for the scaling of an intellectual development instrument not a student development/satisfaction instrument (DeMars &Erwin, 2003). The five benchmarks that are measured by the popular NSSE were created through principal component analysis and don't address item quality despite the wide spread adoption of item response theory models in achievement testing (Kuh et al, 2001). The constructors of the NSSE have concluded that the instrument measures the intended constructs and estimate the stability of the test through three classical approaches: correlation of concordance, comparison of means, and test-retest reliability.

Purpose

The purpose of this study was to determine if student satisfaction can be measured on the KUMC Student Services survey with reliability and validity. The terms "student satisfaction" and "student involvement" are used throughout this study. Student satisfaction is the hypothesized latent variable of interest that is being measured by the KUMC Student Services Survey[1]. There were no preconceived hypotheses about the structure of the underlying variables on the KUMC Student Services Survey. The first step of the study was to identify the structure and nature of the underlying variables. Once the structure was identified, the underlying variables were scaled and then the validity of the scaling was determined based on established student development literature, established literature on survey administration, and a working knowledge of the nature of health professional students at KUMC. This validation process is a continuous one with this study providing a beginning in the exploration of what aspects of student satisfaction the KUMC Student Services Survey measures and how to best understand

[1] Specific definition terms are located in Appendix B.

what those scores mean. There were several variables in this study. The dependent variable was the community satisfaction score. The independent variables were the mode of survey, student academic school and student involvement.

Research Questions

The exact research questions addressed in this study were the following. These research questions were explored in order to assess the reliability of the KUMC Student Services Survey:

1. What underlying variables related to student satisfaction does the KUMC Student Services Survey measure?
2. Can the KUMC Student Services Survey be reliably scaled with an item response theory graded response model?

Hypotheses

The validity of the KUMC SSS in measuring the latent variables will be determined based on established research. Involved students are typically more satisfied with their educational experience (Astin, 1993; Kapp, 1979, Pascarella, 1980). Negligible differences in response between online and paper-based surveys have been determined (Deutsken et al, 2006; Carini et al, 2003; Epstein et al, 2001). Individual item response on the KUMC Student Services Survey has previously detected differences in student academic school (Erwin, 2003; Erwin & Meiers, 2005, 2006; Meiers & Gove, 2007). Specifically, it was hypothesized that:

1. There will be significant differences in student satisfaction related to whether a student is or is not involved in a student organization.
2. There will be no significant differences in student satisfaction related to whether the student completed an online- or paper-based survey.

3. There will be significant differences in student satisfaction related to a student's academic school.

Summary

This research addressed if a student satisfaction survey can be scaled with an item response theory model. Item level analysis is most commonly utilized in achievement testing and rather novel in student affairs research. A study that uses item response theory techniques to assess item quality in student development assessment techniques will have further application in evaluating student satisfaction and development. The instrument and scaling can be used at other academic medical centers for benchmarking purposes and could also possibly have applicability at traditional undergraduate universities. A more sophisticated scaling approach on a student development survey will also contribute to the advancement of quantitative methods in the student development profession.

CHAPTER TWO - REVIEW OF LITERATURE

This study examines the validity of an item response theory scaling of the KUMC Student Services Survey. This chapter is organized in sections. The first sections introduce the constructs of student satisfaction and involvement since they are the constructs that were chosen for part of the validation of the scaling. Information about online- and paper-based surveys are also presented since it is was used for the scaling validation. The conceptual mediation model that is used in this study is also discussed. The models and statistical procedures of reliability estimation, factor analysis, item response theory, and the graded response model are presented since they are part of the methodology. The chapter culminates with a review of literature about establishing reliability and validity with a measurement since it is integral to the research questions of the study.

Student Satisfaction and Involvement

Astin's (1993) research found that the following factors correlated with student satisfaction: relationships with faculty, curriculum and instruction, student life, individual support services, and facilities. Astin also found that student satisfaction of the collegiate experience is not dependent on factors prior to coming to college; rather student satisfaction is dependent on the collegiate environment. Faculty interest in student issues and availability of faculty are highly correlated to student satisfaction. The lack of campus community had the strongest negative effect on campus satisfaction. An institutional diversity emphasis is highly correlated to the student life factor as well as involvement in campus activities.

Astin (1993) also concluded that the following environmental factors are positively associated with student satisfaction: student orientation of the faculty, social change orientation, trust in the administration, diversity orientation, available resources, and reputation of the

institution. Also, according to Astin, involvements with faculty and student peer groups are also positively associated with student satisfaction.

A primary factor that affects student satisfaction is student and faculty relationships. Cultures between faculty and students can encourage or discourage student engagements in activities that will foster a student's academic and personal growth development (Kuh, 1991).

Astin (1999) defines student involvement as the amount of physical and psychological energy that the student devotes to the academic experience. A student who is highly involved spends considerable time studying and preparing for classes, participating in campus activities and organizations, and engages with university faculty and staff. Conversely, a non-involved student puts minimal effort into academic preparation[2], is not involved in campus organizations and has infrequent contact with other students and staff. Astin's involvement theory has five principles:

- Involvement may be highly generalized such as the student's overall experience or specific such as time spent studying for an exam.
- Involvement occurs along a continuum and the same student can manifest different amounts of involvement on different objects and different times.
- Involvement has quantitative and qualitative features.
- The amount of student learning and development is directly proportional to the quality and quantity of student involvement in that program.
- The effectiveness of any educational policy or practice is directly proportional to that policy or practice's ability to increase student involvement.

[2] Astin's research was focused on traditional undergraduate students. In the context of health professional students, student involvement can be related to the amount of time spent in academic preparation. Health professional students that are not involved in student organizations often chose not to get involved because of the amount of time that they are spending on academic preparation.

Astin's involvement theory focuses on the individual needs of the student and involves a flexible and eclectic approach to practice. In terms of curricular involvement, the theory focuses less on course content and pedagogical instruction and what the student does in class in terms of motivation to learn the content. The theory focuses on developmental outcomes and the assessments used to measure student involvement are often focused around the behavioral mechanisms that facilitate student development.

Implications for faculty of Astin's involvement theory include focusing less on content and actually on what the student is doing in terms of academic coursework. In regards to student personnel administrators, the theory provides a frame of reference for working with students. It provides a principle for understanding where a student is focusing their energy and providing policies and resources that foster student involvement.

Students involved in out of classroom activities are also more satisfied with their educational experience (Astin, 1993; Kapp, 1979; Pascarella, 1980) than non-involved students. Involvement in campus institutions and involvement with faculty have been found to assist students in the attainment of broad personal and social benefits (Carnegie Commission on Higher Education, 1974).

Specifically, student involvement has also been identified as important in student developmental factors. Involvement provides opportunities for the development of important skills such as leadership, teamwork, decision making, planning (Schuh & Laverty, 1983). Pascarella, Ethington, and Smart (1988) found that leadership activities are associated with gains in social concerns and altruistic values. Involvement allows for the increased capacity of developing mature, interpersonal relationships (Hood, 1984). Students have also attributed job success to their involvement activities in college (Kapp, 1979). Involvement has also been

attributed with a student's likelihood to provide financial support to an alma mater (Nelson, 1984).

Astin's Input-Environment-Outcome Model

One of the more popular models for assessment activities in higher education is Astin's (1991) input-environment outcome (I-E-O) model. The I-E-O model outlines that outcomes are a function of interactions of inputs and the environment.

The I-E-O model is simple but provides a powerful framework for the design of assessment activities and can accommodate simple or highly sophisticated assessment techniques. Inputs refer to those personal qualities that a student brings initially to the program. The environment refers to the student's actual experience or programming, and the outcomes refer to the end product or attribute that is trying to be obtained by the environment. In the research design process, the outcomes can be referred as the dependent variables or criterion. The model forms a triangle with direct relationships between inputs and the environment, the environment and outputs, as well as a relationship between inputs and outputs. What is critical about the relationship between inputs and outputs is that it can take into account personal attributes that might change regardless of the environment. The model also allows for the adjustment of differences in the inputs in order to get less biased estimates of the effects of different environments on outputs.

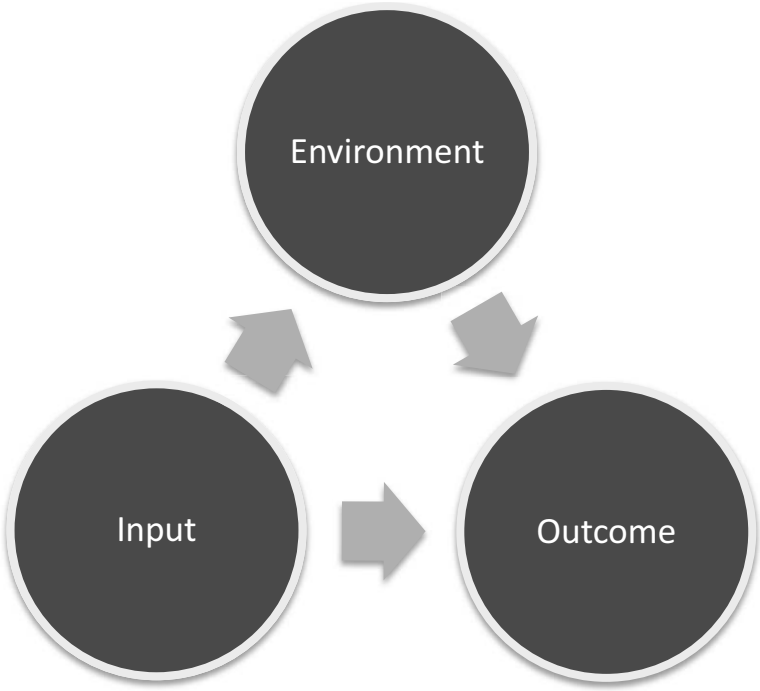

Figure 1: Astin's I-E-O model.

In the context of Astin's involvement theory and satisfaction of the educational environment, the I-E-O model can be applied. A student's level of involvement can be considered a key input. Institutional type, academic program climate and culture, faculty interaction, institutional programs, and services can be considered environmental factors. Student satisfaction can serve as the output. The model will take into account how the level of involvement directly effects satisfaction but will also take into account how the environment acts as a mediator to student satisfaction.

One limitation of the I-E-O model is that it does not distinguish between individual student experiences in the environment and the overall institutional environment (Kim, 2001). Given that each student is unique, there are certain institutional environmental factors that can cause students to have different experiences in the same environment. For example, a student with high social skills might have a better perception of a highly active, social campus than a

student with lower social skills. Shouping and Kuh (2003) outlined a causal model that takes into account individual student environmental perceptions and overall institutional characteristics.

Online and Paper Survey Administration

Online administration of surveys have become more popular due to the emergence of survey construction software, reduced data entry, and overall reduced costs of online surveys. Deutsken, Ruyter, & Wetzels (2006) found in a study of an office equipment manufacturer's clients that online surveys had a higher response rate than paper surveys, respondents to online surveys submitted lengthier answers to open-ended questions, and identical results were found between online and paper surveys. Carini, Hayek, Kuh, Kennedy, & Ouimet (2003) found that there were small distinctions (specifically more favorable results on online surveys) between online and paper submissions of the National Survey on Student Engagement. Epstein, Klinkenberg, Wile, & McKinley (2001) found that mean ratings of their physical attractiveness survey were equivalent between online and paper surveys.

Similar results have been observed with the KUMC Student Services Survey in the context of the length of open-ended comments. During the 2008 administration of the survey, 93 respondents provided general comments. 82 of the comments were from online surveys and 11 of the comments were from paper-based survey. The average length of the online comments was 302 characters whereas the average length of the paper-based comments was 185.

Basics of Factor Analysis

Factor analysis is a statistical technique that is used to discover which variables in a set form latent variables that are independent of one another. The latent variables are referred to as

factors and are intended to reflect a construct that is correlated among the variables (Thompson, 2004). Factor analysis is often referred to as data reduction or the scaling on the instrument.

The common types of research questions that are answered by factor analysis include how many factors are present in a given set of variables and what is the nature of the factors (how can the factors be interpreted)? The importance of the factors can be explained by how much variance in the data set can be explained by the factor(s). Factor analysis can be used to test theory about how well an obtained factor solution fits an expected solution. Once the nature of the factors is determined, factor scores can also be computed. Factor scores are variables of the underlying construct that are computed by various techniques that can be used in further statistical analysis in place of individual measured variables.

The advantage to conducting a factor analysis is that factor scores are often more reliable than scores based on a single variable. In addition, validity of the measurement can occur when predictions can be made based on low and high scores of the factors. In addition to the advantages, there are challenges to performing a factor analysis. First of all, there are no readily available criteria against which to test if a solution is appropriate. The decision is often made by researcher intuition and subjective interpretability of the factors. There are also an infinite number of rotations that are possible. The complexity of variables based on constructs is often challenging and is a reason why factor analysis is often inappropriately used to "bail out" poorly conceived research since if no other statistical procedure is applicable to the data, at least data can usually be factor analyzed. Tabachnick and Fidell (2007) have stated that the power of factor analysis ability to create apparent order out of chaos contributes to its somewhat tarnished reputation as a scientific tool.

There are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is used to look for relationships amongst the data in early phases of research and is used in this study. There is no set structure to the variables within a factor in mind as well as no set theory. The questions often asked by EFA is how many factors, which variables are part of a given factor, and how much variance in a data set is explained by those factors. CFA is used when a structure to the nature of the factors is pre-defined since a theory about the relationship of the variables is in place. CFA is often performed after a structure has been defined by a EFA with a new set of data to confirm that the structure determined by the EFA is the appropriate model. CFA is often referred to as a procedure to test a model and common CFA techniques are done through structural equation modeling. CFA is a method as to how the nature of a set of variables on an instrument can be validated.

For example, a researcher could develop a survey about on-campus housing attitudes and administer it to students at an institution. The researcher would then take the data and perform an EFA to determine the nature of the factors on the instrument. The researcher could then administer the survey at other institutions and use that data to perform a CFA to determine if the model defined by the EFA fits the data that was obtained at other institutions. If the data fits, the researcher can conclude that the survey adequately measures on-campus housing attitudes independent of the sample that was used to determine the nature of the factors.

There are two underlying assumptions regarding factor analysis (Green & Salkind, 2004). Measured variables are linearly related to the factors plus errors. This assumption is often violated especially when the factors have limited response scales that can be found in dichotomously scored items (e.g. right-wrong or true-false). Violation of this assumption can lead to the identification of spurious factors and caution should be taken when determining the

number of factors in the solution. If the assumption is met, the relationship amongst the variables is also determined to be linear and can be determined by evaluating scartterplots for pairs of items. The other assumption is that the chi-squared test for the maximum likelihood solution when it is used assumes that the measured variables are multivariately normally distributed and is often violated[3].

There are several steps to an EFA that are performed (Tabachnick & Fidell, 2007). There is a logical procession, but based on the results of the one of the steps, previous steps might be repeated with modifications based on what was learned. The first step is to select and measure the variables. In behavioral science research, this is often done by the administration of a survey or other methods of data collection. The next step is to develop a correlation matrix among the observed variables that is used to determine the number of factors that are present. There are several criteria that are used to determine the number of factors and are explained in more detail below. The final determination of the number of factors is often made by a culmination of information that is presented by the criteria.

Eigenvalues represent the amount of information represented by a factor. When performing a factor analysis, there is always the same amount of eigenvalues to the number of variables. The eigenvalue divided by the number of variables tells you how much of the total variance is explained by the factor. Guttman (1954) developed the eigenvalue greater than 1.0 rule for determining the number of factors. His rationale was that if a factor only had one variable with a correlation of 1.0 to the factor and the other variables had no correlation to the variable at all, the eigenvalue would be at least 1.0. Due to sampling error, judgment needs to

[3] According to Fabrigar, Wgener, MaCallum, and Strahan (1999), a maximum likelihood solution is most appropriate when the data is normally distributed.

be exercised using this criteria and is often why a factor with an eigenvalue of 1.05 might be rejected whereas a factor with an eigenvalue of 0.99 might be retained.

Cattell (1966) developed the visual inspection of a scree plot procedure to determine how many factors were present in a set of variables. A scree plot graphs eigenvalue magnitudes on the vertical axis and the eigenvalue numbers on the horizontal axis. Factor extraction should be stopped at the point where there is an "elbow" or leveling of the plot and is a visual test.

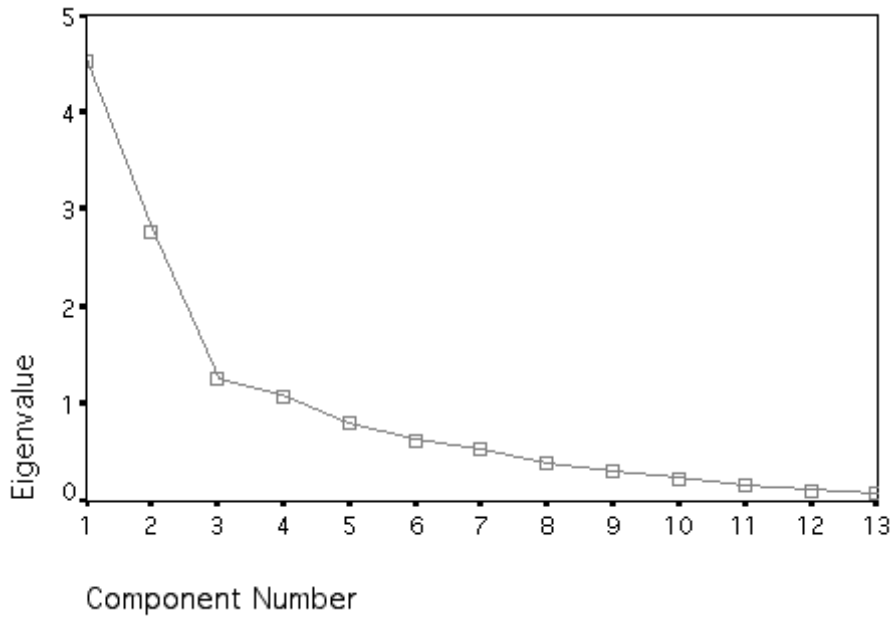

Figure 2: Scree plot for a data set with three factors and 13 variables.

Several researchers have developed objective regression-based methods to determine the number of factors from a scree plot in order to address the subjectivity of visual methods. Common methods found in the literature include the Cattell-Nelson-Gorsuch, the multiple regression, the *t*-value index, and the standard error scree All four procedures, with slight variations, detect the largest drop in the value of the two successive eigenvalues that correspond

to the point on the curve that separates the major from the trivial factors by means of multiple linear regression (Nasser, Benson, & Wisenbaker, 2002).

The next step is to extract the factors by computing factor pattern coefficients and determine which variables load into each factor by looking at the factor loadings. Factor extraction is necessary because in theory, all possible factors (equal to the number of variables) are computed at one time. Since many of the factors are trivial (don't account for much of the variance in the data at all), extraction conceptually allows the researcher to extract one variable at a time, in sequence, according to the number of factors the researcher determined were present. There are multiple statistical theories that can be used to do this which can affect interpretation but the most common default analysis is called principal components analysis (PCA) (Thompson, 2004). PCA is most common when the research purpose is data reduction or exploration and should not be used in causal modeling. PCA attempts to create linear combinations of the variables so that the maximum variance for each variable that is extracted and is often used to help determine the number of factors present. Another method, Principal Axis Factoring (PAF) attempts to create the least number of factors that account for common variances and is used when the research purpose is theory confirmation and causal modeling. PAF can be used when there is pre-conceived notion of the number of factors to extract and is not ideal for exploration purposes (Kline, 1998). The process of creating factors in a PAF is repeated as iterations until the communality estimates stabilize and converge. With an adequate sample size, the analysis should converge. As a rule of thumb, if the convergence takes more than 100 iterations, the data used is often considered problematic (Thompson, 2004). Another common method, maximum likelihood, is most often used once the number of factors has been

identified with a rotation method to interpret the results. This extraction method attempts to reproduce the variance in the population and not the sample.

As part of the extraction process, factor rotation is also completed. Factor rotation involves moving the factor axes measuring the locations of the measured variables in the factor space so that the nature of the underlying constructs become more obvious to the researcher (Thompson, 2004). Rotating the factors maximizes the high correlations between factors and variables and minimizes the low ones. The correlations between the factors and the variables are called factor loadings. Factor rotation is always needed when you have more than one factor and is also needed to better interpret what constructs the factors represent. By rotating the factors, the values of the variables are not changed; rather the scale that the variables are being measured with is being adjusted to enhance interpretation.

Rotation methods fall into two general categories: oblique and orthogonal. Oblique rotation methods allow the factors to correlate with each other whereas orthogonal methods do not allow the factors to correlate with each other. Theoretically, the number of factors that are extracted does not change. In addition, the common variance explained by the factors does not change. What does change based on rotation methods are the item loadings into the factors. Orthogonal rotation methods should be used when there is a preconceived notion that each variable is related to a single factor and does not load into multiple factors. The most common orthogonal rotation is the varimax rotation method developed by Kaiser (1958). Oblique rotations methods are going to lower factor loadings since items will be allowed to share variance between factors. Orthogonal rotation methods are going to have higher factor loadings since the methods are designed to fit items into individual factors. Depending on the nature of the research and the constructs will help determine the appropriate rotation method. In a study

where items are suspected to have loadings into multiple factors, oblique methods are optimal and will make results more interpretable. When oblique rotation is necessary, the promax rotation (Hendrickson & White, 1964) is usually the default choice. In the promax procedure, a number of rotations are conducted where the first rotation is an orthogonal (varimax) rotation and then pattern/structure coefficients are raised to an exponential pivot power[4]. Next, a procrustean rotation is invoked. A procrustean rotation finds the best fit of an actual matrix to some hypothetical target matrix. In the promax method, the hypothetical target is the pattern/structure coefficients derived from the varimax method each raised to the pivot power.

The next step in an EFA is to interpret the results. The variance in each variable that the collection of factors or items explains is referred to as the communalities. The proportion of variance accounted by each factor is often reported. If an oblique rotation method is chosen, the proportion of covariance is also reported. In a model where all the variance is accounted for, the factors will account for 100% of the variance.

The final step often done in a factor analysis is to create factor scores for the factors in order to create a composite numerical score for subsequent values[5]. A factor score can provide a subject's relative spacing or standing on the factor when compared to other subjects. The most common method when using the maximum likelihood extraction method is the regression method (Thompson, 2004) which produces standard z-scores. Most statistical software packages such as SPSS use this method to create factor scores. However, if there is interest in comparing factor scores across different factors, the regression method is not useful since they are based on standard scores. Thompson's method (1993) involves weighting and computing factor scores that can be used to compare factor score means across different factors.

[4] The pivot power can be changed to control the degree of correlation among the factors.
[5] Creating factor scores is not part of this study since the primary focus of performing a factor analysis was to determine the structure of the underlying variables.

Basics of Item Response Theory

Item response theory (IRT) links a test taker's trait (or ability) to responses on individual items. The name IRT is due to the focus on the theory of the item and not the test-level focus of classical test theory and is the preferred method for the development of modern tests like the Graduate Record Examination and Medical College Admission Test admissions tests as well as health professional licensure tests like the United States Medical Licensing Examination and National Council of State Boards of Nursing Licensure Examination. IRT is based on two basic postulates (Hambleton, Swaminathan, & Rogers, 1991):

- performance of an examinee on a test item can be predicted based on their ability; and
- the relationship between item performance and the set of the traits can be described by a monotonically increasing function called an item characteristic curve (ICC).

The ICC is a mathematical expression that connects an examinees probability of success on an item with the trait that is being measured. The ICC is a non-linear (logistic) regression line with item performance regressed on examinee ability (θ)[6]. The x-axis represents ability and the y-axis represents the probability of successfully answering the item correctly. Figure 3 is an ICC for a dichotomous scored item where the probability of success of answering the item correctly.

[6] In the context of the KUMC Student Services Survey, the theta level refers to the level of the psychological construct of student satisfaction that is being measured by the item and the respondent possesses and the probability on the y-axis refers to the probability of endorsing the item.

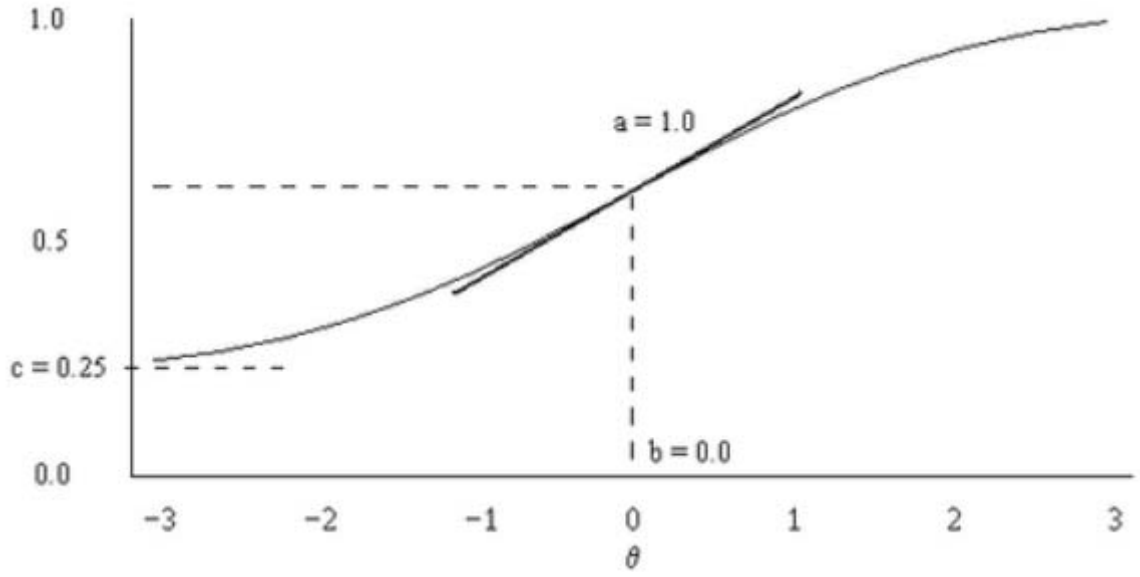

Figure 3: Example item characteristic curve.

The ICC can contain three parameters that are used to build the different models in increasing complexity due to the characteristics of the items. The nature and location of the curves is affected by different values of the parameters. The *b* parameter represents the item's difficulty and the location where θ has the maximum slope. The *a* parameter represents the item's discrimination, that is the degree to how well the item distinguishes between individuals in different regions of ability. For items, such as multiple choice items, the *c* parameter is used in attempt to account for the effects of guessing a correct response and corresponds to the y-intercept on the ICC. Taking the ICC in Figure 3, this item would be considered an item of medium difficulty since $b = 0.0$ which is near the center of the distribution (with theta measured on a standard scale). The item also discriminates very well between individuals of low and high ability since $a = 1.0$. An individual of low ability has approximately a 25% of guessing the answer correctly since the *c* parameter equals 0.25.

In IRT, there are a number of models to choose from depending on the nature of the data and three of the more popular models are the one-parameter logistic (1PL), two-parameter logistic (2PL), and three-parameter logistic (3PL) models. The 1PL model only takes the *b* parameter into account, the 2PL model takes the *a* and *b* parameters into account, and the 3PL model takes all three parameters into account. For example, the probability of a correct response to an item *i* in the 3PL model is:

$$p_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta - b_i)}} \quad \textbf{(1)}$$

The models above are helpful for comparing across items. For example, Figure 4 represents two items that have the same discrimination but different item difficulties (*b* parameter values):

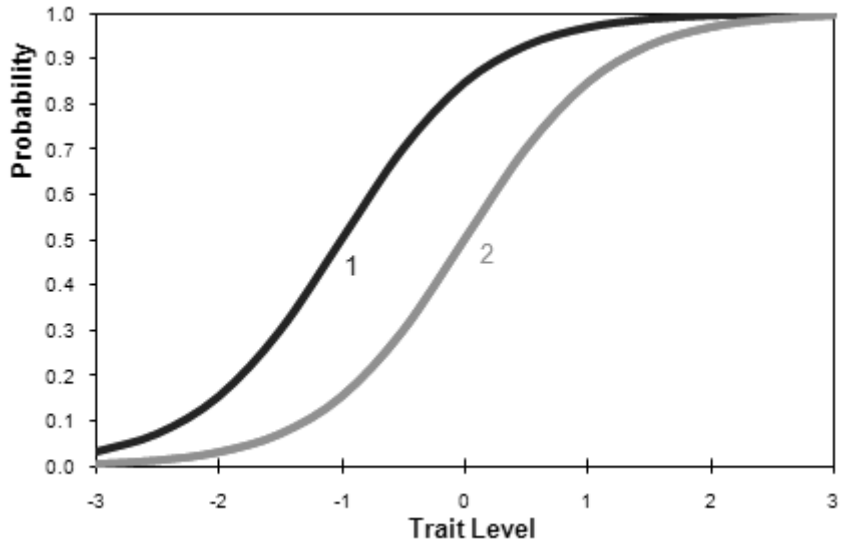

Figure 4: Item characteristic curves for two example items with different difficulty values.

Item 2 would be considered a more difficult item since the respondent needs to be at a higher ability level than item 1 in order to have a greater chance of answering the question correctly.

Figure 5 represents two example items where the item difficulty is the same but the items have different discrimination values. If the intention is to better discriminate between individuals of lower and higher ability, item 2 is a preferable item to use based on the steepness of the slope.

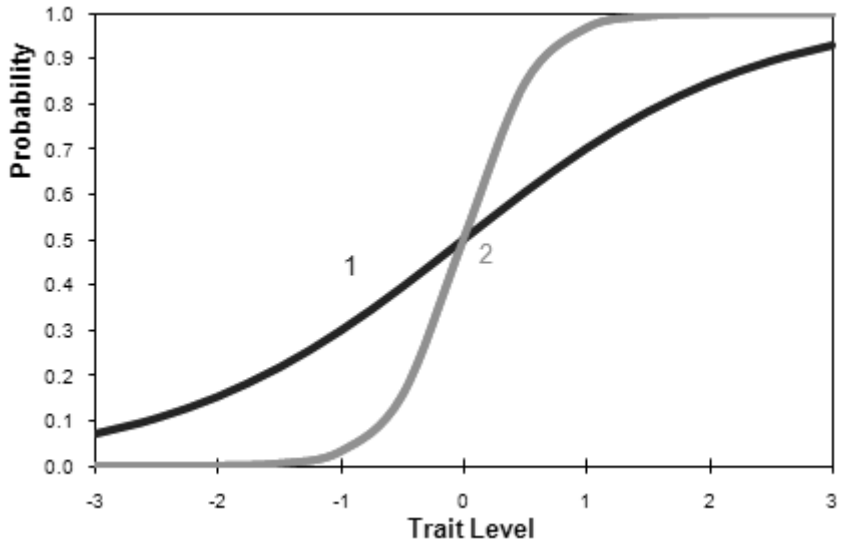

Figure 5: Item characteristic curves for two example items with different discrimination values.

In the context of a survey such as the KUMC Student Services Survey, the probability of "endorsing" the item is used instead of the probability of answering the item "correctly." The item is measuring the latent trait of "student satisfaction" instead of "ability" level. In the models described above, there are no allowances for partially endorsing an item but there are polytomous IRT models that allow for partial endorsement that is often present in Likert scales. Some researchers believe that ordinal scales like Likert scales are more informative and reliable than dichotomous scoring (Embretson & Reise, 2000). Due to this, the graded response model is used in this study.

There are several important assumptions made when using item response theory. First, it was assumed that each factor measured a single latent trait (unideminsionality). Second, it was assumed that within a group of respondents with the same value of the latent trait, the distributions of the item responses are independent of each other (Lord & Novick, 1968).

Basics of the Graded Response Model

When item responses are characterized as ordered categorical responses, the graded response model (GRM) is appropriate (Samejima, 1969, 1972, 1997). The GRM is often referred to as a difference model because the probability of an examinee responding in a particular category is found by subtraction (Thissen & Steinberg, 1986). In the GRM, each item is described by one item discrimination *a* parameter and multiple item difficulty *b* parameters similar to the 2PL model. The number of item difficulty parameters is determined by J = K -1 where K equals the number of response categories and J equals the number of category thresholds. For example, in a survey item with five response options (e.g., strongly disagree, disagree, neutral, agree, strongly agree, each scored one to five, respectively), the number of thresholds, or boundaries between categories, is four. The GRM does not include a guessing *c* parameter since guessing is assumed not to be present in endorsement surveys. In the example above, four item response functions (IRFs) are created which can be interpreted similarly to ICCs. For a polytomous item without a guessing parameter, the GRM formula is:

$$P_{jx}^{*}(\theta) = \frac{e^{Da_j(\theta - b_{jx})}}{1 + e^{Da_j(\theta - b_{jx})}} \quad \textbf{(2)}$$

Where j = 1,2,…n, n = number of items, x = number of response thresholds, and D is a constant equal to 1.702.

For a given item, the *a* parameter is constant for the IRFs and each curve will have a different *b* parameter. The curves do not cross due to the constant slope. The curves that are derived from this formula represent the conditional probabilities for which an examinee with a given ability level will score at those category levels or higher and are referred to as the cumulative category characteristic functions.

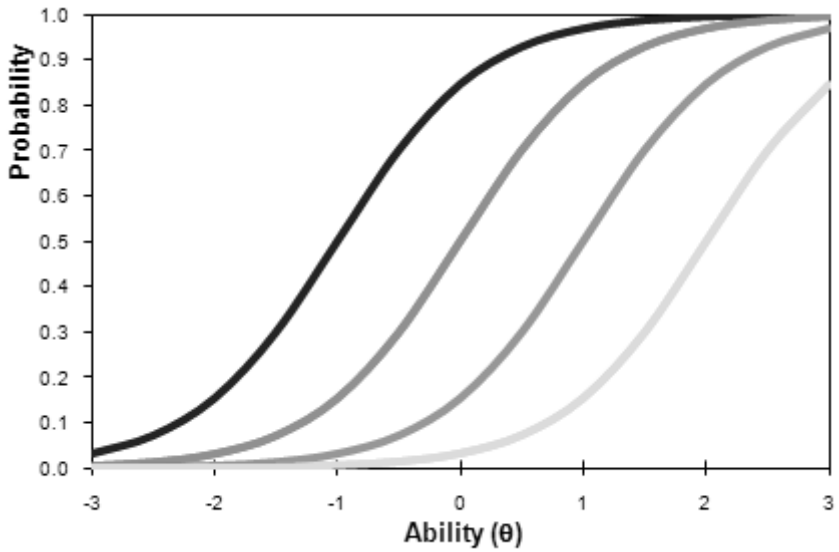

Figure 6: Cumulative category characteristic functions for an example with 5 response options.

When comparing different items, the *a* values can vary and represent the relationship between the latent trait (θ) and the item response where larger *a* values represents a stronger response for the item in relationship to θ.

Each curve in Figure 6 represents a category threshold. For example, in a survey with the following response options:

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

The first threshold is between categories 1 and 2, the second threshold is between categories 2 and 3, the third threshold is between categories 3 and 4, and the fourth threshold is between category 4 and 5. The resulting b_{jx} parameters for each threshold can be interpreted as the amount of trait level necessary to respond at or above the threshold with a 50% probability. By definition, the probability of responding above the highest category is zero and the probability of responding in the lowest category or higher is 1.0.

The cumulative category characteristic functions are used to calculate score category response functions. The score category response function is the probability with which an examinee with a given ability will endorse a particular category and is derived by subtracting the difference in probabilities between the cumulative category characteristic functions of adjacent categories.

$$P_{jx}(\theta) = P^*_{jx}(\theta) - P^*_{j(x+1)}(\theta) \quad \textbf{(3)}$$

As noted previously, the cumulative probability of scoring the lowest category is 1.0 so the probability of scoring in the second lowest category is calculated (for illustration purposes the scale above is used where category 1 corresponds to strongly disagree and category 2 corresponds to disagree):

$$P_{j1}(\theta) = P^*_{j0}(\theta) - P^*_{j1}(\theta) = 1.0 - P^*_{j1}(\theta) \quad \textbf{(4)}$$

the score category response functions for categories 2, 3 (agree), and 4 (strongly agree) are calculated in a similar fashion:

$$P_{j2}(\theta) = P^*_{j1}(\theta) - P^*_{j2}(\theta) \quad \textbf{(5)}$$

$$P_{j3}(\theta) = P^*_{j2}(\theta) - P^*_{j3}(\theta) \quad \textbf{(6)}$$

$$P_{j4}(\theta) = P^*_{j3}(\theta) - P^*_{j4}(\theta) \quad \textbf{(7)}$$

and since the probability of scoring above the highest category is zero, the probability of scoring a 5 (strongly agree) for a given ability is calculated by the following formula:

$$P_{j5}(\theta) = P_{j4}^{*}(\theta) - P_{j5}^{*}(\theta) = P_{j4}^{*}(\theta) - 0 \quad \textbf{(8)}$$

Also of note is that the cumulative probabilities for each of the score category response functions at a given ability level equal 1.0.

$$\sum_{x=0}^{m_j} P_{jx}(\theta) = 1.0 \quad \textbf{(9)}$$

Plotting the curves of the score category response functions can provide the researcher with a good deal of information about an item. The item parameters dictate the shape and location of the curves. The discrimination parameter dictates the steepness and how the narrow curve will be and is an indicator of how well the response categories differentiate between individuals. The difficulty parameters determine the location of the curves. Curves located at the lower ability levels indicate that lower levels of ability are needed to endorse a particular category.

Figure 7 (Rennie, 1982) is an ICC for an item with 4 response categories (strongly agree, disagree, agree, and strongly agree).

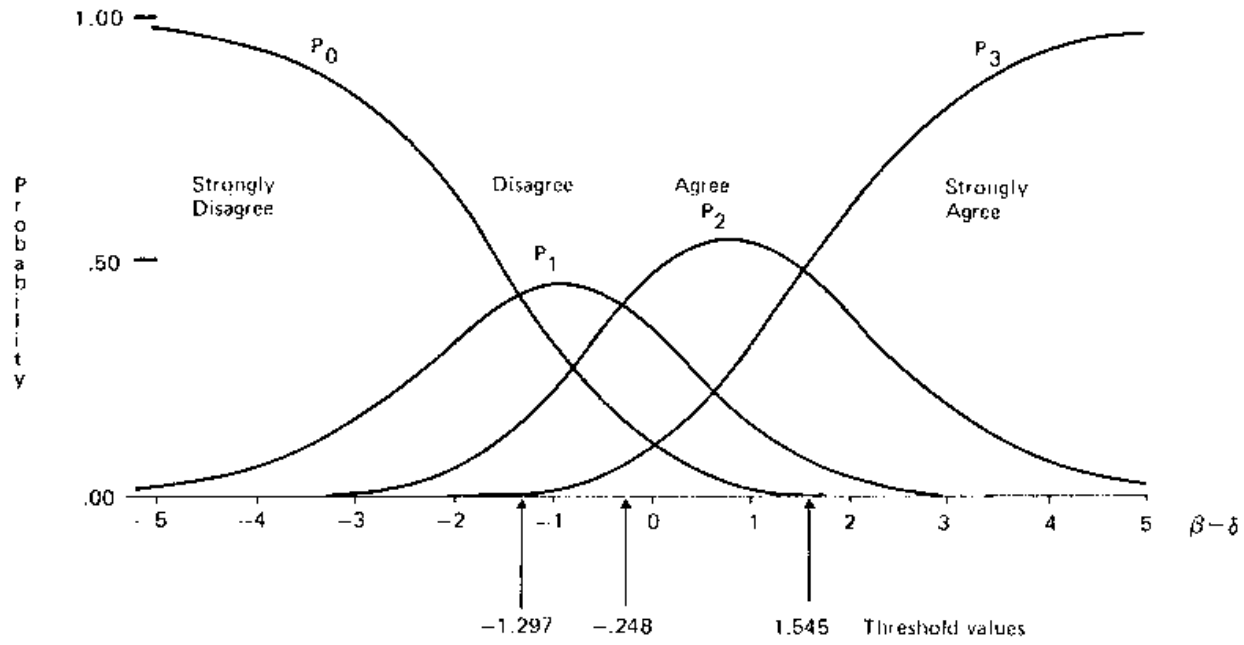

Figure 7: Graded response model item characteristic curves.

An individual with a θ value of -3.0 has approximately an 85% probability of endorsing the strongly disagree category. This is determined by locating the θ value of -3.0 on the x-axis and determining which curve has the highest value. This individual would have approximately a 15% probability of endorsing the disagree, and essentially no probability of endorsing categories agree or strongly agree. The other curves are interpreted in a similar fashion. In the context of the KUMC Student Services Survey, a favorable result of this study would be to find similar curves. Based on relevant research, a student who does not endorse an item related to student satisfaction would be expected to have a higher probability of selecting one of the non-endorsement categories than one of the endorsing categories.

The probability information can be combined into one function for the expected score at a given ability level and is referred to as an item characteristic function and is analogous to an item characteristic curve for a dichotomous item:

$$E(X_j|\theta) = \sum_{x=1}^{m_j} xP_{jx}(\theta) \quad \textbf{(10)}$$

The expected score can function is then derived into an information function that is able to describe how much information can be determined from the item at a given theta level. Generally, more information can is obtained from an item when the *a*-parameter is high and the b-parameter is close to theta. The information provided by the curve is analogous to the concept of the reliability found in classical test theory. For example, the item information curve in Figure 8 can be interpreted as a very reliable item where the respondent does not need to possess a high level of ability in order to endorse the item and is more discriminating between levels.

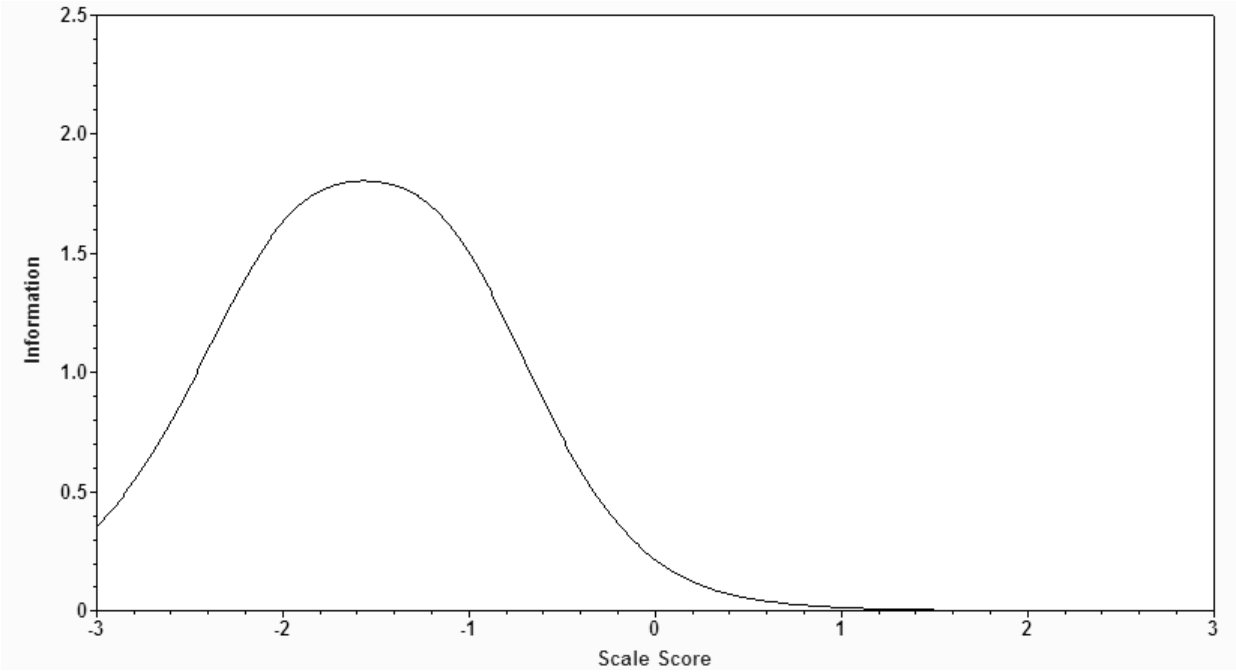

Figure 8: Example item information curve for a high discriminating item.

A broader curve as depicted in Figure 9 indicates an item that is less discriminating and does not provide as much information when compared to Figure 8.

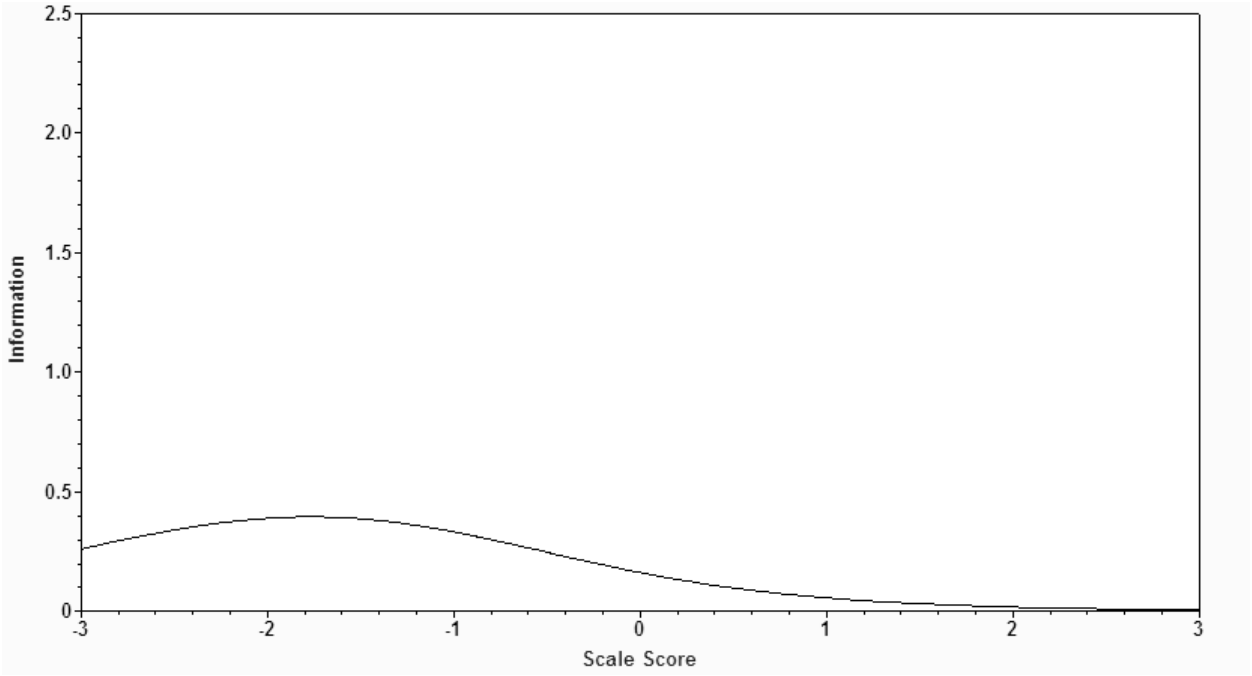

Figure 9: Example item information curve for a low discriminating item.

The sum of information functions for a set of items and plotting those values with the respective theta values will produce the test information curve. Plotting this function along with the standard can provide information about at which score levels the test is a more reliable measure of the construct of interest. For example, the item in Figure 10 is a more reliable measure of ability in the lower regions than the higher regions.

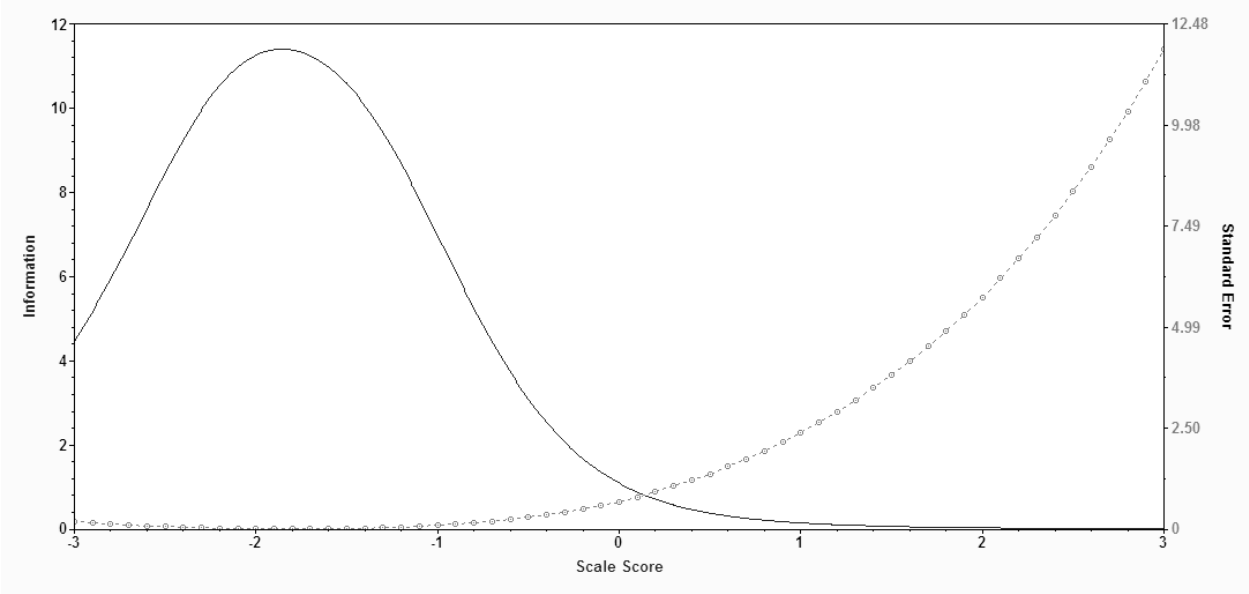

Figure 10: Example test information characteristic function and standard error.

There are number of ways to determine if the items fit a GRM model. The goodness of fit of each item can be determined by the likelihood-ratio chi-squared statistic and the sum of these chi-squared statistics provides the goodness of fit for the whole test (Samejima, 1997). Large chi-squared values are concluded that the assumptions of the model have been violated. When large chi-squared values are clustered around a relatively small number of items, exclusion of these items will often improve the fit.

Establishing the Reliability of an Instrument

A very important property of any measurement is that it is reliable. Reliability in the terms of psychometrics is similar to how it is used in everyday conversation. A reliable measurement is consistent or stable. In terms of testing[7], a test is reliable when an individual receives similar scores on different administrations of the test (test, retest reliability), with

[7] For the purpose of discussing reliability, testing is referenced but is analogous to the measurement that is being conducted by the KUMC Student Services Survey.

different versions of a test (alternate-form), or with different subsets of questions from a test (internal consistency) (Friedenberg, 1995).

The reliability of a test can be determined by using two sets of scores and calculating the Pearson correlation coefficient from those test scores. The two sets of scores can be from a test-retest scenario or alternate forms testing. In a test-retest scenario, the reliability is an index of the stability of scores. Enduring characteristics that are suspected not to change over time are appropriate for test-retest reliability. The coefficient is representative of the proportions of variance. For example, a coefficient of 0.95 indicates that 95% of the differences in scores are due to individual differences in ability and 5% of the differences in scores are due to measurement error.

Alternate forms reliability is an index of similarity of test items across different forms. The most apparent application of this reliability is ensuring that different forms of a test are equivalent and is critical in large scale testing. The coefficient that is computed is an index of equivalence between forms and can be interpreted similarly to the example above except that differences are due to error between the terms.

Some instruments, like the KUMC Student Services Survey, only have only one form or are administered once yet it is still important to establish the reliability of the test. The internal consistency approach to a test is estimating if the items are tapping into the same domain of knowledge. One procedure is the split-half procedure where the test items are split into half and treated like two separate tests where a correlation coefficient is calculated. The split-half procedure works under the notion that all of the items are measuring the same domain and can be used to predict performance on other tests that measure the same domain.

The split-half procedure is not always practical since splitting the items into half can be quite arbitrary. A more sophisticated measure towards establishing internal consistency is Cronbach's coefficient alpha (1951). Cronbach's alpha (Formula 11) is computed from the individual item variances and estimates the average split-half coefficient for all possible divisions of the test:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K}\sigma_{Y_i}^2}{\sigma_X^2}\right) \quad \textbf{(11)}$$

where K=number of items, σ_X^2=variance of the observed total test scores, and $\sigma_{Y_i}^2$=variance for each item.

Establishing the Validity of Test Scores

In the context of test scores, validity is the integrative evaluation of the degree to which the empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on scores (Messick, 1995). Validity is the evaluation of whether the test is actually measuring the intended construct whereas reliability is more a measurement of the stability or precision of the scores. Reliability is a prerequisite in determining the validity of test scores.

The unitary concept of validity (Messick, 1995) refers to the integration of the three main types of validity that were mentioned above. In order to make any conclusions about the validity of a measurement, all three have to be considered in a conceptual framework. Only by considering all of the aspects of validity can you make a conclusion about the measurement's ability to represent the construct.

The unitary concept of validity introduces the concept of interpretation versus use as the functional outcome of the testing. Interpretation refers to describing an examinee's level on the construct. Use refers to making decision of the basis of scores. Messick's also introduces the

concept of the evidential versus consequential basis of validity. Evidential refers to the ability to relate empirical evidence to the score meaning and consequential refers to the outcomes are associate with the purpose of test scores.

These four concepts are the basis of Messick's matrix for progressive validity (Figure 11). The test interpretation and evidential basis help determine the construct validity of a measurement. Once construct validity is established, determining how the test is going to be used helps establish the relevance and utility of test scores. Evaluating the test interpretation and consequential basis of test scores further helps establish construct validity and the value implications of the test score. Putting all of these considerations together helps determine the ultimate consideration of the social consequences of the measurement. By building on top of construct validity in a progressive fashion like this, the unitary concept of validity can be established in test construction.

	Test Interpretation	**Test Use**
Evidential Basis	Construct Validity (CV)	CV+Relevance/Utility (R/U)
Consequential Basis	CV+Value Implications (VI)	CV+R/U+VI+Social Consequences

Figure 11: Messick's Progressive Matrix for Validity.

Establishing Validity in the Research Process

Cook and Campbell (1979) outlined aspects of validity that a researcher must take into account during a study. Essentially, establishing the validity of a study is ensuring that whatever conclusions are being made about the research question has some meaning to the outside world.

Careful consideration of all four aspects must be taken into account in the research process to help ensure that results and conclusions have some form of utility.

The first aspect of validity outlined by Cook and Campbell is statistical validity. Statistical validity refers to the establishment of a statistical relationship between the independent and dependent variables of a study. When statistically significant results are found in a study, the establishment of statistical control validity helps substantiate the quantitative results. Threats to statistical control validity include inadequate sample sizes. Increasing the sample size of a study has a direct effect on the power of a test which increases the certainty of a conclusion. "Fishing" for significant results in data is also a threat and should be avoided. Going into a research study, the researcher should have an idea about the data analysis that will be employed. If non-significant results are found, looking for differences in the data for the sake of reporting significant results has a detrimental effect on the validity of the study.

Violations of the assumptions of a statistical test are also a threat to statistical validity since the significance values of the test are partially based on these assumptions. Assumption violations result in the increased possibility of making a Type I error. Researchers should always check assumption tests and make the necessary adjustments when violations are found are use alternative tests that relax or correct for the standards of the assumption. Unreliable measurements are also a major threat to the statistical validity since the validity of a study cannot be establishment with inconsistent measurement of variables. Reliable measurement techniques and appropriate analysis of the reliability of the measures should always be employed in a study.

Internal validity refers to the establishment of causality between the independent and dependent variables. Internal validity helps answer the question about if the treatment, manipulation, or the technique in question on an independent variable helped initiate a change in

the dependent variable. Significant results on a statistical test do not infer causality and internal validity is established by determining that change occurred in the dependent variables based on something to do with the independent variables.

The biggest threats to internal validity, especially in social sciences, revolve around the subjects of a study. Since the subjects are what are of interest, the confounding of any of their behavioral aspects during a study threatens the internal validity. Improper selection of subjects, mortality of subjects, imitation behaviors during testing, and maturation during longitudinal testing are all threats to establishing causality. Proper subject selection and testing protocols that help minimize these threats and ensure that the behavior is being adequately measured can help establish the internal validity of a study.

Content validity refers to the ability to draw inferences from examinee test scores to a larger domain of items that are similar to the test but not being measured. Often referred as face validity, this can be conceptualized as the ability to determine if the items on the test are adequate enough to make a conclusion that a score on the test is an indication of an individual's ability, mastery, etc. The evidence that is generally collected is by using qualified experts to determine if the constructs are being measured based on certain specifications. This type of evidence is not necessarily empirically based but essential in determining if the constructs are being measured. It also can help determine if bias is present and if test performance and item data need to be considered in test construction.

Arguably, the most important type of validity is construct validity. Cronbach and Mehl (1955) defined construct validity as the interpretation of a measure that cannot be operationally defined. An instrument is considered to have construct validity when it valid measures appropriate characteristics of a test taker (Friedenberg, 1995). This is very important especially

when there is no acceptable criterion that can be directly measured. Construct validity emphasizes the overlap that exists in validity and many argue that construct validity and overall validity is the same thing.

Convergent and divergent validity are important in determining construct validity. Convergent validity is the confirmation that intended measure by different procedures and divergent validity is established when different methods yield different expected results on the same construct. Additionally, consideration of the traits and methods that are being measured are crucial to construct validity. This helps determine if the instrument is measuring the trait and if the trait is even a functional unity of the construct.

Construct validity data is collected by formulating a hypothesis about the construct, developing your instrument, gathering empirical data which allows you to test the relationships in your hypotheses, and determining if the data is consistent with the hypotheses. Since construct validity is such an essential piece of validity, it is vital in determining if the construct can be measured which allows for interpretation and use of the instrument.

The biggest threats to construct validity are construct underrepresentation and overrepresentation. In terms of underrepresentation, are there other confounding constructs that are also being measured that are not being taken into account is a common question that the researcher must answer. With respect to overrepresentation, is the construct of interest overstating its effect on the variables is the question that needs to be answered. Construct validity is very difficult to establish and the researcher can employ techniques to protect against this threat. Using multiple measures and different established tests that measure the construct are ways to help determine if construct validity is being obtained in addition to using longer testing

procedures since longer tests are more reliable and will provide multiple sources of data about the constructs of interest.

External validity refers to the establishment of representation to the population of interest. Ultimately, the goal of a research study that employs sampling is to be able to draw conclusions to the general population. The biggest threat to external validity is that the sample does not represent the population. The principles of random sampling and the central limit theorem ensure that conclusions about a population can be made from random sampling. By employing adequate sample sizes and random sampling of an adequate, representative population, the researcher can counteract the threats to the external validity of a study.

The aspects of validity outlined by Cook and Campbell and their threats need to be addressed in any study. In essence, the establishment of validity is the reduction of doubt about the conclusions of the study. Many of the ways to protect against these threats complement each other. Adequate sample sizes, random sampling of subjects, proper testing techniques, and reliable measurements are the fundamental principles to establishing validity.

Summary

Astin (1985) stated, "True institutional excellence is in the institution's ability to effect a student's intellectual, personal, and scholarly development" (pp. 60-61). It has been well established in student affairs literature that the non-curricular contexts of the educational experience are more important in encouraging student engagement and learning opportunities than curricular contexts of the educational experience (Pascarella & Terenzini, 1991). The adequate assessment of these factors is vital to understand how the campus environments influence student behavior and academic performance (Kuh, 1991).

This research addresses if student satisfaction can be scaled under an item response theory model with reliability and validity. The KUMC Student Services Survey has been long believed to measure student satisfaction based on the face validity of the instrument. Even though the items do not specifically ask a student to rate their level of satisfaction with their academic and non-academic experiences, based on the nature of the items, a respondent who agrees to the item is assumed to be satisfied with their experience. If an item response theory model can be specified, the theta estimates produced by the analysis will represent the student's level of satisfaction with the underlying variable. The reliability of the instrument will be established through classical test theory statistics of internal consistency, as well as from the test information functions derived from IRT.

Construct validity of an item response model towards measuring student satisfaction with the KUMC Student Services Survey will be established based on established literature and the application of Astin's I-E-O model. As mentioned previously, research has found negligible differences between other online and paper- based versions of surveys including the National Survey of Student Engagement that is administered to college students. If the specified item response theory model for the KUMC Student Services Survey is appropriately scaled, the theta estimates between online and paper-based versions of the survey will be statistically non-significant. One of the fundamental findings of student development literature over the years has been that involved students are more satisfied with their collegiate experiences. If the specified item response theory model for the KUMC Student Services Survey is appropriately scaled, the theta estimates between involved and not-involved students will be statistically significant with involved students indicating higher levels of satisfaction. Based on previous results of the KUMC Student Services Survey and a working knowledge of the KUMC environment, School

of Medicine students have historically been less satisfied with their educational experience than School of Allied Health and School of Nursing students. If the specified item response theory model for the KUMC Student Services Survey is appropriately scaled, the satisfaction estimates between School of Medicine students and School of Allied Health/Nursing students will be statistically significant with School of Allied Health/Nursing students indicating higher levels of satisfaction.

It is also plausible that a student's academic school can have a mediating effect where it with student involvement in regards to student satisfaction. Figure 12 outlines a conceptual framework using Astin's I-E-O model. If the specified item response theory model for the KUMC Student Services Survey is appropriately scaled, a mediating effect between student involvement and academic school might be present when looking at levels of student satisfaction.

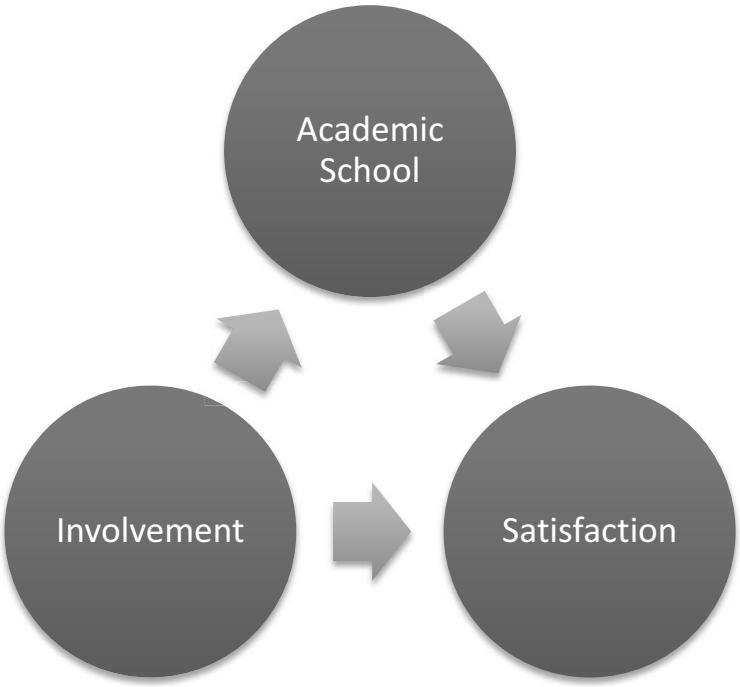

Figure 12: Conceptual framework for student involvement, academic school, and satisfaction.

CHAPTER THREE – METHODS

The purpose of this study was to determine if a student satisfaction survey can be validly scaled with an item response theory (IRT) graded response model (GRM). Evidence for the validity of the scaling is provided by an investigation into the relationships between scores and survey mode, student academic school, and student involvement. This chapter describes the participants, instrument, scaling procedures, statistical analyses, and limitations of the study.

Participants

The participants of the study were students at an academic medical center. An academic medical center is generally defined as an institution that contains at least a medical school and is owned by or affiliated with a clinical facility where faculty instruct physicians in training. Activities that are often conducted at an academic medical center include didactic teaching, clinical education, patient care, and research. Most academic medical centers contain additional schools for the health professions (nursing, pharmacy, dentistry, allied health professions) and other clinical entities such as faculty group practices and community health centers (Blumenthal, Campbell, & Weissman, 1997).

Specifically, the participants were undergraduate, graduate, and medical students at the University of Kansas Medical Center (KUMC) who completed the Student Services Survey (SSS) in 2006, 2007, 2008, and 2009. In 2006, 2007, and 2008, all KUMC students were mailed a copy of the survey requesting that they complete the survey as well as multiple email messages with a link where they could complete the survey online. Due to budget constraints in 2009, students were only sent email messages requesting that they complete the survey online but were also given the opportunity to pick up a paper survey in case they had privacy concerns. No

KUMC students submitted a paper survey in 2009. Overall, 1,515 surveys were used in this study.

Table 1
Demographic Information for Participating KUMC Students

	KUMC Students (n = 1,515)
Survey Mode	
Paper	35.0%
Online	65.0%
Involvement in a Student Organization	
Yes	52.0%
No	48.0%
Academic School	
School of Allied Health	18.0%
School of Medicine	58.8%
School of Nursing	23.2%
Mean Age (S.D.)	24.39 (11.72)
Year of Study	
1st	34.5%
2nd	24.2%
3rd	18.4%
4th	16.4%
5th	3.5%
6th	1.9%
7th or more	1.0%
Sex	
Male	24.7%
Female	75.3%
Ethnicity	
American Indian/Alaskan Native	0.7%
Hispanic	3.0%
White, non-Hispanic	83.9%
Asian/Pacific Islander	7.4%
Black, non-Hispanic	3.7%
Other	1.4%
U.S. Citizenship	
Yes	93.4%
No	6.6%

KUMC is the academic medical center campus of the University of Kansas. KUMC has two locations: the main location is in Kansas City, Kansas and a second location in Wichita, Kansas. KUMC is comprised of three academic schools (School of Allied Health, School of Nursing, and School of Medicine) and a Graduate Studies program. KUMC operates alongside the University of Kansas Hospital which provides clinical experiences for its students and residency programs for medical school graduates. The primary mission areas of KUMC are instruction, research, and service (About the University of Kansas Medical Center). KUMC enrolls approximately 1200 graduate, 800 medical students, and 250 undergraduate students each academic year. Approximately 700 residents and fellows participate in graduate medical education experiences at one or both campuses (Official Headcount Enrollment by School and Program: Fall 2005 to Fall 2009)

All survey responses on the KUMC Student Services Survey are collected on a confidential basis and analyzed as group data. Since the survey responses are non-identifiable and previously gathered for non-research purposes, informed consent is not required to conduct the study. Approval from the University of Kansas Human Subjects Committee has been obtained.

KUMC Student Services Survey

The KUMC Student Services Survey was designed by the KUMC Division of Student Services to assess student experiences as a KUMC student (the 2009 version is in Appendix A). The KUMC Student Services Survey has been in existence since 1993. From 1993-2001, the survey was conducted biannually. Beginning with 2003, the survey has been conducted annually and students were given the choice of completing it online instead of the paper survey. A complete online administration was first conducted in 2009 due to budgetary constraints.

Students were given the option to submit a paper application in 2009 but no students elected this option. Since 2003, the survey has contained the same 30 items related to student satisfaction on a five point Likert agreement scale. Beginning with the 2005 administration of the survey, a five point Likert importance scale was added for each of the 30 items. A special section of interest is added to the survey each year. The special interest topics have included smoking on campus, campus safety, student wellness, and student rights and responsibilities. The importance scale, special sections, and various demographic items were not part of this study.

In 2004, a section related to information resources was added to the survey to assess information technology needs of KUMC students. Due to comments from respondents about the survey length, the survey was split into two separate surveys in 2007. The second survey, the KUMC Student Technology Survey, only contains the items related to student information technology.

The respondents of the survey are also asked to provide demographic information:

- academic school;
- level of degree;
- year of study;
- gender;
- ethnic category;
- age;
- citizenship;
- involvement in KUMC student organizations; and
- location of the majority of classes.

Procedure

Any survey submission with no data was removed from the data set. Missing responses to the academic school and student involvement items were removed from the descriptive statistics. Responses with missing individual item, academic school and involvement responses were not completely removed from the internal consistency and factor analysis procedures since the SPSS default is to exclude these cases. Missing data was also not excluded from the IRT procedures since PARSCALE will skip over these responses. Missing demographic data and item response was cause for removing the response from the ANOVA procedures. The following rubric was used to assign a score to each item's response:

- Strongly disagree = 1;
- Disagree = 2;
- Neutral = 3;
- Agree = 4; and
- Strongly Agree = 5.

Descriptive statistics for the 30 items were produced and reported.

Exploratory Factor Analysis Procedure

An exploratory factor analysis using the program SPSS was used to identify the number of factors of the KUMC Student Services Survey and which factor each item loaded into according to the procedures outlined by Thompson (2004) and Green and Salkind (2004). An exploratory factor analysis was conducted as opposed to a confirmatory factor analysis since the purpose was to simply identify the structure of the survey. The number of factors were determined by a combination of several procedures which included the eigenvalue greater than 1.0 rule (Guttman, 1954), visual inspection of the scree plot (Cattell, 1966), the objective

standard error scree method (Zoski, 1996), and the interpretability of the factor solution. The standard error scree method was chosen over other objective methods because it has been found to be more accurate than other objective measures in determining the number of factors (Nasser, 2002). The method of principal components analysis was used to extract the factors since there was no pre-conceived notion as to the nature of the items and the purpose of this process was for data reduction and exploration. The factors were rotated with the orthogonal varimax rotation method. Respondents with missing data in any of the items were excluded from the factor analysis. For the purposes of this study, the factor structure of the underlying variables was assumed to be the same for all subgroups that were studied.

Classical Test Theory Reliability Procedure

The internal consistency estimate, coefficient alpha (Cronbach, 1951) was computed for each of the underlying variables that were identified by the factor analysis procedure. Coefficient alpha was also computed for the community subscale for each year the survey was administered in order to justify using all of the survey responses for an adequate sample size.

Graded Response Model Procedure

The computer program PARSCALE (Muraki & Bock, 2003) was used to attempt to scale each of the constructs identified by the exploratory factor analysis procedure. This was accomplished using the graded response model calibration option in PARSCALE. The syntax for each of the scaling procedures is located in Appendix B. Because the data was polytomous in form, the estimation process in PARSCALE for the parameters is more complex. The discrimination and difficulty parameters are estimated one at a time and then the iteration process is repeated for each item. After obtaining stable item parameters estimates, the estimation process for the category threshold parameters is repeated until their values becomes stable at a

specified level of precision (Muraki, 1990). PARSCALE estimates the parameters by marginal maximum likelihood assuming a latent distribution with a mean of zero and a standard deviation of one. Once the model converged for the construct using PARSCALE, each item's item characteristic curve and item information curve was analyzed to assess the quality of the item and likelihood ratio chi-squared statistic was computed. For each underlying variable that was identified by the exploratory factor analysis, the likelihood ratio chi-squared statistic and theta estimates for each respondent was computed in addition to the Gaussian (i.e., normal) fit to the latent trait distribution. Chi-squared statistics were analyzed at the 0.001 level of significance.

Score Transformations

In order to make the raw scores for the score more interpretable for the community satisfaction subscale, a conversion table was created. The new community satisfaction score is determined by summing the responses for the eight community items (strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, and strongly agree = 5) and finding the corresponding value in Table 12. A community satisfaction score of 50 corresponds to a theta estimate of zero and the scale has a standard deviation of 4. The community satisfaction score was derived by the following procedure. For each item, the categorical response probabilities were calculated for a given range of theta. These response probabilities were multiplied by the corresponding value (strongly disagree = 1, disagree = 2, neutral = 3, agree = 4, and strongly agree = 5) and summed for the item. The summed response probabilities for all eight items were summed up and plotted with the new community satisfaction score for a given raw score that was created by linearly transforming the theta value where a theta value (Figure 33). The actual community satisfaction score that matches the integer level raw scores was determined by a Microsoft Excel formula that is located in Appendix D.

Statistical Analyses

A 2x2x3 analysis of variance (ANOVA) was conducted to determine the effect of student involvement, mode of survey, student academic school, and interaction effects dependent variable of community satisfaction. A 2x3 ANOVA was also conducted to determine the effect of student involvement, student academic school, and interaction effects on the dependent variable of community satisfaction for the 2008-2009 survey administrations since those administrations were completely online. All procedures were tests at the 0.05 level of significance.

Design

For the purposes of the scaling procedures, all survey responses were analyzed as a single group. For the purposes of the scaling validation, there were three independent variables which are outlined in Appendix B and one dependent variables which was the community satisfaction scores.

Summary of Procedures and Statistical Analyses

The primary purpose of this study was to determine if an academic medical center student satisfaction survey can be scaled using an item response theory graded response model with reliability and validity. While the current analysis of the survey administration and staff valuable information about the campus environment and the needs of students for programming purposes, the survey has never undergone reliability and validity analysis to determine if it is adequately measuring its intended constructs. Descriptive statistics and a classical test theory reliability statistic were computed, exploratory factor analysis was used to identify the underlying variables, a graded response model was used to scale the underlying variables, and analysis of variance was used to establish the construct validity of the model.

Limitations

There are four main limitations to this study. First, an individual student could submit multiple versions of the survey during a given administration. In other words, a student could submit multiple online and paper-based versions of the survey. So, it is likely that some of the individuals in the sample of 1515 are actually the same people measured on more than one occasion. It was impossible to indentify these individuals because each time a person takes the survey it is confidential. It is assumed that these instances are relatively rare and therefore would not adversely affect the results of these analyses. Second, students are also free to choose how they submit the survey (online versus paper) and are not randomly assigned to groups. Thus, any mode of administration effects present may not be simply due to paper versus online, but may be related to the choice of mode. Third, the item used to identify involved versus non-involved students only asks the respondent if they are involved in a student organization as a dichotomous item (yes or no), and does not address their level of involvement. Organizations can range from an intramural team to a student governing organization. Students can also have different levels of involvement in an organization ranging from if they just attend meetings to being an active leader in multiple student organizations Due to this, the involvement variable as it is measured by the survey may overlook differences in satisfaction related to student involvement. Finally, several of the items have limited response patterns since most respondents respond in the higher category levels which will produce some skewness in the estimated parameters.

CHAPTER FOUR – RESULTS

The purpose of this study was to determine if constructs related to student satisfaction can be measured with reliability and validity at an academic medical center using an existing student services survey instrument. This chapter is organized in sections. Descriptive statistics about the items are first presented, followed by results of the exploratory factor analysis procedure, internal consistency estimates, graded response model scaling procedures, the raw score transformation procedures, and the results of the validation procedures.

Item Descriptive Statistics

Table 2 summarizes the frequency distributions of all the items. In order for the graded response model to estimate parameters, at least one response is needed in each response category. The majority of responses are in the upper response categories for the first 26 items and the lower response categories for the last four items. Table 3 summarizes the descriptive statistics for the 30 items. Most of the items have a mean score in the low to mid 4.0 range. The last four items have lower scores because they are considered reverse scored items when compared to the first 26 items. The standard deviations for the items are in the 0.7 to 1.1 ranges.

Table 2
Frequency Distributions for the 30 Items on the KUMC Student Services Survey (n = 1,515)

Item	Response Category					
	1	2	3	4	5	No Response
The environment at KUMC is friendly and supportive of students	10	29	61	480	899	36
Campus security/safety is adequate	41	142	135	499	644	54
Students have a voice in decision-making at KUMC	16	61	263	591	516	68
KUMC has a positive image in the Kansas City area	5	24	102	532	801	51
There is racial/ethnic harmony at KUMC	6	36	183	493	737	57
Students at KUMC are free from sexual harassment	11	19	150	391	880	64
Male and female students are treated equally at KUMC	11	45	142	384	880	53
There is a feeling of community at KUMC	15	70	164	590	631	45
Students have a sense of pride at KUMC	15	39	129	567	727	38
It is easy to approach faculty at KUMC	12	46	68	493	856	40
It is easy to approach staff at KUMC	12	44	117	539	749	54
KUMC is committed to resolving diversity-related problems	12	27	344	426	600	106
I am comfortable raising diversity related problems to faculty, administration, and staff	8	7	247	325	824	104
Students are treated the same regardless of race	16	32	172	348	874	73
Students are treated the same regardless of color	14	32	174	343	879	73
Students are treated the same regardless of creed	7	25	181	351	872	79
Students are treated the same regardless of religion	8	31	184	357	859	76
Students are treated the same regardless of national origin	16	24	179	353	867	76
Students are treated the same regardless of ancestry	11	14	184	357	874	75
Students are treated the same regardless of gender	10	46	152	368	868	71
Students are treated the same regardless of age	13	37	158	379	857	71
Students are treated the same regardless of sexual orientation	18	20	204	351	839	83
Students are treated the same regardless of disability	8	21	202	349	843	92
Students are treated the same regardless of veteran status	8	7	247	325	824	104
It has been beneficial for me to be a member of a student organization while at KUMC	12	30	57	373	1017	26
The Student Wellness Program provides services and programs to meet my needs	37	85	99	329	928	37
Drug use is a problem for students in my program at KUMC*	655	327	374	41	11	107
Alcohol use is a problem for students in my program at KUMC*	280	221	248	143	36	587
Drug use is a problem for me*	1363	32	45	5	5	65
Alcohol use is a problem for me*	1304	79	45	10	5	72

Notes. Items marked with a * are reversed scored items.
Response Categories: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree

Table 3
Descriptive Statistics for the 30 Items on the KUMC Student Services Survey (n = 1,515)

Item	*Mean*	*SD*
The environment at KUMC is friendly and supportive of students	4.51	0.73
Campus security/safety is adequate	4.07	1.08
Students have a voice in decision-making at KUMC	4.06	0.90
KUMC has a positive image in the Kansas City area	4.43	0.72
There is racial/ethnic harmony at KUMC	4.32	0.82
Students at KUMC are free from sexual harassment	4.45	0.79
Male and female students are treated equally at KUMC	4.42	0.84
There is a feeling of community at KUMC	4.19	0.89
Students have a sense of pride at KUMC	4.32	0.82
It is easy to approach faculty at KUMC	4.45	0.79
It is easy to approach staff at KUMC	4.35	0.88
KUMC is committed to resolving diversity-related problems	4.12	0.90
I am comfortable raising diversity related problems to faculty, administration, and staff	3.94	0.99
Students are treated the same regardless of race	4.41	0.86
Students are treated the same regardless of color	4.42	0.86
Students are treated the same regardless of creed	4.43	0.81
Students are treated the same regardless of religion	4.41	0.83
Students are treated the same regardless of national origin	4.41	0.85
Students are treated the same regardless of ancestry	4.44	0.81
Students are treated the same regardless of gender	4.41	0.85
Students are treated the same regardless of age	4.41	0.85
Students are treated the same regardless of sexual orientation	4.38	0.87
Students are treated the same regardless of disability	4.40	0.83
Students are treated the same regardless of veteran status	4.38	0.83
It has been beneficial for me to be a member of a student organization while at KUMC	3.78	0.97
The Student Wellness Program provides services and programs to meet my needs	3.65	0.95
Drug use is a problem for students in my program at KUMC*	1.88	0.95
Alcohol use is a problem for students in my program at KUMC*	2.39	1.18
Drug use is a problem for me*	1.11	0.47
Alcohol use is a problem for me*	1.15	0.52

Notes. Items marked with a * are reversed scored items.
Response Categories: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree

Exploratory Factor Analysis Results

The dimensionality of the 30 items from the KUMC Student Services Survey was analyzed using principal components factor analysis. Four criteria were used to determine the number of factors to extract and rotate: eigenvalue greater than 1.0 rule, visual inspection of the scree plot, the standard error score method, and the interpretability of the factor solution.

Based on the eigenvalues greater than 1.0 rule, scree test criteria, and the standard error scree the hypothesis that the KUMC Student Services Survey is comprised of multiple latent constructs related to student satisfaction was supported. The eigenvalue greater than 1.0 (Table 4) and the visual inspection of the scree plot (Figure 13) indicated at least six factors. The results of the standard error scree method indicated that there were at least 11 factors present (Table 5). Based on all of these findings, it was determined to start the review of the interpretability of at least a six factor solution.

Table 4
Initial Eigenvalues of the KUMC Student Services Survey

Component	Eigenvalue
1	13.64
2	2.37
3	2.11
4	1.26
5	1.10
6	1.02
7	0.89
8	0.77
9	0.77
10	0.68
11	0.63
12	0.58
13	0.51
14	0.44
15	0.41
16	0.39
17	0.32
18	0.30
19	0.27
20	0.26
21	0.25
22	0.23
23	0.18
24	0.16
25	0.13
26	0.11
27	0.10
28	0.07
29	0.06
30	0.02

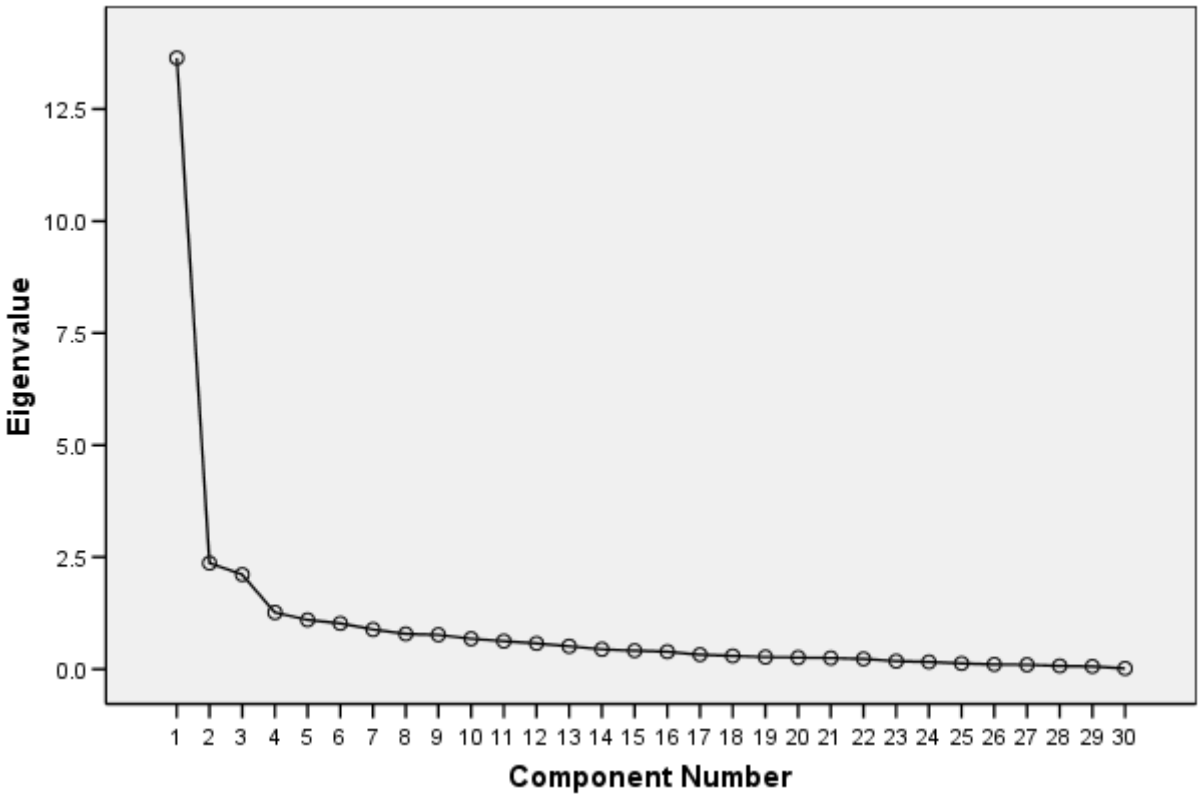

Figure 13: Initial Scree Plot of the KUMC Student Services Survey.

Table 5
Standard Error Scree Scores of the KUMC Student Services Survey

Order	Eigenvalue	SE	Factor*
1	13.64	2.177	Nontrivial
2	2.37	0.300	Nontrivial
3	2.11	0.222	Nontrivial
4	1.26	0.102	Nontrivial
5	1.10	0.083	Nontrivial
6	1.02	0.072	Nontrivial
7	0.89	0.058	Nontrivial
8	0.77	0.051	Nontrivial
9	0.77	0.048	Nontrivial
10	0.68	0.040	Nontrivial
11	0.63	0.034	Nontrivial
12	0.58	0.029	Trivial
13	0.51	0.021	Trivial
14	0.44	0.016	Trivial
15	0.41	0.015	Trivial
16	0.39	0.014	Trivial
17	0.32	0.010	Trivial
18	0.30	0.011	Trivial
19	0.27	0.011	Trivial
20	0.26	0.011	Trivial
21	0.25	0.011	Trivial
22	0.23	0.011	Trivial
23	0.18	0.011	Trivial
24	0.16	0.011	Trivial
25	0.13	0.008	Trivial
26	0.11	0.009	Trivial
27	0.10	0.010	Trivial
28	0.07	0.011	Trivial
29	0.06	0.010	Trivial
30	0.02	0.013	Trivial

*critical value = .033

The interpretability of a six, five, four, and three factor solution was reviewed based on the face validity of the factor loading of the items. Based on this analysis, it was determined that a three factor solution was the appropriate model. The three factors were named community, diversity, and substance abuse. Once the number of factors was determined, the factor analysis was repeated with a varimax rotation. The varimax rotated solution is shown in Table 6. The bolded factor loadings indicate the item's factor loading. Diversity accounted for 34.4% of the item variance, community accounted for 18.5% of the item variance, and substance abuse accounted for 7.5% of the item variance for a cumulative amount of 60.4% of the item variance being accounted for by a three factor solution. Five of the items loaded on both the diversity and community factors. Two of the items did not load on any of the factors.

Table 6
Initial Rotated Factor Loadings for the KUMC Student Services Survey (n = 1,515)

Item	Diversity	Community	Substance Abuse
The environment at KUMC is friendly and supportive of students	.25	**.70**	-.05
Campus security/safety is adequate	.28	**.50**	-.10
Students have a voice in decision-making at KUMC	.27	**.70**	-.00
KUMC has a positive image in the Kansas City area	.28	**.57**	-.06
There is racial/ethnic harmony at KUMC	**.59**	**.40**	-.09
Students at KUMC are free from sexual harassment	**.52**	**.44**	-.08
Male and female students are treated equally at KUMC	**.57**	**.43**	-.06
There is a feeling of community at KUMC	.23	**.76**	-.08
Students have a sense of pride at KUMC	.20	**.73**	-.14
It is easy to approach faculty at KUMC	.27	**.72**	-.06
It is easy to approach staff at KUMC	.26	**.68**	-.02
KUMC is committed to resolving diversity-related problems	**.56**	**.43**	-.02
I am comfortable raising diversity related problems to faculty, administration, and staff	**.53**	**.45**	-.02
Students are treated the same regardless of race	**.92**	.25	-.03
Students are treated the same regardless of color	**.92**	.23	-.05
Students are treated the same regardless of creed	**.90**	.24	-.06
Students are treated the same regardless of religion	**.86**	.24	-.06
Students are treated the same regardless of national origin	**.89**	.25	-.07
Students are treated the same regardless of ancestry	**.90**	.24	-.06
Students are treated the same regardless of gender	**.80**	.30	-.09
Students are treated the same regardless of age	**.79**	.33	-.05
Students are treated the same regardless of sexual orientation	**.86**	.25	-.07
Students are treated the same regardless of disability	**.81**	.28	-.07
Students are treated the same regardless of veteran status	**.82**	.23	-.07
It has been beneficial for me to be a member of a student organization while at KUMC	.03	.35	.116
The Student Wellness Program provides services and programs to meet my needs	.19	.35	-.02
Drug use is a problem for students in my program at KUMC	-.15	-.05	**.71**
Alcohol use is a problem for students in my program at KUMC	-.13	-.05	**.66**
Drug use is a problem for me	.01	.04	**.77**
Alcohol use is a problem for me	.00	.01	**.78**

The five items that loaded across factors and the two items that did not load into any factor were removed and the factor analysis with a varimax rotation was repeated. This was considered the final model for the graded response model scaling (Table 7). The bolded factor loadings indicate the item's factor loading. The cumulative variance explained increased to 67.7%. Diversity accounted for 37.8% of the item variance, community accounted for 20.5% of the item variance, and substance abuse for 9.7% of the item variance.

Table 7
Final Factor Loadings for the KUMC Student Services Survey (n = 1,515)

Item	Diversity	Community	Substance Abuse
C1 The environment at KUMC is friendly and supportive of students	.24	**.72**	-.02
C2 Campus security/safety is adequate	.28	**.49**	.09
C3 Students have a voice in decision-making at KUMC	.26	**.69**	-.01
C4 KUMC has a positive image in the Kansas City area	.26	**.59**	-.04
C5 There is a feeling of community at KUMC	.21	**.78**	-.06
C6 Students have a sense of pride at KUMC	.18	**.76**	-.11
C7 It is easy to approach faculty at KUMC	.27	**.74**	-.03
C8 It is easy to approach staff at KUMC	.24	**.71**	-.00
D1 Students are treated the same regardless of race	**.91**	.27	-.03
D2 Students are treated the same regardless of color	**.92**	.25	-.06
D3 Students are treated the same regardless of creed	**.90**	.27	-.06
D4 Students are treated the same regardless of religion	**.86**	.27	-.06
D5 Students are treated the same regardless of national origin	**.90**	.27	-.06
D6 Students are treated the same regardless of ancestry	**.90**	.27	-.07
D7 Students are treated the same regardless of gender	**.81**	.31	-.08
D8 Students are treated the same regardless of age	**.80**	.34	-.05
D9 Students are treated the same regardless of sexual orientation	**.85**	.26	-.07
D10 Students are treated the same regardless of disability	**.82**	.28	-.07
D11 Students are treated the same regardless of veteran status	**.82**	.25	-.07
S1 Drug use is a problem for students in my program at KUMC	-.16	-.06	**.70**
S2 Alcohol use is a problem for students in my program at KUMC	-.11	-.09	**.65**
S3 Drug use is a problem for me	.02	.04	**.79**
S4 Alcohol use is a problem for me	.01	.01	**.79**

Note. A variable name has been added for each item that corresponds to its factor loading.

Internal Consistency

The internal consistency estimate of coefficient alpha was computed for each of the three factors. The value for coefficient alpha was 0.86 for community, 0.98 for diversity, and 0.68

for substance abuse indicating satisfactory reliability for (Helmstadter, 1964)[8]. Individual coefficient alpha statistics were also computed for each year of the survey for the community subscale. The values for coefficient alpha were comparable for each year (2006 = 0.82, 2007 = 0.84, 2008 = 0.84, and 2009 = 0.82). Based on this finding, it was determined that using the data from all four survey administrations in order to have an adequate sample size was not problematic even though the same student could have possibly submitted a survey multiple years.

Graded Response Model Scaling Results

The scaling results are organized by treating each underlying variable as a separate scaling exercise. The parameter estimates, assessment of fit, item characteristic curves, item information curves, and total information curves are presented for the models that were able to converge.

Community

Parameter Estimates

Parameter estimates were obtained for the eight community items following the graded response model. The model converged after 49 iterations. The estimated parameters are displayed in Table 8. The category response thresholds are depicted by c_i for all the items and are not to be confused with the *c* parameter that is part of the three-parameter item response theory logistic model.

[8] According to Helmstadter, reliable attitude scales typically have a median reliability of 0.79 and 0.98 is considered highly reliable.

Table 8

Discrimination (a), Difficulty (b), and Category Response Thresholds (c_i) Parameters for the Convergent Graded Response Model of the 8-Community Items

Item	*a*	*SE*	*b*	*SE*	*c*(1)	*c*(2)	*c*(3)	*c*(4)
C1 The environment at KUMC is friendly and supportive of students	1.415	0.054	-1.937	0.027	-1.240	-0.469	0.119	1.590
C2 Campus security/safety is adequate	0.697	0.019	-1.665	0.040	-1.838	-0.362	0.293	1.907
C3 Students have a voice in decision-making at KUMC	1.079	0.031	-1.496	0.028	-1.773	-0.717	0.512	1.979
C4 KUMC has a positive image in the Kansas City area	0.974	0.029	-2.338	0.034	-1.888	-0.675	0.419	2.145
C5 There is a feeling of community at KUMC	1.466	0.047	-1.436	0.025	-1.431	-0.467	0.270	1.628
C6 Students have a sense of pride at KUMC	1.488	0.046	-1.605	0.025	-1.264	-0.547	0.211	1.599
C7 It is easy to approach faculty at KUMC	1.648	0.060	-1.689	0.026	-1.187	-0.349	0.107	1.429
C8 It is easy to approach staff at KUMC	1.479	0.479	-1.652	0.026	-1.318	-0.487	0.221	1.584

Assessment of Fit

As shown in Table 9, among the fit indices for the 8 community items, four of the items fit and the total test did not fit.

Table 9
Likelihood Ratio Chi-Square Statistics for Assessing Fit in Community Items

Item	χ^2	*df*	*p*
C1 The environment at KUMC is friendly and supportive of students	20.85	15	0.141*
C2 Campus security/safety is adequate	232.58	34	0.000
C3 Students have a voice in decision-making at KUMC	150.30	27	0.000
C4 KUMC has a positive image in the Kansas City area	51.04	19	0.000
C5 There is a feeling of community at KUMC	75.55	24	0.000
C6 Students have a sense of pride at KUMC	33.88	19	0.019*
C7 It is easy to approach faculty at KUMC	37.13	17	0.003*
C8 It is easy to approach staff at KUMC	42.00	19	0.002*
Total	643.33	174	0.000

Figure 14[9] presents the distribution of ability estimates that were obtained. The area under the normal curve equals the total area of the histogram. The satisfaction level estimates appear to have a normal distribution with some exceptions especially the higher ability estimates.

[9] The X-axis is labeled as ability but actually represents satisfaction level.

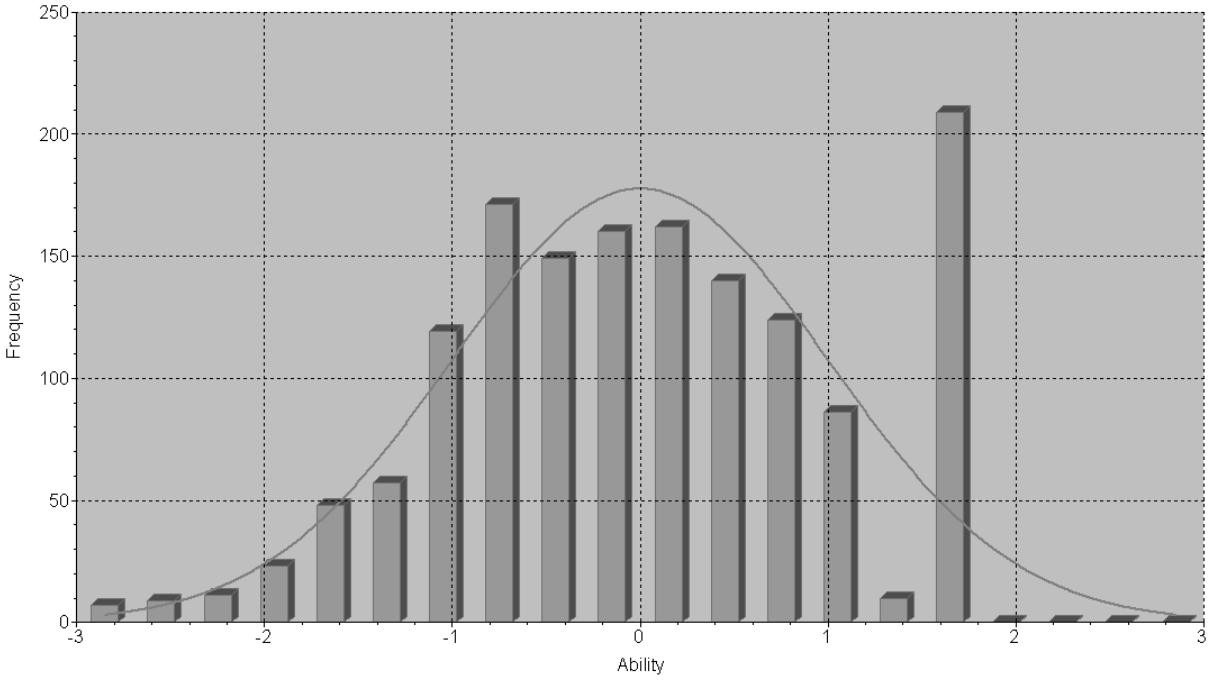

Figure 14: Gaussin fit to satisfaction scores for the eight community items.

Item Characteristic Curves

The following section individually discusses the eight community item characteristic curves in detail. The items are presented in order of easiest to most difficult to endorse.

C4 KUMC has a positive image in the Kansas City Area (Figure 15) is considered the easiest item to endorse meaning that the individual needs to have a low level of satisfaction in order respond favorably to the item. The item adequately discriminates between the five response categories.

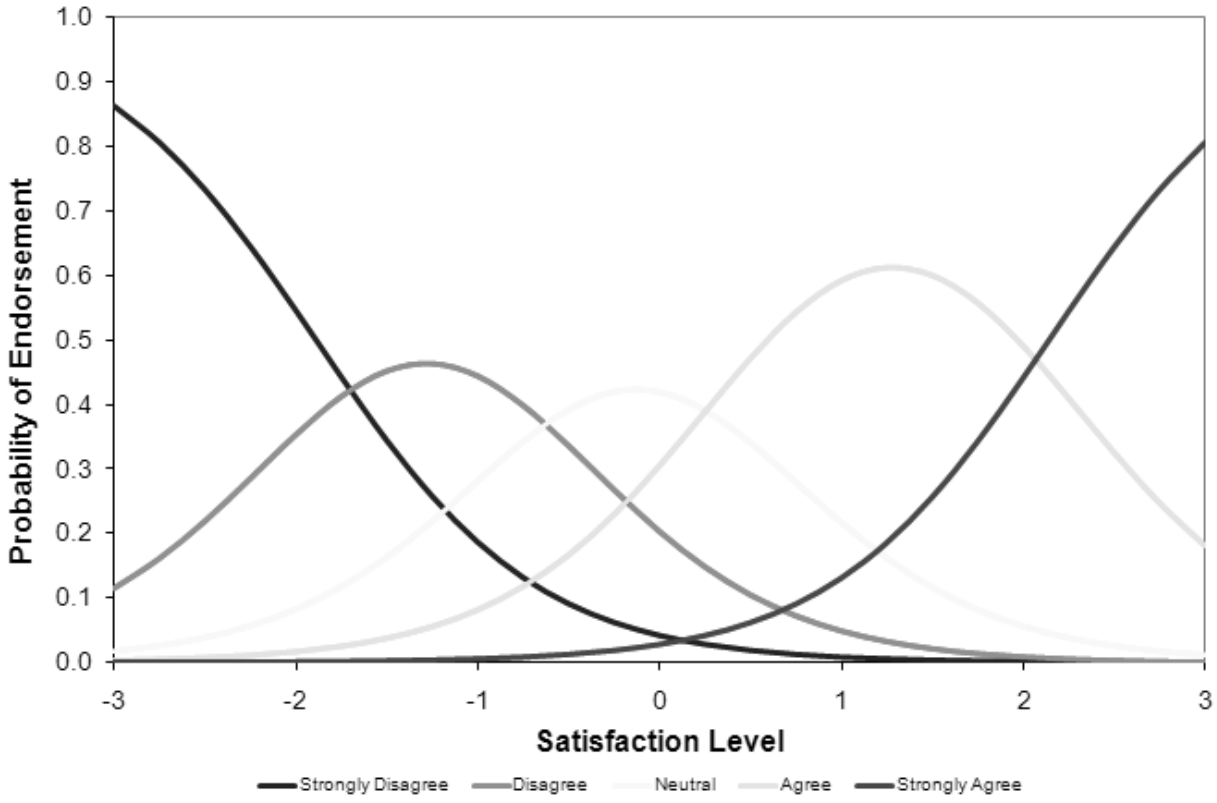

Figure 15: C4 KUMC has a positive image in the Kansas City area item characteristic curve.

C1 The environment at KUMC is friendly and supportive of students (Figure 16) is considered a more difficult item to endorse than C4 and is evidenced by the shifting of the curves to the right of the satisfaction level axis. The item is able to discriminate between strongly disagree, disagree, agree, and strongly agree, but the neutral category tends to be underutilized.

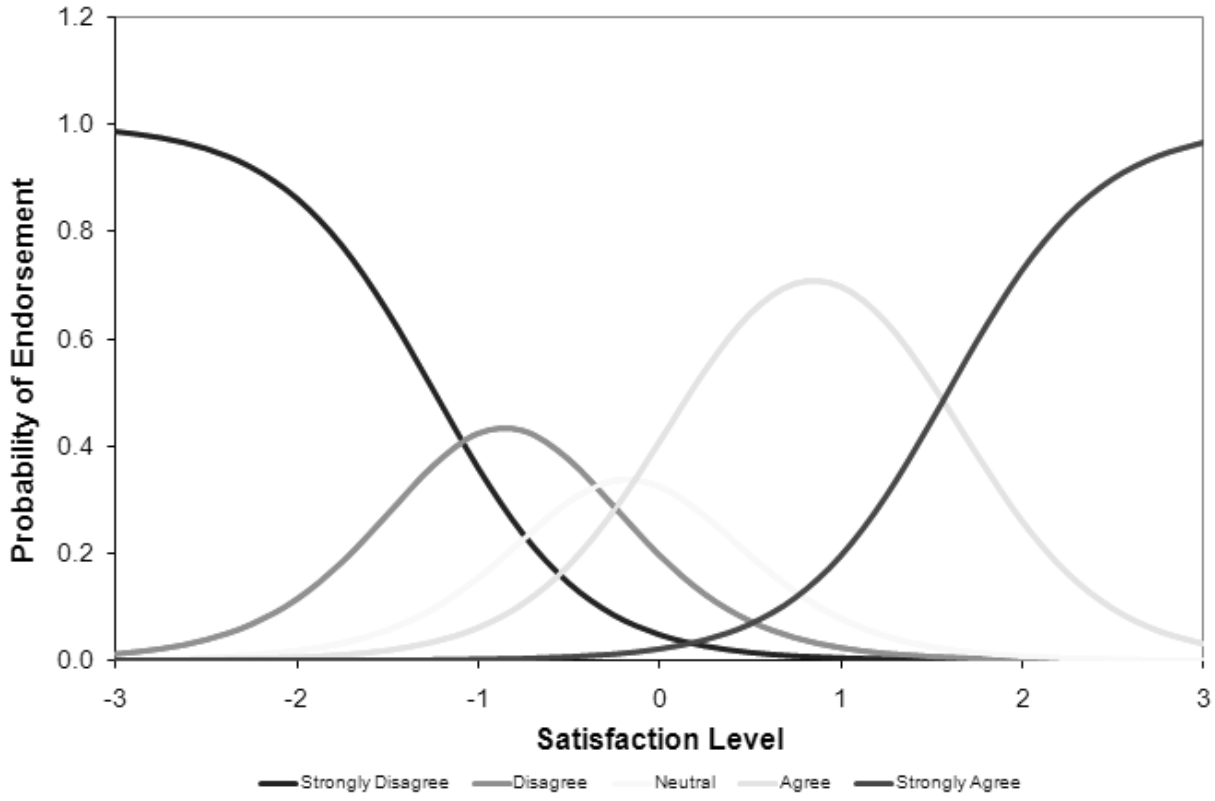

Figure 16: C1 The environment at KUMC is friendly and supportive of students item characteristic curve.

C7 It is easy to approach faculty at KUMC (Figure 17) is the third easiest item to endorse. The item is able to discriminate between strongly disagree, disagree, agree, and strongly agree, but the neutral category is underutilized.

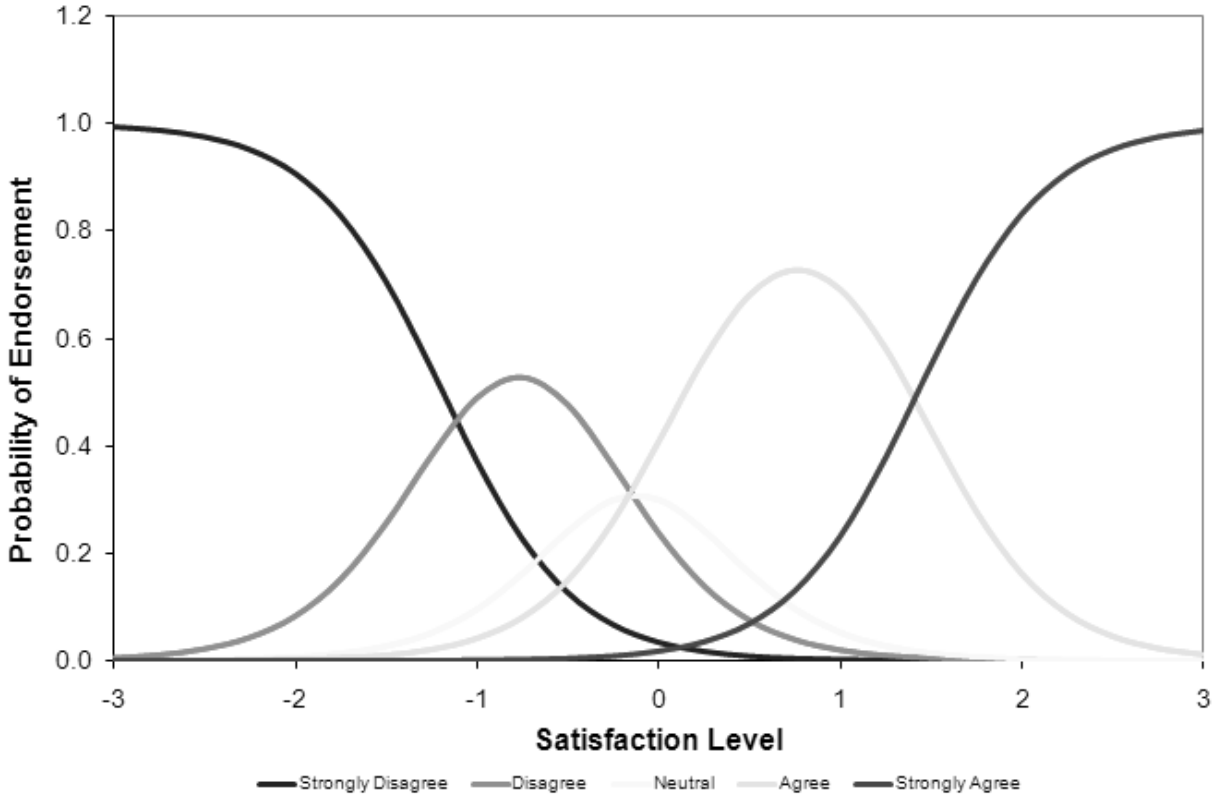

Figure 17: C7 It is easy to approach faculty at KUMC item characteristic curve.

C2 Campus safety/security is adequate (Figure 18) is the fourth easiest item to endorse. The item is able to discriminate between strongly disagree, disagree, agree, and strongly agree, but the neutral category is underutilized. When compared to the other seven items, C2 also has broader curves for the response categories suggesting that is not as discriminating item between response categories than the other seven community items.

Figure 18: C2 Campus security/safety is adequate item characteristic curve.

C8 It is easy to approach staff (Figure 19) is the fourth hardest item to endorse. The item is able to discriminate between all five categories with some slight underutilization of the neutral category.

Figure 19: C8 It is easy to approach staff at KUMC item characteristic curve.

C6 Students have a sense of pride in KUMC (Figure 20) is the third hardest item to endorse. The item is able to discriminate between all five categories except for underutilization of the neutral category.

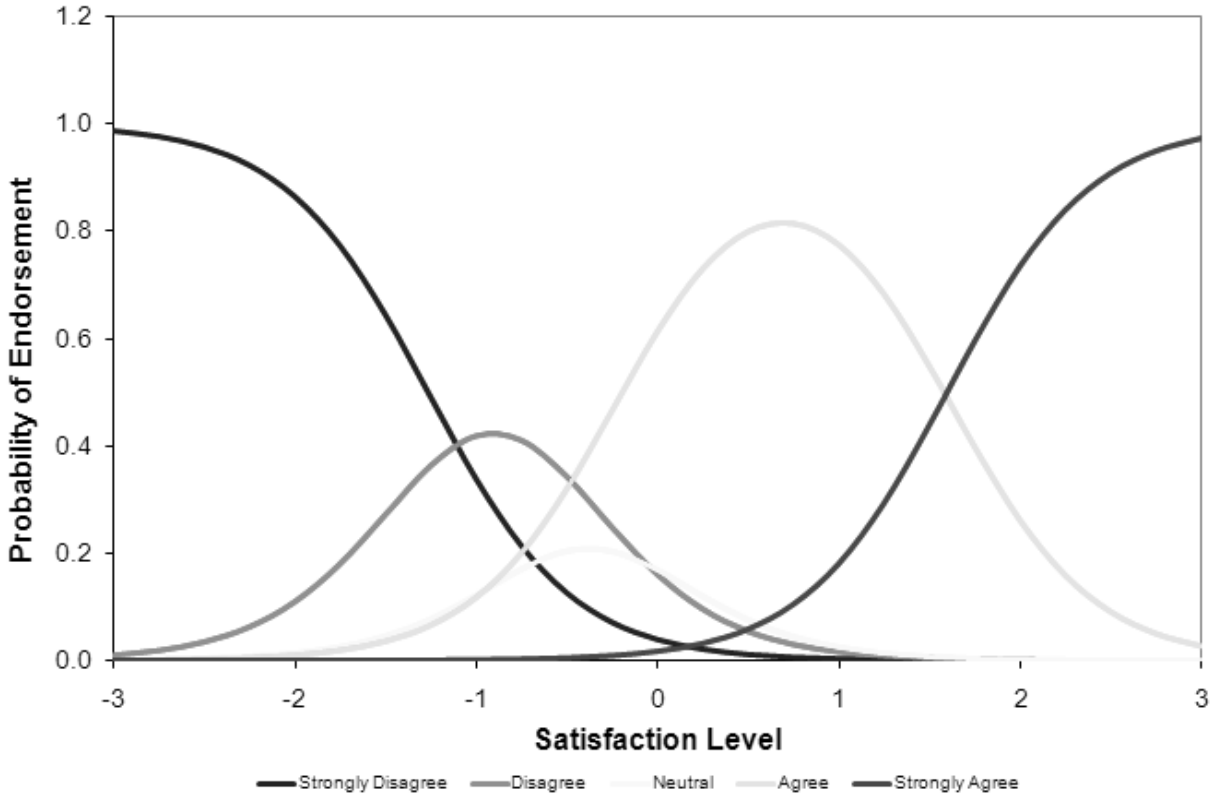

Figure 20: C6 Students have a sense of pride at KUMC item characteristic curve.

C3 Students have voice in decision-making at KUMC (Figure 21) is the second hardest item to endorse. The item adequately discriminates between the five response categories.

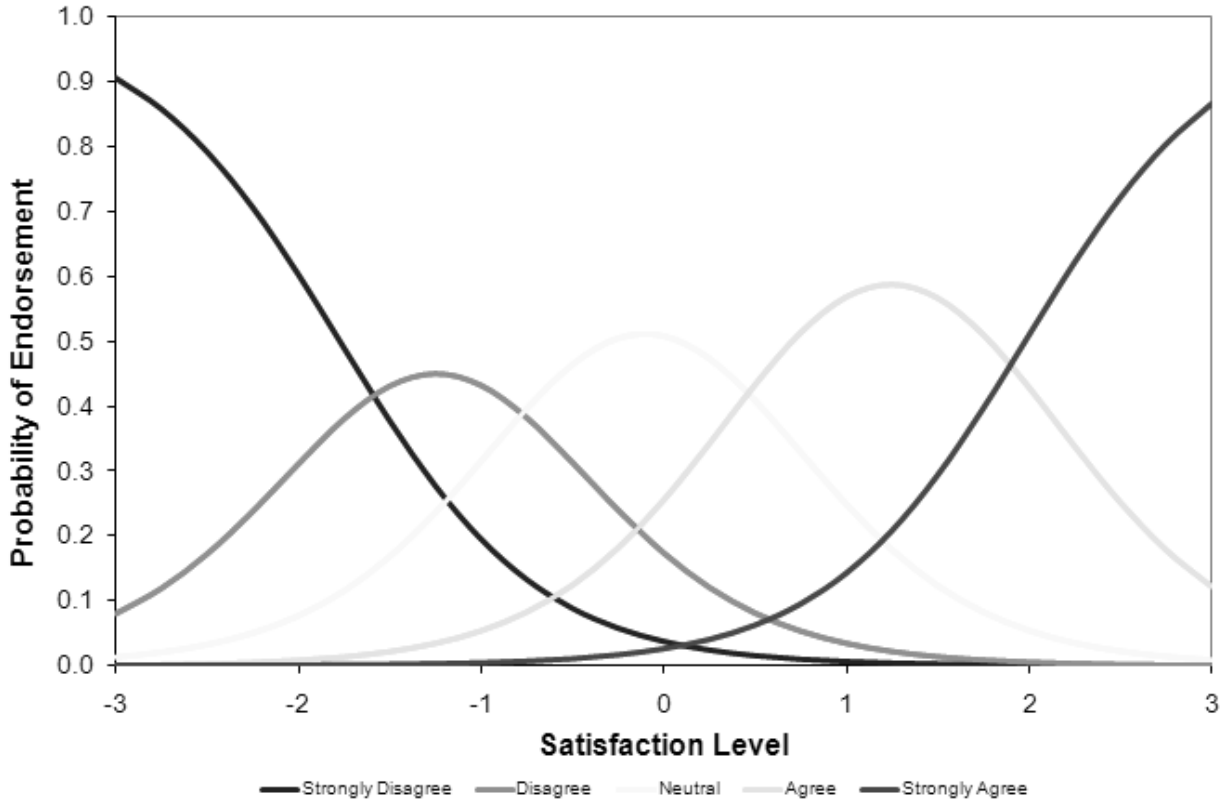

Figure 21: C3 Students have a voice in decision-making at KUMC item characteristic curve.

C5 There is a feeling of community at KUMC (Figure 22) is the hardest item to endorse. The item adequately discriminates between the five response categories.

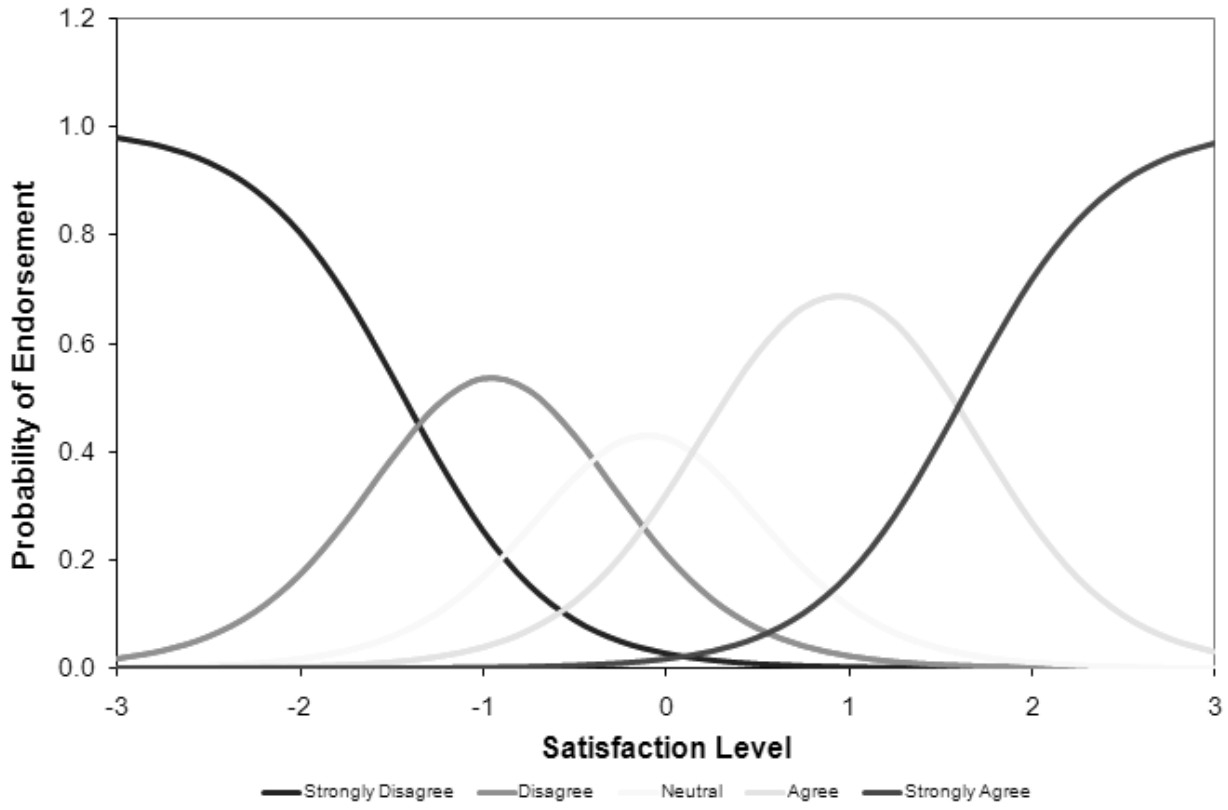

Figure 22: C5 There is a feeling of community at KUMC item characteristic curve.

Item Information Curves

The following section individually discusses the eight community item information curves in detail. The items are presented in order of least to most amount of information provided by the item. The order was determined by visually inspecting the item information curves.

C2 Campus security/safety is adequate provides the least amount of information (Figure 23). The information that it provides is fairly constant across all satisfaction levels but provides the least amount of information at low and high levels of satisfaction.

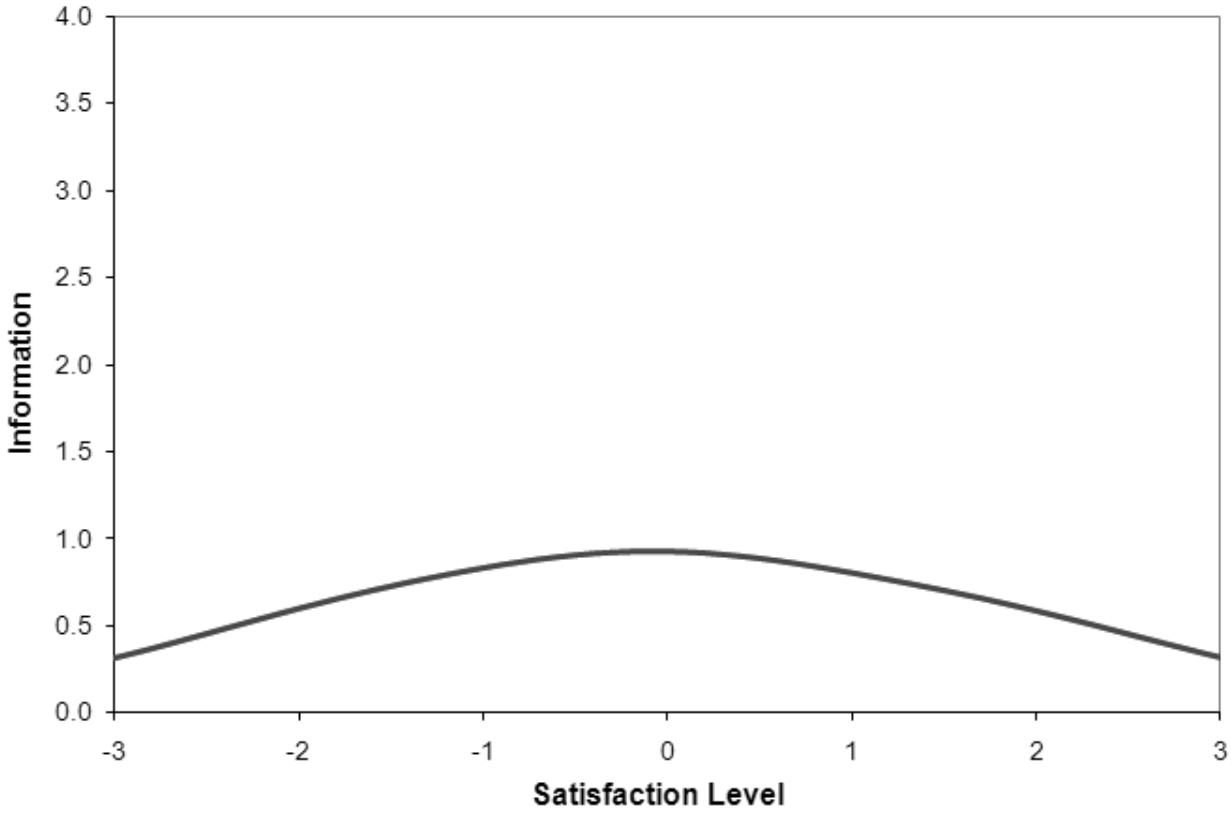

Figure 23: C2 Campus security/safety is adequate item information curve.

C4 KUMC has a positive image in the Kansas City area provides the second least amount of information (Figure 24). The information that it provides is fairly constant across all satisfaction levels but provides the least amount of information at low and high levels of satisfaction.

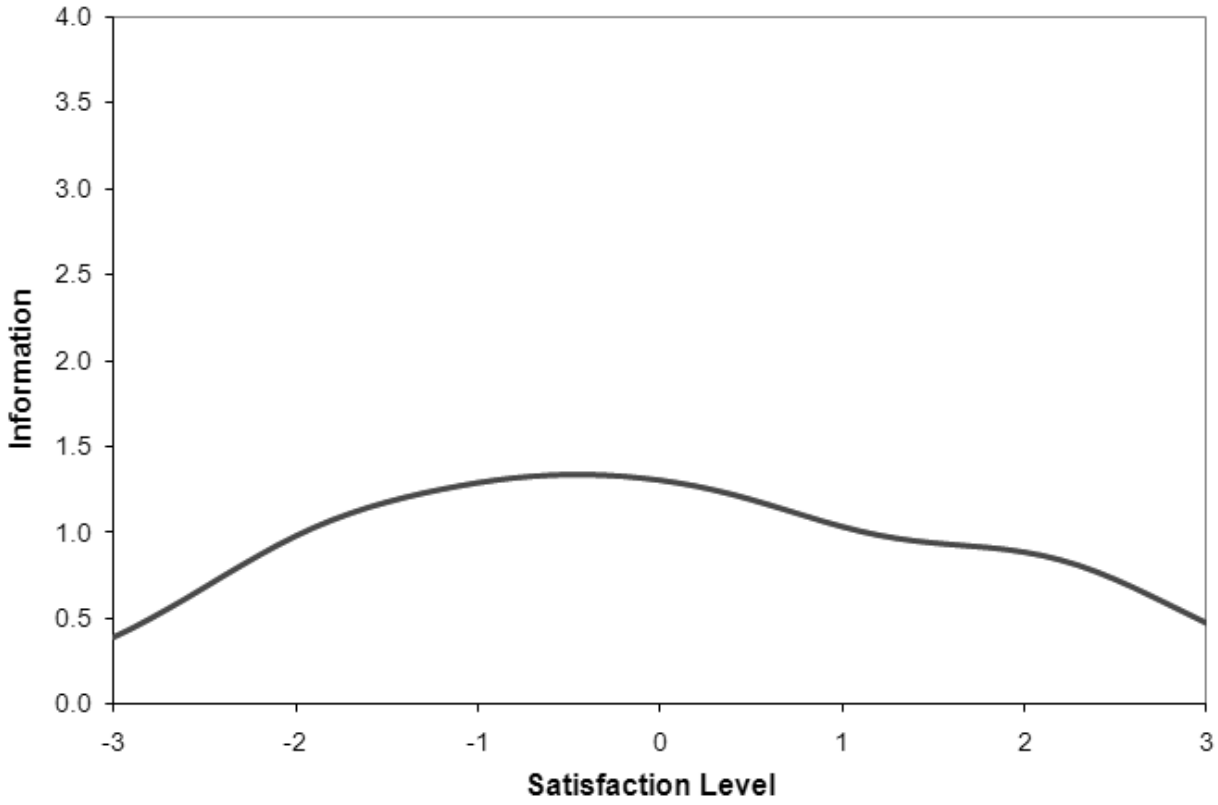

Figure 24: C4 KUMC has a positive image in the Kansas City area item information curve.

C3 Students have a voice in decision0making at KUMC provides the third least amount of information (Figure 25). The information that it provides is fairly constant across all satisfaction levels but provides very little information at the low levels of satisfaction. This is evidenced by the sharp decline on the left side of the graph. The decline in the higher levels of satisfaction is not as steep.

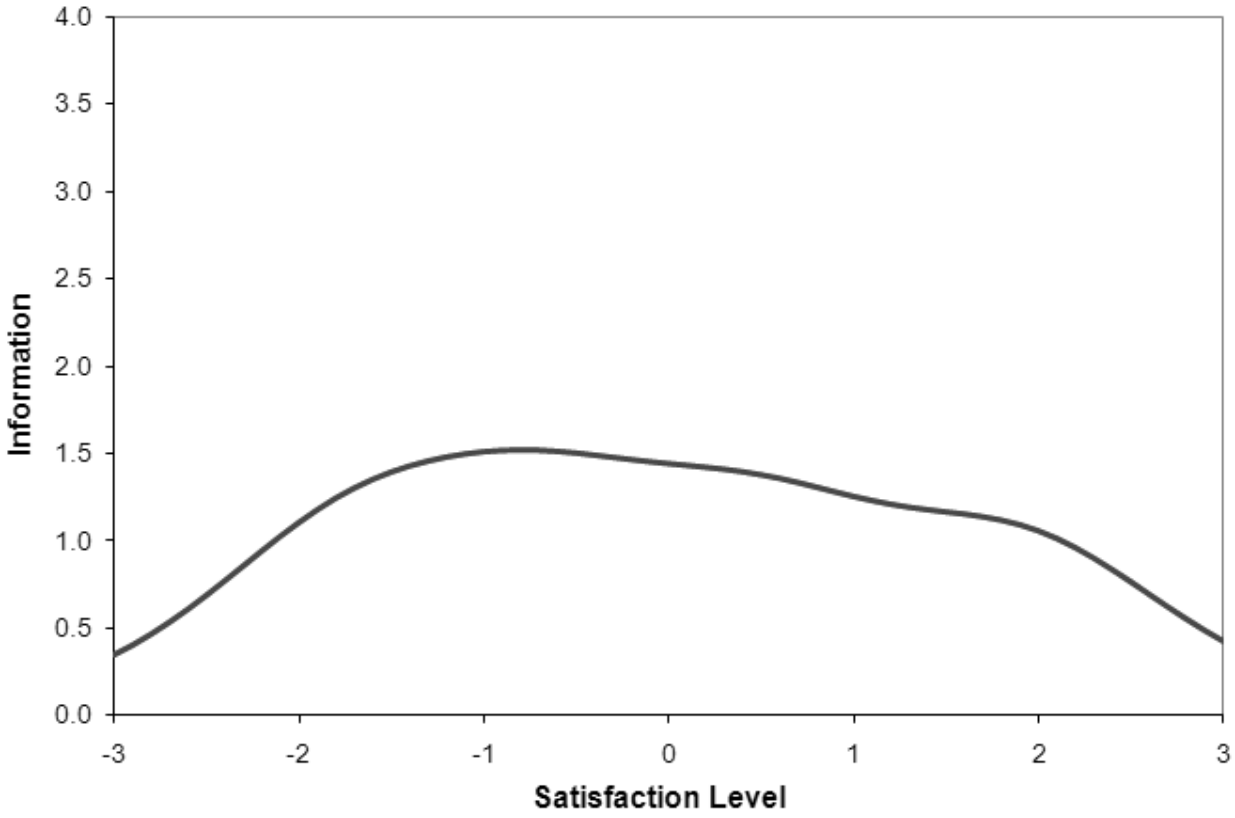

Figure 25: C3 Students have a voice in decision-making at KUMC item information curve.

C1 The environment at KUMC is friendly and supportive of students provides the fourth least amount of information (Figure 26). C1 provides most of its information in the -1 to 2 range of satisfaction which is evidenced by the sharp peak and plateau in that range.

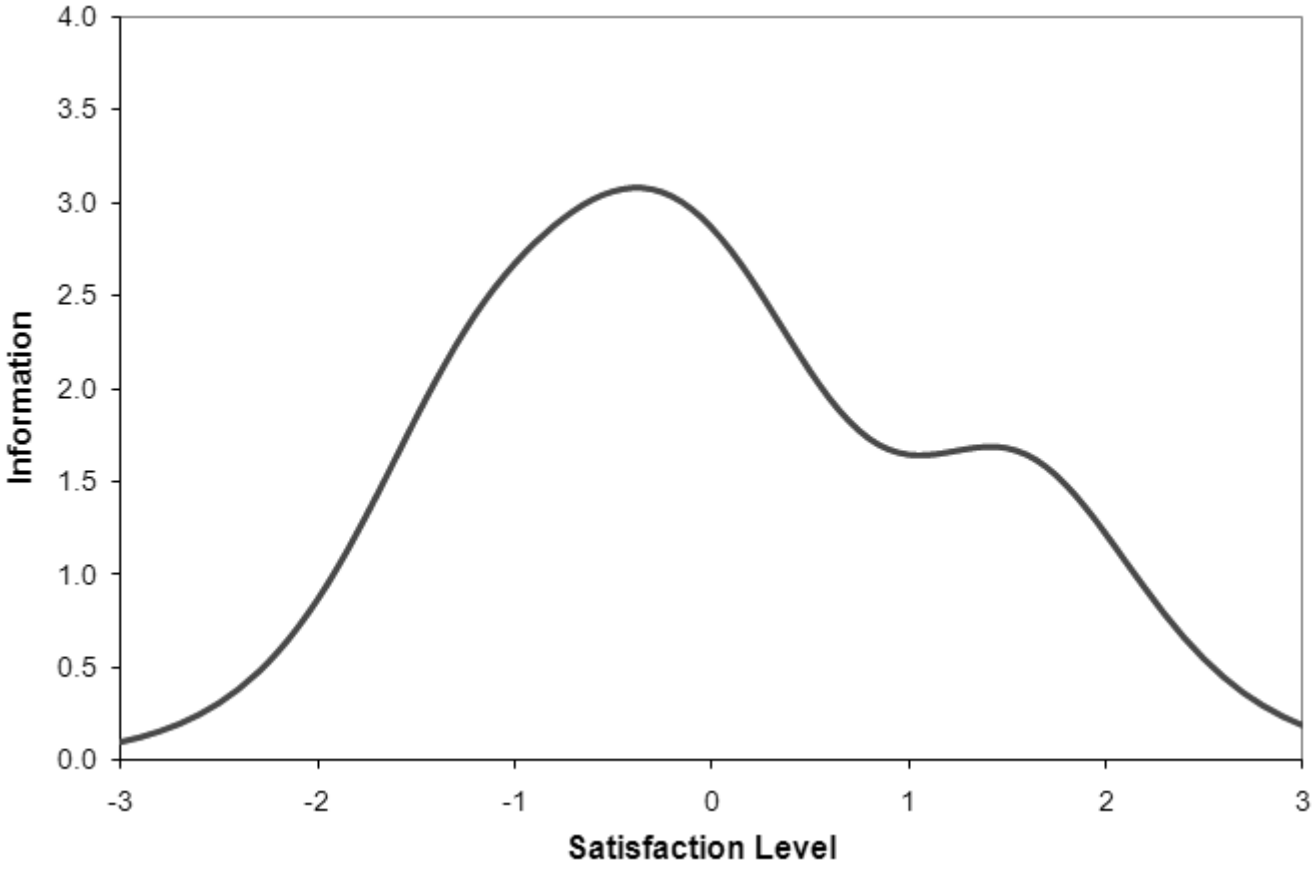

Figure 26: C1 The environment at KUMC is friendly and supportive of students item information curve.

C8 It is easy to approach staff at KUMC provides the fourth most amount of information (Figure 27). C8 provides most of its information in the -2 to 2 range of satisfaction which is evidenced by the peak and plateau in that range.

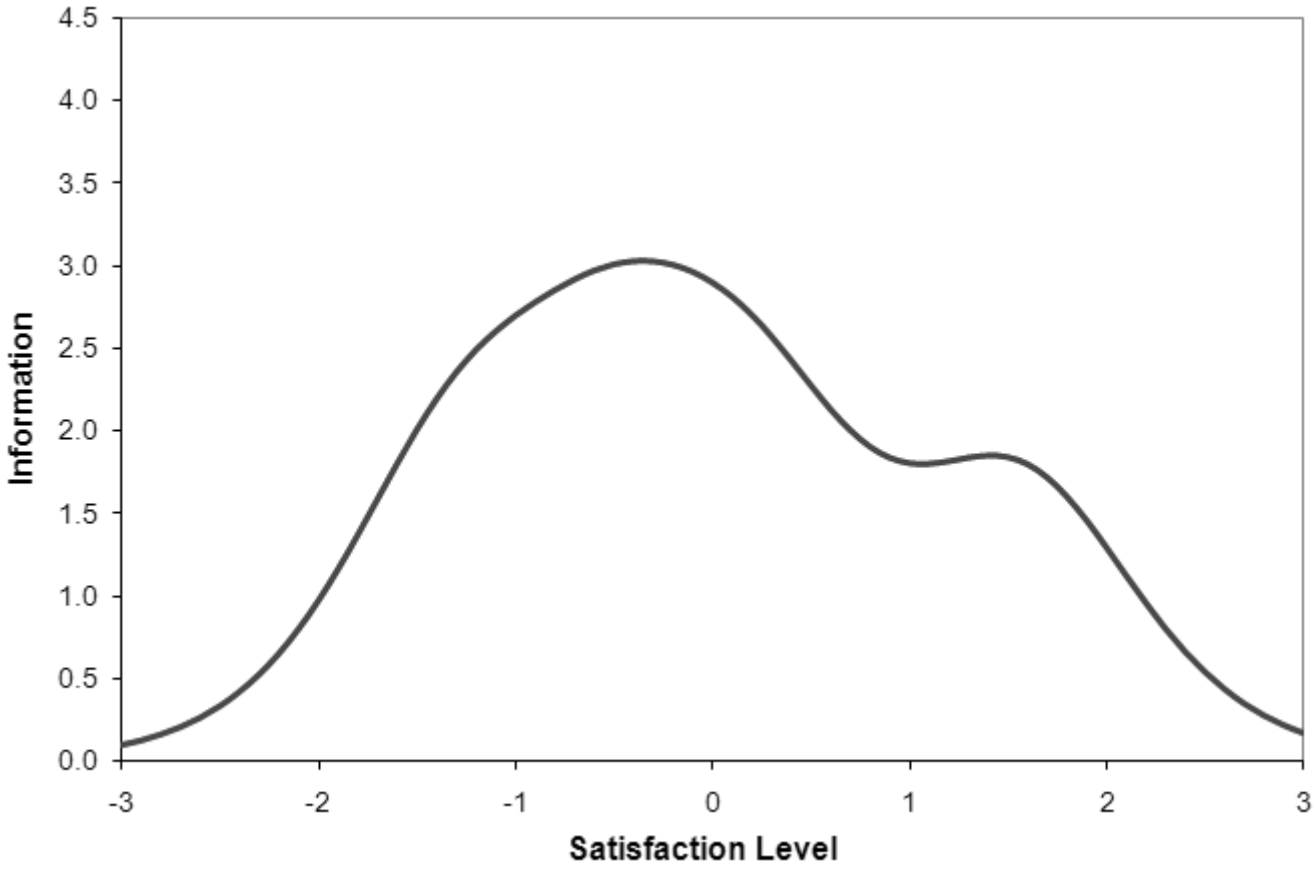

Figure 27: C8 It is easy to approach staff at KUMC item information curve.

C5 There is a feeling of community at KUMC provides the third most amount of information (Figure 28). C5 provides most of its information in the -2 to 2 range of satisfaction which is evidenced by the peak and plateau in that range.

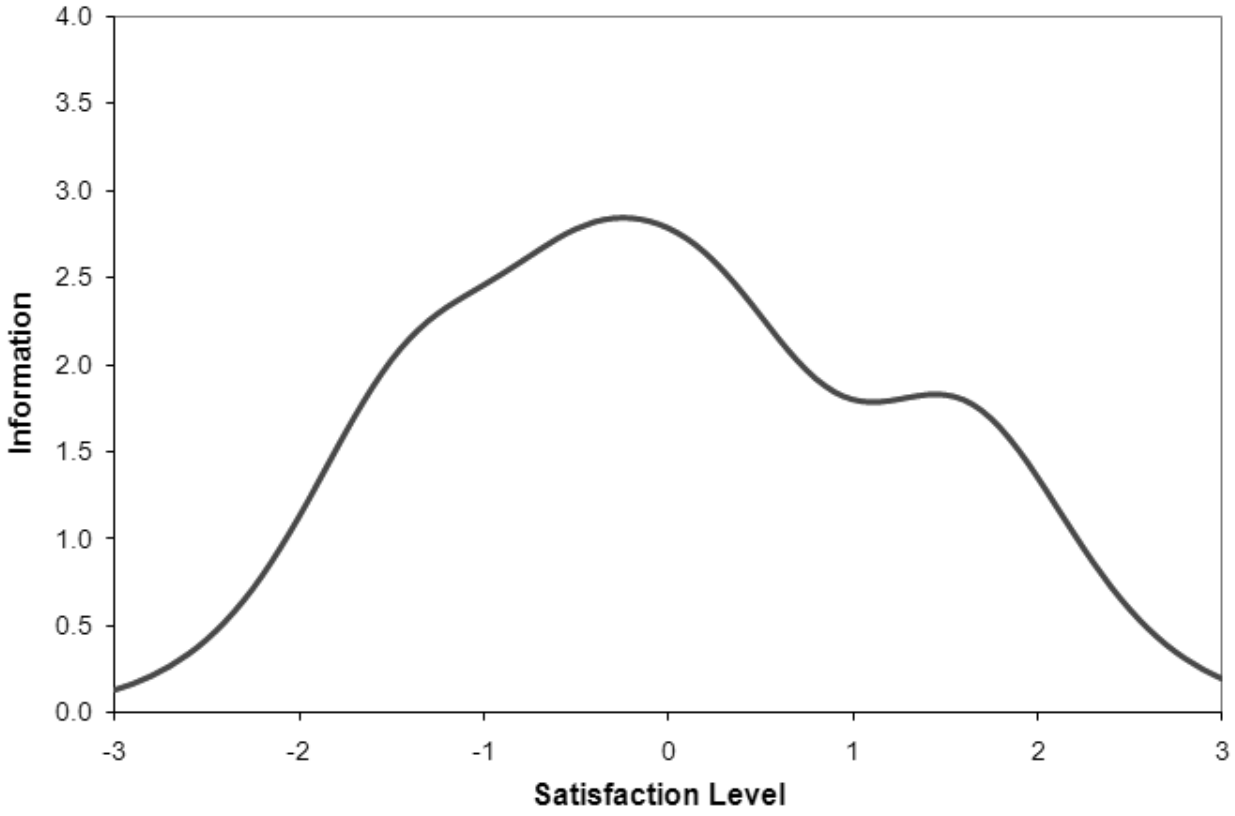

Figure 28: C5 There is a feeling of community at KUMC item information curve.

C6 Students have a voice in decision making at KUMC provides the second most amount of information (Figure 29). C6 provides most of its information in the -2 to 2 range of satisfaction which is evidenced by the peak and plateau in that range.

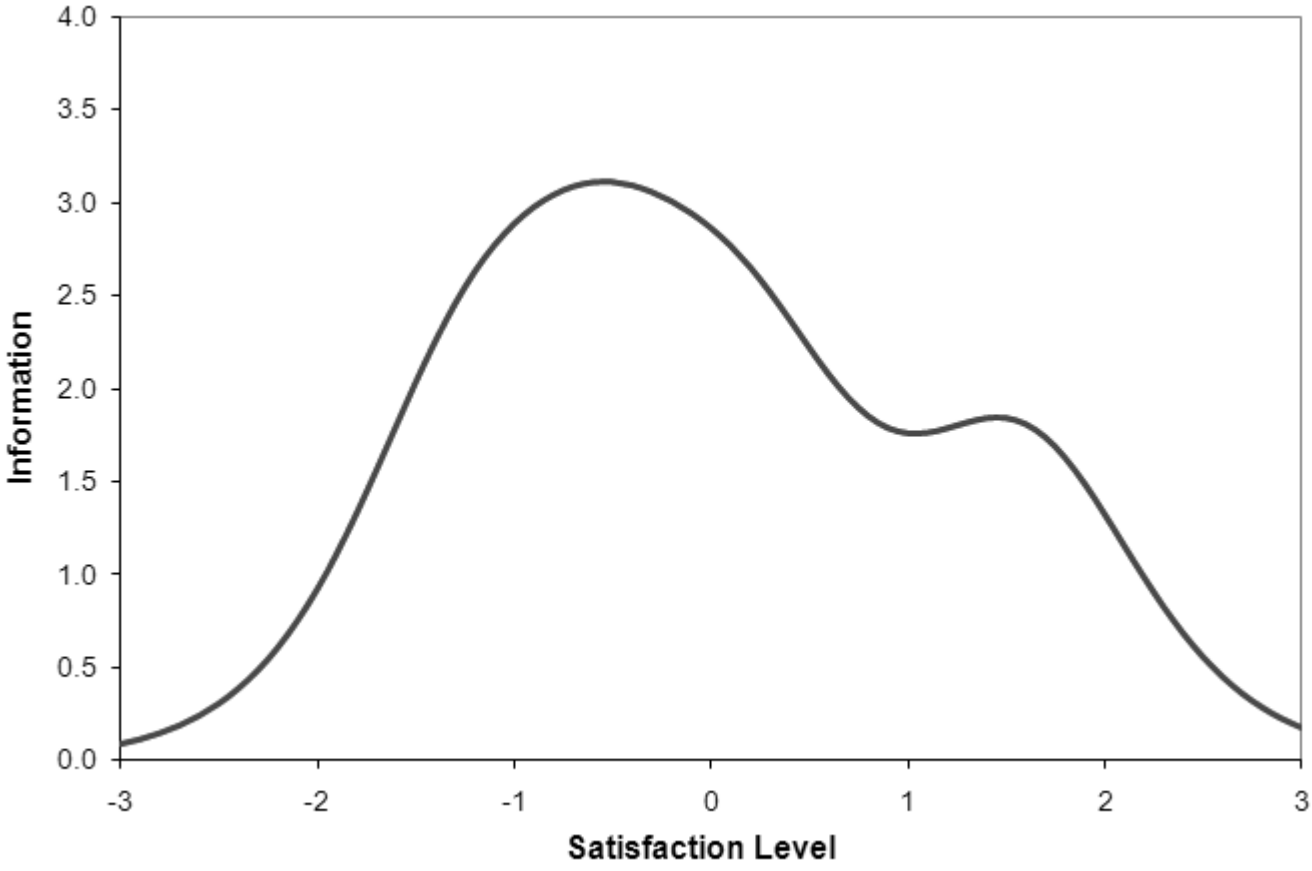

Figure 29: C6 Students have a sense of pride at KUMC item information curve.

C7 It is easy to approach faculty at KUMC provides the most amount of information (Figure 30). C7 provides most of its information in the -2 to 2 range of satisfaction which is evidenced by the peak and plateau in that range.

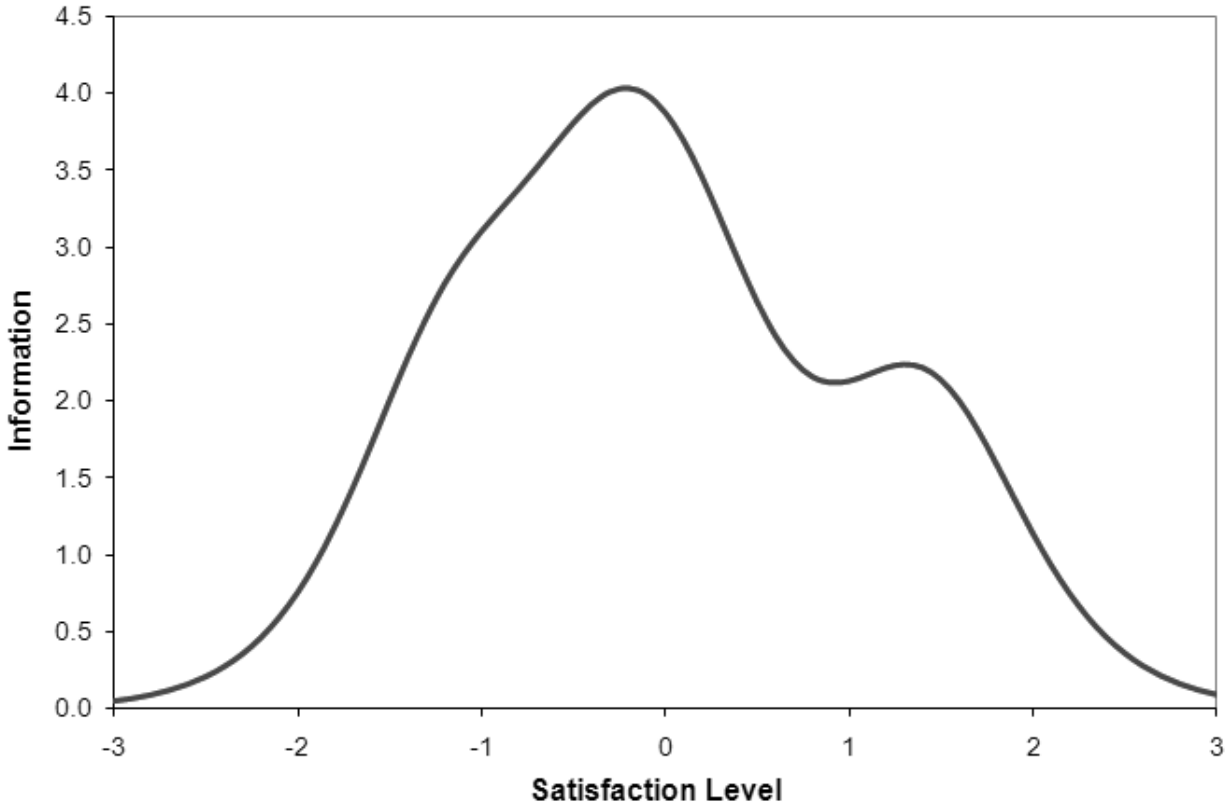

Figure 30: C7 It is easy to approach faculty at KUMC item information curve.

Figure 31 represents the total information of the eight community items and the standard error with respect to satisfaction level (The scale for total information is on the left y-axis and the scale for standard error is on the right y-axis). The most information about a participant is acquired in the -2 to 2 ranges of satisfaction. Analogous to reliability in the classical test theory sense, the lower levels of standard error in these ranges indicate that a respondent's level of satisfaction is measured with more precision in the -2 to 2 range than outside of that range.

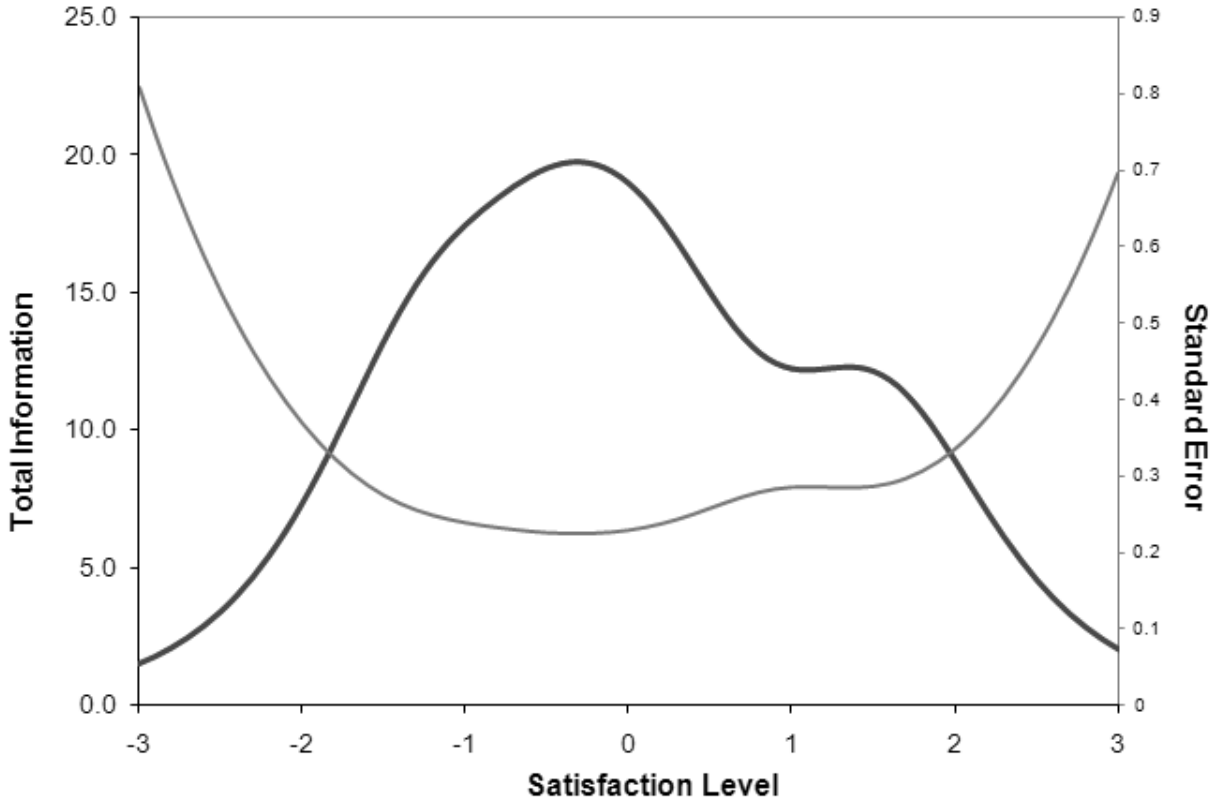

Figure 31: Total information and standard error for the eight community items.

Diversity

Parameter Estimates

Parameter estimates were not obtained for the 11 diversity items since PARSCALE could not get the model to converge. Inspection of the output files indicated that D6 (Students are treated the same regardless of ancestry) was preventing the model to converge. D6 was omitted and estimation stopped with the message "matrix is singular." Changing the constant from 1.7 to 1 in accordance with DeMars (2005) allowed the estimation to run properly. DeMars indicated that changing the constant to 1 should not have an effect on the fit statistics. The model converged after 19 iterations. The *a* parameter for each item that was provided by PARSCALE was divided by 1.7 to adjust for the change in the constant that was necessary to get PARSCALE to complete the estimation. The estimated parameters are displayed in Table 10. The category response thresholds are depicted by c_i for all the items and are not to be confused with the *c* parameter that is part of the three-parameter item response theory logistic model.

Table 10

Discrimination (a), Difficulty (b), and Category Response Thresholds (c_i) Parameters for the Convergent Graded Response Model of the 10-Diversity Items

Item	*a*	*SE*	*b*	*SE*	*c*(1)	*c*(2)	*c*(3)	*c*(4)
D1 Students are treated the same regardless of race	5.706	0.303	-1.032	0.013	-0.628	-0.310	0.152	0.786
D2 Students are treated the same regardless of color	5.990	0.303	-1.042	0.014	-0.652	-0.298	0.166	0.784
D3 Students are treated the same regardless of creed	5.893	0.327	-1.090	0.012	-0.745	-0.316	0.204	0.857
D4 Students are treated the same regardless of religion	4.196	0.170	-1.108	0.014	-0.771	-0.340	0.205	0.906
D5 Students are treated the same regardless of national origin	5.480	0.266	-1.038	0.012	-0.623	-0.347	0.155	0.815
D7 Students are treated the same regardless of gender	3.614	0.131	-1.096	0.014	-0.747	-0.290	0.157	0.880
D8 Students are treated the same regardless of age	3.322	0.125	-1.101	0.015	-0.724	-0.347	0.158	0.912
D9 Students are treated the same regardless of sexual orientation	3.500	0.144	-1.068	0.015	-0.633	-0.448	0.196	0.884
D10 Students are treated the same regardless of disability	3.495	0.140	-1.157	0.015	-0.633	-0.440	0.268	0.959
D11 Students are treated the same regardless of veteran status	3.031	0.118	-1.216	0.017	-0.788	-0.614	0.383	1.019

Assessment of Fit

As shown in Table 11, no items in addition to the overall test were able to obtain a satisfactory likelihood ratio chi-square statistic.

Table 11*Likelihood Ratio Chi-Square Statistics for Assessing Fit in Diversity Items*

Item	χ^2	*df*	*p*
D1 Students are treated the same regardless of race	1108.53	9	0.000
D2 Students are treated the same regardless of color	1186.11	9	0.000
D3 Students are treated the same regardless of creed	1018.70	8	0.000
D4 Students are treated the same regardless of religion	728.93	9	0.000
D5 Students are treated the same regardless of national origin	990.42	9	0.000
D7 Students are treated the same regardless of gender	666.73	9	0.000
D8 Students are treated the same regardless of age	600.62	10	0.000
D9 Students are treated the same regardless of sexual orientation	610.44	10	0.000
D10 Students are treated the same regardless of disability	500.20	8	0.000
D11 Students are treated the same regardless of veteran status	403.96	8	0.000
Total	7814.91	89	0.000

Figure 32[10] presents the distribution of ability estimates that were obtained. The area under the normal curve equals that total area of the histogram. The ability estimates appear to not have a normal distribution.

[10] The X-axis is labeled as ability but actually represents satisfaction level.

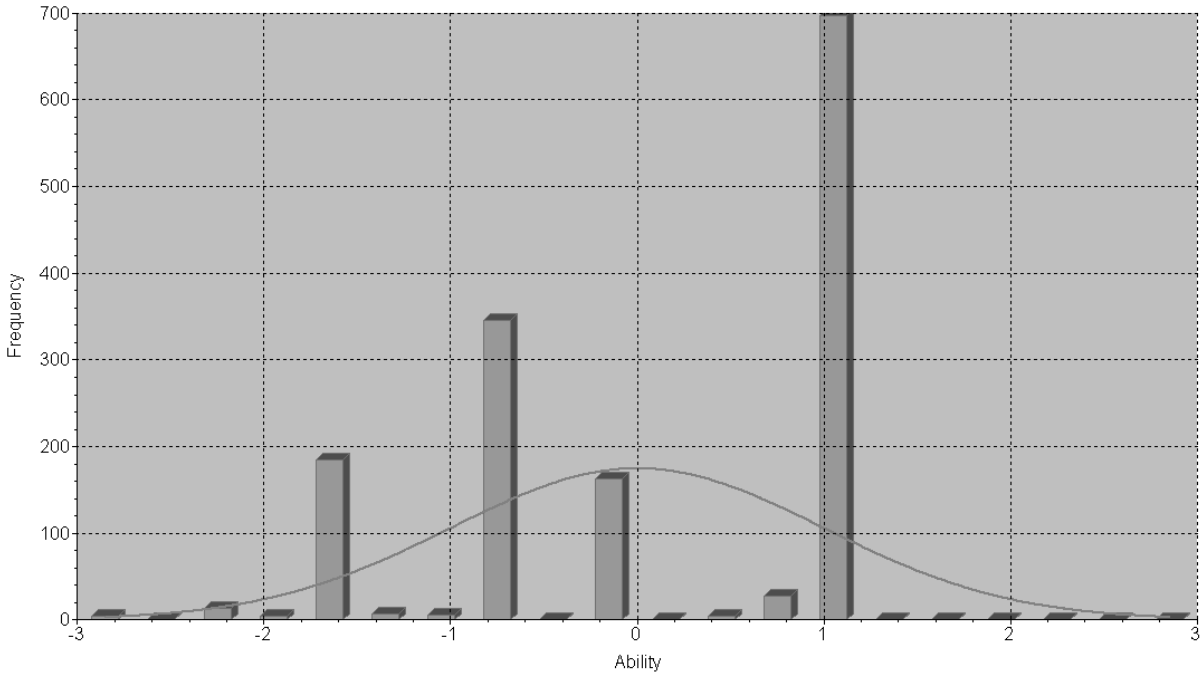

Figure 32: Gaussin fit to satisfaction scores for the 10 diversity items.

Based on the difficulties getting PARSCALE to converge, the fit statistics, and the lack of normality of the ability estimates, it was decided that creating a reliable scale with the diversity items was problematic. The individual item characteristic curves, item information curves, and total information curves for the ten diversity items are presented in Appendix E.

Substance Abuse

Parameter estimates for the four substance abuse items were not obtained due to PARSCALE's inability to converge the model.

Community Satisfaction Score Transformation

For interpretation purposes, a raw score conversion table was created for the community satisfaction subscale. The new score is called the community satisfaction score. The raw score has a range from 8 to 40 where 8 corresponds to a respondent who indicated strongly disagree on every item and 40 corresponds to a respondent who indicated strongly agree on every item. For the community satisfaction score, a score of 50 corresponds to a theta estimate of zero and the scale has a standard deviation of 4.

Table 12
Community Satisfaction Raw Score Conversion Table

Raw Score	Community Satisfaction Score
8	16
9	39
10	41
11	42
12	43
13	44
14	45
15	45
16	46
17	46
18	47
19	47
20	48
21	48
22	49
23	49
24	50
25	50
26	51
27	51
28	52
29	52
30	53
31	54
32	54
33	55
34	56
35	56
36	57
37	58
38	59
39	61
40	93

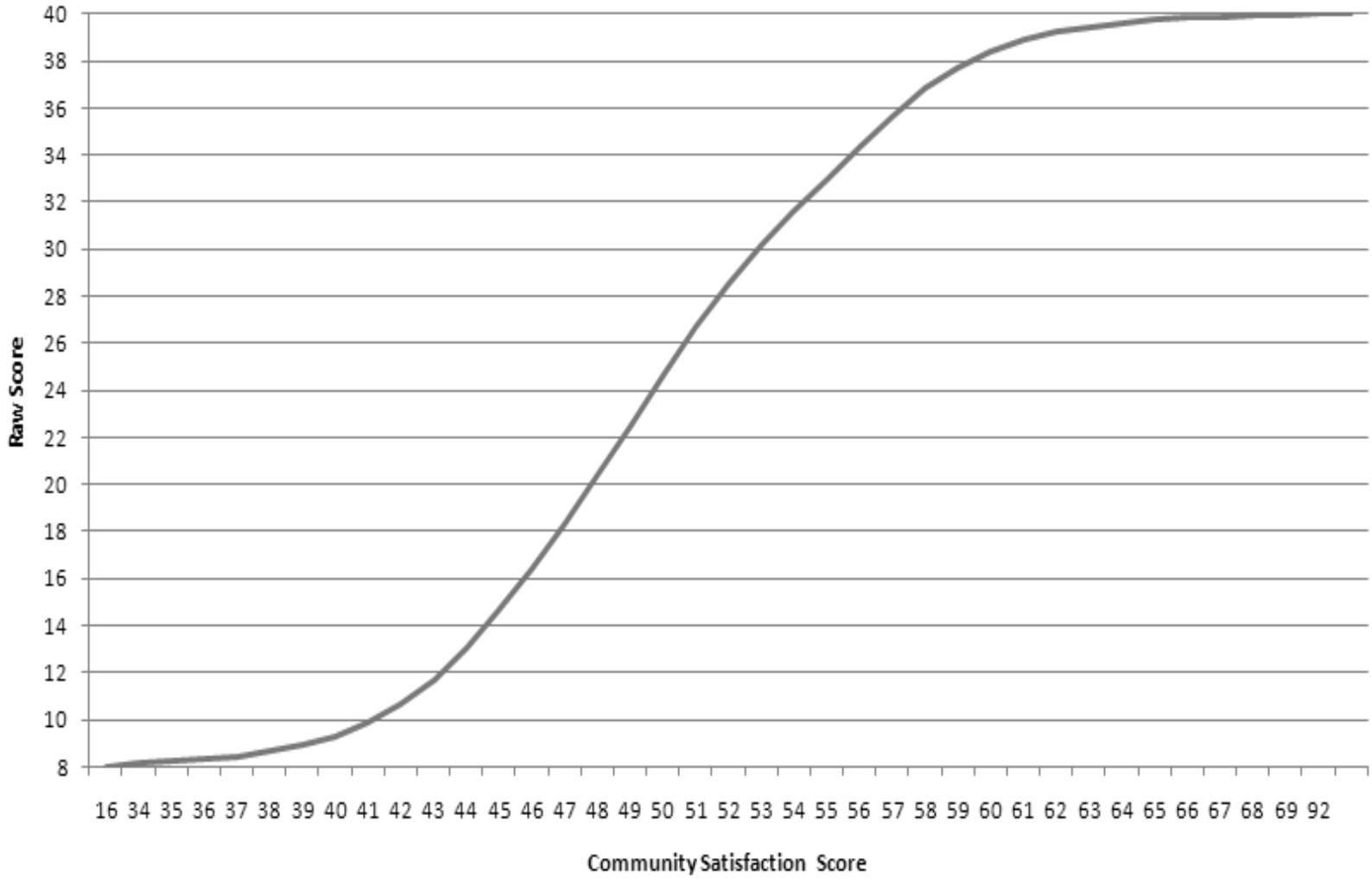

Figure 33: Community satisfaction score conversion chart.

Validity Evidence

A 2x2x3 analysis of variance (ANOVA) was conducted to determine the effect of student involvement, mode of survey administration, and student academic school on the dependent variable, community satisfaction score, as well as any interaction effects. Table 13 presents the means and standard deviations for the community satisfaction score for each of the independent variables.

Table 13
Means and Standard Deviations on the Community Satisfaction Score for Student Involvement, Mode of Survey Administration, and Student Academic School

	N	*M*	*SD*
Student Involvement			
Involved	728	61.4	13.9
Not Involved	614	60.8	13.4
Mode of Survey Administration			
Online	894	57.7	10.2
Paper	448	68.1	16.8
Student Academic School			
Allied Health	238	60.7	13.4
Nursing	798	61.0	13.1
Medicine	306	61.4	14.0

The results of the ANOVA are presented in Table 14. Significant differences in the community satisfaction score were detected for mode of survey administration, $F(1,1342) = 126.9$, $p < 0.00$, $\eta^2 = 0.09$ (Figure 34), and non-significant differences in the community satisfaction score were detected for student involvement and student

academic school. No significant interactions were detected.

Table 14
Community Satisfaction Score Analysis of Variance Results for Student Involvement, Mode of Survey Administration, and Student Academic School

Effect	*MS*	*df*	*F*	*p*
Student Involvement (SI)	181.78	1	1.11	0.29
Mode of Survey Administration (M)	2,070.31	1	129.96	0.00
Student Academic School (S)	316.75	2	1.93	0.14
SIxM	115.44	1	0.71	0.40
SIxS	53.32	2	0.33	0.72
MxS	8.38	2	0.05	0.95
SIxMxS	404.66	2	2.48	0.08
Error	163.36	1,330		

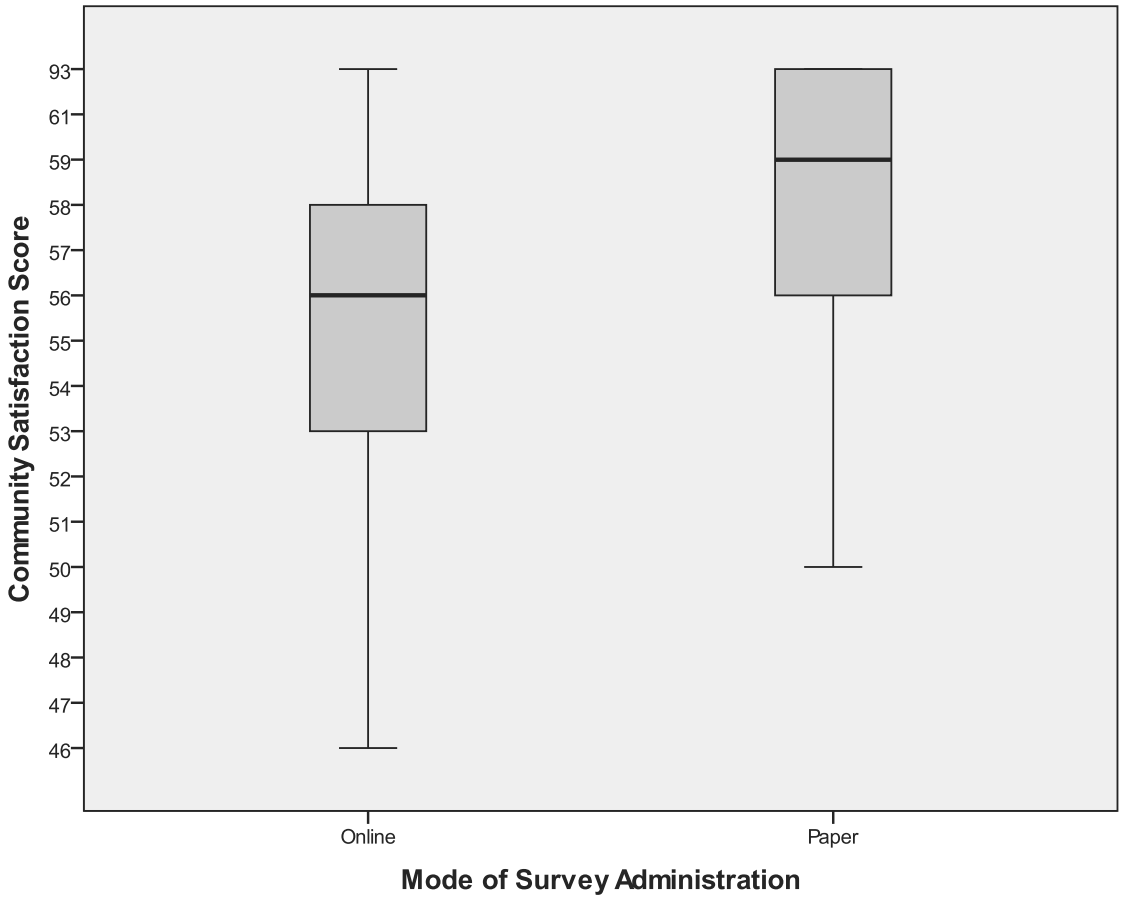

Figure 34: Distributions of community satisfaction scores for the mode of survey administration.

A 2x3 analysis of variance (ANOVA) was conducted to determine the effect of student involvement and student academic school on the dependent variable, community satisfaction score, as well as any interaction effects for the 2008 and 2009 administration of the surveys since 2008 and 2009 were the administrations that had only one survey mode (online). Table 15 presents the means and standard deviations for the community satisfaction score for each of the independent variables.

Table 15
Means and Standard Deviations on the 2008-2009 Community Satisfaction Score for Student Involvement, Mode of Survey Administration, and Student Academic School

	N	*M*	*SD*
Student Involvement			
Involved	332	58.9	11.5
Not Involved	262	57.4	9.4
Student Academic School			
Allied Health	112	57.5	9.8
Nursing	148	59.2	10.9
Medicine	334	58.1	10.8

The results of the ANOVA are presented in Table 16. Significant differences in the community satisfaction score were detected for student involvement, $F(1,594) = 5.02$, $p = 0.03$, $\eta^2 = 0.01$ (Figure 35), and non-significant differences in the community satisfaction score were detected for student academic school. The interaction was not significant.

Table 16
2008-2009 Community Satisfaction Score Analysis of Variance Results for Student Involvement, Mode of Survey Administration, and Student Academic School

Effect	*MS*	*df*	*F*	*p*
Student Involvement (SI)	566.25	1	5.02	0.03
Student Academic School (S)	198.68	2	1.76	0.17
SIxS	19.81	2	0.18	0.84
Error	112.84	588		

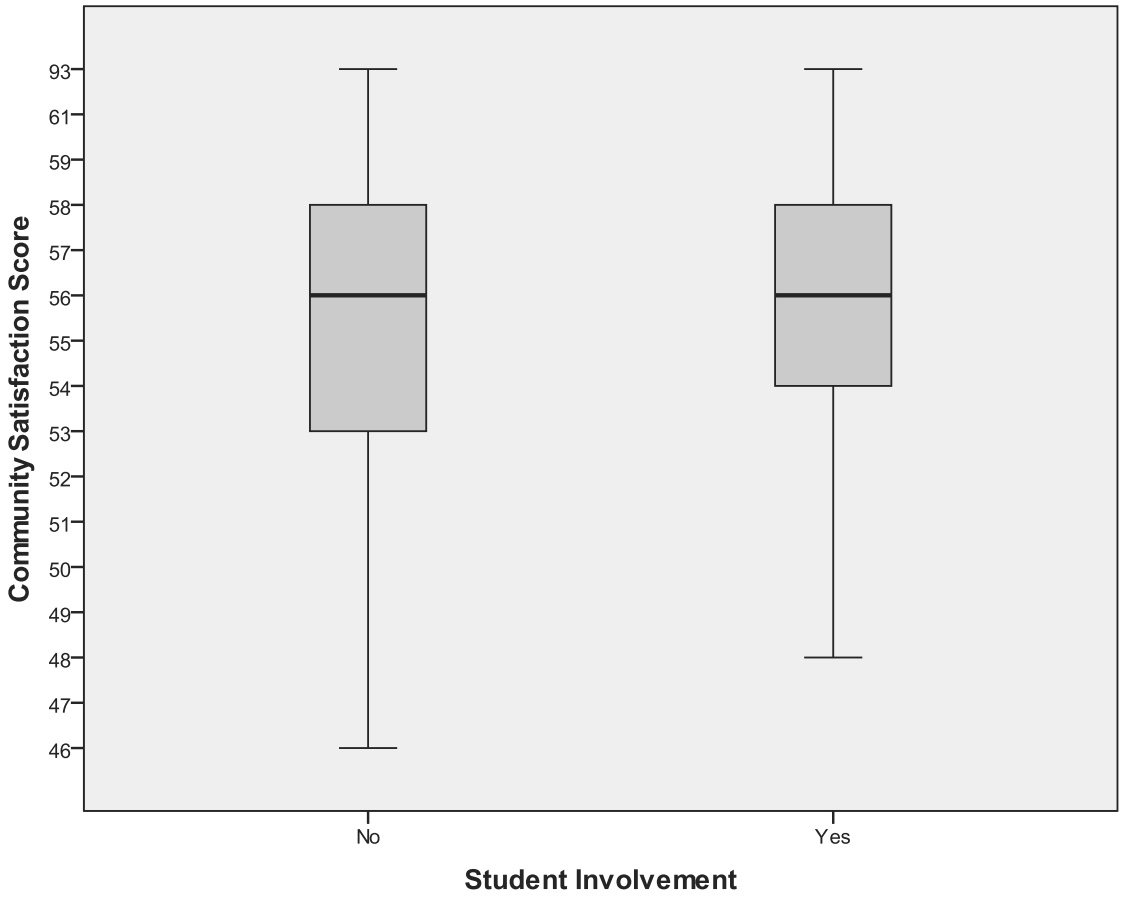

Figure 35: Distributions of the 2008-2009 community satisfaction scores for student involvement.

Summary

Three underlying variables (community, diversity, and substance abuse) related to student satisfaction were identified on the KUMC Student Services Survey. The community subscale was able to be scaled while the diversity and substance abuse scale were unable to be scaled with the graded response model. In general, the community subscale is highly reliable in the middle ranges of student satisfaction and less reliable in

the extreme ranges of student satisfaction. Significant differences were detected for mode of survey administration with students who submitted a paper survey being more satisfied than students who submitted an online survey. Non-significant differences were detected for student involvement, student academic school, and the mediating effect of student academic school and student involvement.

Due to the significant result of mode of survey administration, only the 2008-2009 results were analyzed since the survey was uni-modal for 2008-2009. Significant differences were detected for student involvement with involved students being more satisfied. Non-significant differences were detected for student academic school and the mediating effect of student academic school and student involvement.

CHAPTER 5 – DISCUSSION AND CONCLUSIONS

The purpose of this study was to determine if the KUMC Student Services Survey measured student satisfaction with reliability and validity under an graded response model. This chapter discusses the results of this study. In addition, it also provides suggestions for further areas of study.

Dimensionality

The KUMC Student Services Survey adequately measures student satisfaction amongst three subscales: community, diversity, and substance abuse. The items related to community are congruent with Astin's (1993) definition of student satisfaction:

- The environment at KUMC is friendly and supportive of students;
- Campus security/safety is adequate;
- Students have a voice in decision-making at KUMC;
- KUMC has a positive image in the Kansas City area;
- There is a feeling of community at KUMC;
- Students have a sense of pride at KUMC;
- It is easy to approach faculty at KUMC; and
- It is easy to approach staff at KUMC

Astin's definition focuses on the relationship that the student has with faculty, staff, and the institution. Factors like faculty interest in students and trust in the administration have been found to be positively associated with student satisfaction whereas the lack of

a campus community has a negative effect on campus satisfaction. The notion of a community can be a nebulous one to define but what is important for student development is that students find a community to belong to while at an institution. A community can be the larger campus, their academic department, social group, or student organizations to which they belong. Institutions should focus efforts towards developing environments and support structures that encourage student communities to flourish. Measuring the community aspect of student satisfaction on a survey like the KUMC Student Services Survey is vital to any campus administration. The results can be used to identify subgroups of students that are need of help identifying a role in the campus community or developing their own community.

The subscale of diversity is measured by 10 items that are worded similarly except the last portion of the item. Each item focuses on a protected class of students at KUMC: race, color, creed, religion, national origin, ancestry, gender, age, sexual orientation, disability, and veteran status. All of these items load extremely high into the diversity factor that was performed as part of the factor analysis but one of the challenges with the data set that was utilized is that there is little variation in item response. The majority of respondents indicate a 4 or 5 on these items. In addition, it would be very difficult for a respondent to know if a student was mistreated based on some of the protected classes since identification of individuals in that class would be difficult if you

did not actually know a student in that protected class[11]. It is plausible to conclude that most of the students respond to a 4 or 5 to these items mainly since they have never witnessed mistreatment whereas a student under one of these classes that was actually mistreated would be difficult to detect from the survey due to the low number of students that fit some of these protected classes. One area of improvement to the survey would be to include all of the protected classes as demographic items in order to identify specific responses of individuals that are part of the protected classes.

Five items cross-loaded into the community and diversity factor that were excluded from the final structure:

- There is racial/ethnic harmony at KUMC;
- Students at KUMC are free from sexual harassment;
- Male and female students are treated equally at KUMC;
- KUMC is committed to resolving diversity-related problems; and
- I am comfortable raising diversity related problems to faculty, administration, and staff

From a face validity standpoint, it makes sense that all of these items cross-loaded into both factors. In order to have a thriving campus community, the community needs to be harmonious when it comes to issues of diversity. These items were removed from the final structure that was used for the graded response model (GRM) scaling since item

[11] For example, only 14 KUMC students receive VA benefits during the Fall 2009 semester.

response theory procedures like the GRM require unideminsionality. One area of further research would be to attempt a scaling that doesn't require unideminsionality between subscales so that these items could be included in the scaling. There is also some redundancy in the items. For example, "Male and female students are treated equally at KUMC" and "Students are treated the same regardless of gender" are essentially asking about the same aspect of gender differences. Further analysis should be conducted to look at if both items should stay on the survey or if one should be removed.

Four items related to wellness were part of the final third factor, substance abuse:

- Drug use is a problem for students in my program at KUMC
- Alcohol use is a problem for students in my program at KUMC
- Drug use is a problem for me
- Alcohol use is a problem for me

Like the diversity items, the variances in responses to these items suggest problems with the reliability of scaling these items for all ability levels. It has also been believed by KUMC administration that students do not honestly answer these questions due to the high stakes possible consequences related to their academic programs despite every assurance students are given that their responses are not identifiable. Since most of a health professional student's evaluation is done through what can be considered at times subjective measures (observations in a clinical setting), students are often concerned that

all aspects of their academic career, including taking a student survey, are being evaluated.

Two items did not load into any factor. The items did load into the community factor at a very low level, but the loading levels were not high enough to include in the final structure.

- It has been beneficial for me to be a member of a student organization at KUMC
- The Student Wellness Program provides services and programs to meet my needs

Almost all of KUMC's student organizations are focused towards the student's academic program and are not the traditional extra-curricular activities that are found at more traditional institutions. It is possible that student responses to this student organization item might be problematic due to the student's inability to distinguish between a student organization and an activity in their academic department. In addition, over the last 5 years at KUMC, the overall percentage of students in programs that traditionally have high levels of involvement in student organizations (medicine and undergraduate programs) have decreased since the number of post-professional and graduate programs have increased. Students in post professional and graduate programs are less likely to get involved in student organizations since they are more likely to be part-time and/or working students. The non-loading of the item related to the Student Wellness Program is not surprising since this a specific program that only a few students participate in and is not recognizable campus-wide by students.

The final structure that was used for the GRM scaling did demonstrate relatively high levels of reliability when looking at coefficient alpha. This finding suggests that the scales are measuring student responses accurately, but these high levels of reliability can also be due to the limiting response patterns of many of the items especially the diversity items.

Item Response Theory Scaling

Reliable parameter and satisfaction level estimates were able to be successfully obtained for the community subscale but not the diversity and substance abuse subscale using the graded response model. Each of three scaling exercises are discussed in more detail.

Community

For the most part, the items of the community subscale are able to adequately discriminate between the response categories. There are a few item where it is difficult to discriminate between the neutral category but that can also be due to underutilization of that category for the item. The items on the community subscale suggest that students do not need to have a high level of satisfaction in order to endorse the item and the subscale is able to provide the most useful information about a student's level of satisfaction in the middle ranges of the latent trait.

Specifically, items related to decision making and the friendly and supportive nature of the campus require very low levels of satisfaction to endorse whereas the items relating to if students have a voice on campus and believe there is a feeling of community

require slightly higher levels of satisfaction in order to endorse. This finding suggests that there is more objectivity in the community items that require a stronger feeling of community from the student in order to feel satisfied with the community than other items that have a more subjective interpretation like the friendly nature of a campus.

One aspect worth discussion is that the KUMC Student Services Survey has also experienced a trending increase in the favorable responses on the items, especially for the years that were used in this study. During the early administrations of the survey, there were many items where respondents did not have high levels of agreement. For example, Figures 36-39 demonstrate how agreement response[12] in several items has longitudinally increased since 1993 (Meiers and Gove, 2009).

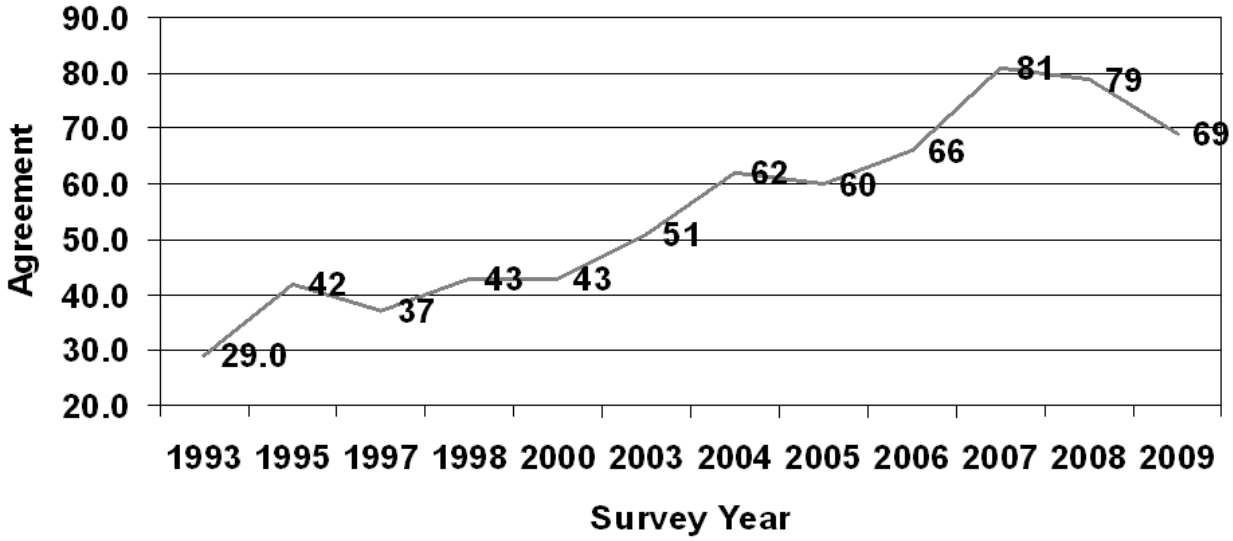

Figure 36: Agreement response for students have a voice in decision making at KUMC 1993-2009.

[12] For summary purposes, respondents who indicated a 4 (agree) or 5 (strongly agree) are collapsed into a general agreement category and the Y-axis represents percentage of agreement.

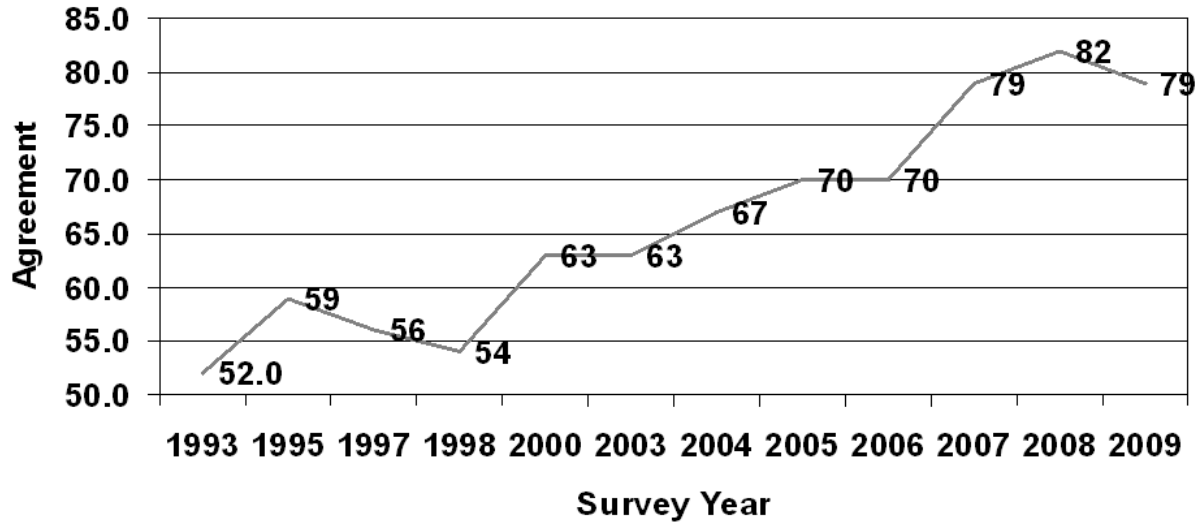

Figure 37: Agreement response for there is a feeling of community at KUMC 1993-2009.

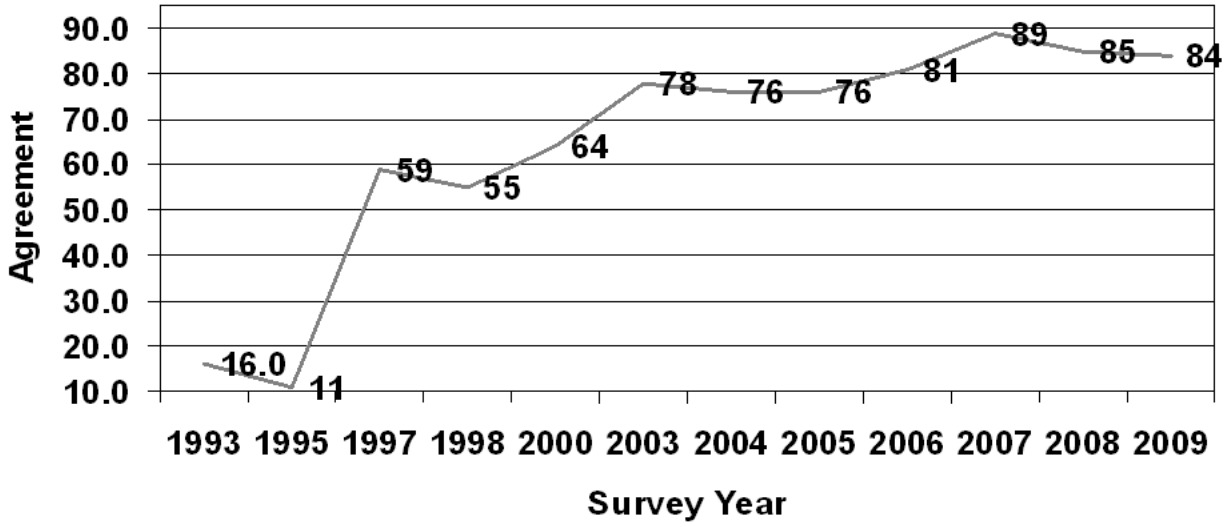

Figure 38: Agreement response for students are free from sexual harassment 1993-2009.

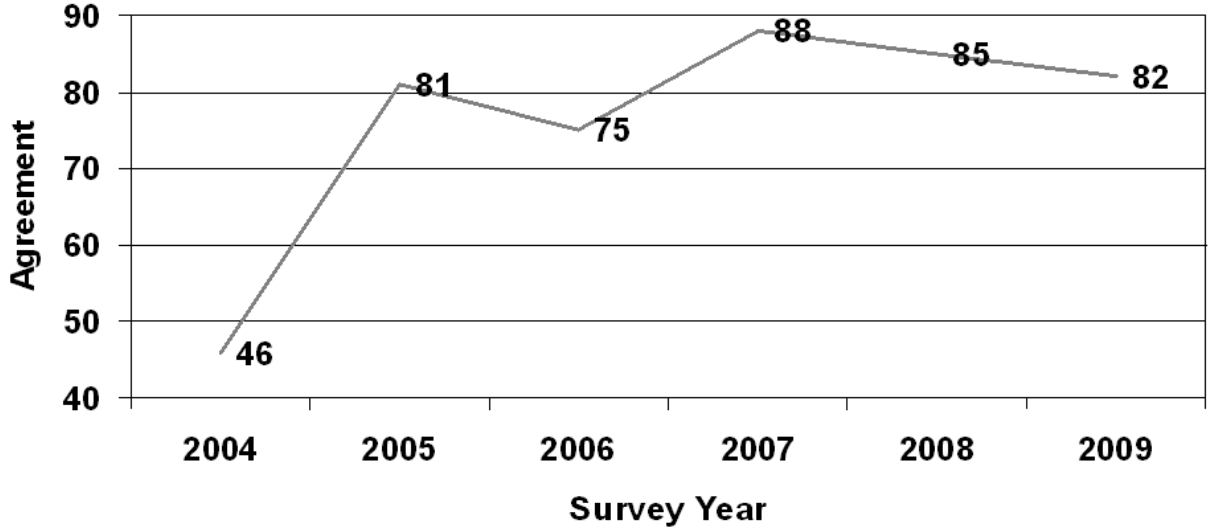

Figure 39: Agreement response for students are treated the same regardless of color 2004-2009.[13]

The long-term increase in satisfaction is mainly attributed to KUMC administration's willingness to create definitive action from the survey, but is also why the range of responses for the data that was used in this study is limited. While highly satisfied results are desirable to campus administration, it would be more preferable to have wider ranges of survey response for scaling purposes. One area of further research would be to obtain the survey responses from all the years of the survey and repeat the scaling exercises to see how parameter estimates change. It is plausible to hypothesize that by incorporating data from less satisfied respondents that more information could be obtained from all ranges of student satisfaction and improve the overall reliability of the scale.

[13] This item was introduced on the survey in 2004

The community satisfaction score that was created through linear transformation is a useful application of item response theory in student development assessment. The raw score conversion chart (which is more common in achievement testing) can be used by student affairs practitioners administering this survey in order to create a reliable estimate of student satisfaction for comparison purposes. As mentioned previously, the most predominant method towards establishing the reliability of student development surveys is through classical methods. The advantage to the method in this study is that it allows researchers to determine at which levels of student satisfaction[14] the items on the survey are most useful

Diversity

There were some challenges in getting PARSCALE to converge a model for the diversity items. The item related to ancestry was problematic and once it was removed from the model, PARSCALE was able to easily converge the model. Ancestry could be considered a vague concept and also redundant to ethnicity so removing the item was not considered problematic. Even though PARSCALE was able to converge a model, creating a reliable scale for the diversity items was considered problematic mainly due to complete lack of normality to the satisfaction estimates.

One limitation of assessing the quality of diversity items is that many respondents are not part of the protected classes that the items are assessing and then to not have an opinion on the item. One area of further research would be to include demographic

[14] The method could be applied to other traits that are desired to be measured in student development research.

questions related to these items to determine if students are part of the protected classes.[15] Another area of possible research is to conduct a differential item functioning study to determine if the diversity items have some bias based on the protected classes. A study such as this would further assess the quality of these items for measuring diversity constructs related to student satisfaction.

Based on how the diversity items are worded where the only phrase on the item that changes is the last, there could also be a possible stem effect that is preventing the items to be useful for measuring student satisfaction. Several of the items are also redundant or vague. One area of further research would be to re-write the diversity questions to minimize any possible stem effects.

Substance Abuse

It is not surprising that the model would not converge for the substance abuse subscale mainly for three reasons. First, the subscale only consisted of four items. The items related to the self scale have been suspected to not be honestly answered by students over the years. And finally, the nature of the items does not appear to be unidemensional which is an assumption of item response theory scaling. Two of the items are focused towards personal use of drug and alcohol and the other two items are related to perceptions of drug and alcohol use amongst classmates. A discussion about the utility of how to use these items on the KUMC Student Services Survey is warranted. There might still be value in asking these types of items on the survey for early detection purposes of

[15] Race, gender, and age are currently asked on the KUMC Student Services Survey

drug and alcohol issues, but developing a reliable graded response model scale from these items is not appropriate. If campus administration is interested in obtaining comprehensive data about student wellness, there are more established wellness surveys such as the National College Health Assessment administered by the American College Health Association (About ACHA-NCHA) that can be utilized.

Survey Validation

Based on established research, several hypotheses about survey response were analyzed to determine if the KUMC Student Services Survey validly measures the construct of student satisfaction. During the initial analysis, none the three hypotheses was supported, but once the analysis was repeated with the survey years where there was only one mode of survey administration, one of the three hypotheses was supported. It is concluded that the validation of the KUMC Student Services Survey will be a continuous process with further areas of research. Each of the hypotheses is discussed in further detail in this section.

Mode of Survey Administration

Perhaps the most surprising result was that significant differences in community satisfaction scores between online and paper administrations of the survey were detected. Online respondents indicated lower levels of satisfaction on both the community and diversity scale. Item level comparisons between online- and paper-based surveys have previously indicated no significant differences in item response (Meiers & Gove, 2007). Prior to the analysis, the multi-modal nature of the survey was considered problematic

and further research into this finding is warranted. This could be easily accomplished by administering the survey to students at other institutions with random assignment between paper and online versions for a more experimental setting to determine if item response differs between different modes of the survey.

Student Involvement

Involvement in a student organization was not found to show significant differences in community satisfaction scores which was a surprising result based on established research about student involvement. Since KUMC has had a significant increase in enrollment of online-only students who are not typically involved in student organizations, online students were completely excluded from the survey beginning in 2008 (Meiers & Gove, 2008). The analysis was repeated just for the 2008 and 2009 administrations of the survey since the respondents would have been entirely on-campus students, and significant differences in community satisfaction scores were detected. As expected, involved students are more satisfied than non-involved students which is suggests that the community satisfaction scale has some validity in measure student satisfaction.

Student Academic School

A student's academic school did not have a significant effect on community satisfaction scores for the complete and modified analyses. One of the fundamental practices of the Division of Student Services at KUMC is that all students are treated equally without regard to their academic program and this has been evidenced on recent

administrations of surveys. The non-significant result is more than likely due to the overall favorable recent results of the survey that has been mentioned previously. During the early years of this survey, it was common for many of the items to have significant differences between student academic programs especially medical students. These differences have almost become non-existent in recent administrations of the survey (Meiers & Gove, 2009) which suggests that students don't feel as marginalized as their predecessors. One area of further research would be to use the survey responses from earlier administrations of the survey and convert the raw scores to the new community satisfaction score where individual item differences were present. Detecting differences in earlier administrations would provide more validity evidence that the survey reliably measures student satisfaction.

Conclusion

It has been well established that environmental assessment is by far the most difficult and challenging aspects in the practice of student affairs. The development of assessment tools that measure student satisfaction in addition to student engagement are warranted to assist student affairs practitioners in improving overall campus life. Even though this current research's finding provided conflicting information about the validity of a graded response model for the KUMC Student Services Survey, it does not suggest that the instrument is flawed. It is possible that some of the unexpected validity evidence is due to using health professional students for the scaling and the graded response model can very well be an adequate model with further research.

Another aspect that has been discussed in this study is the need for assessment techniques specifically related to measuring student satisfaction with health professional students. One limitation of the current survey in regards to generalizing the results to health professional education is that none of the items are specifically focused towards the unique nature of health professional education. Items specifically related to academic medicine should be explored for further administrations of the survey. The National Survey on Student Engagement has done similar modifications to its survey for specific groups of students such as law students and buy-in from health accreditation bodies about the utility of the KUMC Student Services Survey for measuring student satisfaction of health professionals would be easier if there were specific items that could be validly proven to be focused specifically to health professional students.

KUMC has been able to take the results of the survey and improve student life through program and facility improvements over the last 17 years. Assessment activities are only useful if substantive actions are taken from the findings and the Division of Student Services at KUMC has proven that by actively engaging administration and students about the survey results, real changes can occur on a campus. Applying more sophisticated statistical techniques usually reserved to achievement testing to student development surveys like the ones demonstrated in this study are needed to help advance the science of assessment in student affairs.

REFERENCES

About ACHA-NCHA. (n.d.). Retrieved March 15, 2010, from http://www.acha-ncha.org/overview.html

About the national survey of student engangement. (n.d.). Retrieved February 27, 2010, from http://nsse.iub.edu/html/about.cfm

About the University of Kansas Medical Center. (n.d.). Retrieved February 15, 2010, from http://www.kumc.edu/Pulse/aboutkumc.html

Aitken, N. D. (1982). College student performance, satisfaction, and retention: spcification and estimation of a structural model. *The Journal of Higher Education, 53*(1), 32-50.

Aldridge, S. & Rowley, J. (1998). Measuring customer satisfaction in higher education. *Quality Assurance in Higher Education, 6*(4), 197-204.

Astin, A. W. (1985). *Achieving educational excellence* (1st ed.). San Francisco: Jossey-Bass.

Astin, A. W. (1991). *Assessment for excellence: the philosophy and practice of assessment and evaluation in higher education*. New York: MacMillian Publishing Company.

Astin, A. W. (1993). *What matters in college?: four critical years revisited* (1st ed.). San Francisco: Jossey-Bass.

Astin, A. W. (1999). Student involvement: a developmental theory for higher education. *Journal of College Student Developoment, 40*(5), 518-529.

Athiyaman, A. (1997). Linking student satisfaction and service quality perceptions: the case of university education. *European Journal of Marketing, 31*(7), 528-540.

Blumenthal, D., Campbell, E. G., & Weissman, J. S. (1997). The social missions of academic health centers. *New England Journal of Medicine, 337*(21), 1550-1553.

Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: does mode matter? *Research in Higher Education, 44*(1), 1-19.

Carnegie Commission on Higher Education. (1974). *The demise of diversity? A comparative profile of eight types of institutions*. Berkeley, CA: Pace, C. R.

Cattell, R. B. (1966). The scree test for the number of behaviors. *Multivariate Behavior, 1*, 245-276.

Cook, T. D., & Campbell, D. T. (1979). Validty. In Peracchio, L., Dunnette, M. D, & Hough, L. M. (Eds.), *Quasi-experimentation, handbook of industrial and organizational psychology.* Vol 1 (2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika, 16,* 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

DeMars, C. E., & Erwin, T. D. (2003). Revising the scale of intellectual development: application of an unfolding model. *Journal of College Student Development, 44*(2), 168-184.

DeMars, C. E. (2005). Type I error rates for PARSCALE's fit index. *Educational and Psychological Measurement 65*(1), 42-50.

Deutskens, E., Ruyter, K., & Wetzels, M. (2006). An assessment of equivalence between online and mail surveys in service research. *Journal of Service Research, 8*(4), 346-355.

Elliot, K. M., & Shin, D. (2002). Student satisfaction: an alternative approach to assessing this important concept. *Journal of Higher Educatin Policy and Management, 24*(2), 197-209.

Embretson, S. E., & Reise, S. (2000). *Psychometric methods*. Mahwah, NJ: Lawrence Erlbaum.

Epstein, J., Klinkenberg, W. D., & McKinley, L. (2001). Insuring sample equivalence across internet and paper-and-pencil assessments. *Computers in Human Behavior, 17*(3), 339-346.

Erwin, L. (2003). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Erwin, L., & Meiers, C. (2005). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Erwin, L., & Meiers, C. (2006). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Pschological Methods, 4*(3), 272-299.

Friedenberg, L. (1995). *Psychological testing: design, analysis, and use*. Needham Heights, MA: Allyn & Bacon.

Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evauation in Higher Education, 24*(4), 365-377.

Green, S. B., and Salkind, N. J. (2004). *Using SPSS for windows and macintosh: analyzing and understanding data* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Helmstadter, G. C. (1964). *Principles of psychological measurement.* Englewood, Cliffs, NJ: Prentice Hall.

Hendrickson, A. E., and White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17*, 65-70.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kapp, G. J. (1979). *College extracurricular activities: who participates and what are the benefits?* Doctoral Dissertation. University of California, Los Angeles.

Kim, M. M. (2001). Institutional effectiveness of women only colleges: cultivating students' desires to influence social conditions. *Journal of Higher Education, 72*, 287-321.

Kline, R. B. (1998). *Principles and practice of structural equation modeling.* New York: Guilford Press.

Knoll, D., & Erwin, L. (2001). Student needs assessment. *NASPA's NetResults*. Retrieved February 15, 2010, from http://www.naspa.org/membership/menu/nr/PrinterFriendly.cfm?id=465

Kuh, G. D. (1991). *Involving colleges : successful approaches to fostering student learning and development outside the classroom* (1st ed.). San Francisco: Jossey-Bass Publishers.

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., & Kennedy, J. (2001). *NSSE technical and norms report.* Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison Wesley.

Meiers, C., & Gove, R. (2007). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Meiers, C., & Gove, R. (2008). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Meiers, C., & Gove, R. (2009). *Student services survey results*. Kansas City, KS: University of Kansas Medical Center, Division of Student Services.

Messick, S. (1995). Validity of psychological assessment: validations of inferences from persons' reposnes and performances as scientific inquiry into score meaning. *American Psychologist,* 50(9), 741-749.

Muraki, E. (1990). Fitting a polytomous item response model to likert-type data. *Applied Psychological Measurement*, *14*(1), 59-71.

Muraki, E., & Bock, R. D. (2003). *PARSCALE: IRT based test scoring and item analysis for rating scale and graded response.* Chicago, IL: Scientific Software International.

National Academy of Sciences. (2003). *Health professions education: bridge to quality*. Washington, DC.

Nasser, F., Benson, J., & Wisenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of factors. *Educational and Psychological Measurement, 62*(3), 397-419.

Nelson, T. A. (1984). *A comparison of selected undergraduate experiences of alumni who financially support their alma mater*. Unpublished doctoral dissertation. Indiana University.

Official headcount enrollment by school and program: fall 2005 to fall 2009. (n.d.). Retrieved February 15, 2010, from http://www2.kumc.edu/aa/ir/reports/students/Fall2009_20th-dayReport.pdf

Pascarella, E. T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research, 50*(4), 545-595.

Pascarella, E. T., Ethington, C. A., & Smart, J. C. (1988). The influence of college on humanitarian/civic inolvement values. *Journal of Higher Education, 59*(4), 412-437.

Pascarella, E. T., and Terenzini, P. T. (1991). *How college affects students: findings and insights from twenty years of research*. San Francisco: Jossey-Bass.

Quick facts about NSSE. (n.d.). Retrieved February 27, 2010, from http://nsse.iub.edu/ html/quick_facts.cfm

Rennie, L. J. (1982). Detecting a response set to likert-style attitude items with the rating model. *Education Resarch and Perspectives 9*(1), 114-118

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph* (Vol. 17). Iowa City, IA.

Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph* (Vol. 18). Iowa City, IA.

Samejima, F. (1997). Graded reponse model . In van der Linden, W. J., & Hambleton, R. K. (Eds.), *Handbook of modern item response theory*. New York, Springer.

Schmidt, P. (2009, November 7). Researcher harpoons the 'nessie' survey of students [Electronic version]. *Chronicle of Higher Education*.

Schuh, J. H., & Laverty, M. (1983). The perceived long-term influence of holding a significant student leadership position. *Journal of College Student Personnel, 24*(1), 28-32.

Shouping, H., & Kuh, G. D. (2003). Maximizing what students get out of college: testing a learning productivity model. *Journal of College Student Developoment, 44*(2), 185-203.

Tabachnick, B. G., and Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.

Thissen, D., & Steinberg, L. (1986). Taxonomy of item response models. *Psychometrika, 51*, 567-577.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: understanding concepts and applications*. Washington, DC: American Psychological Association.

Thomspon, B. (1993). Calculation of standardized, noncentered factor scores: An alternative to conventional factor scores. *Perceptual and Motor Skills, 77*, 1128-1130.

Upcraft, M. L., & Schuh, J. H. (2000a). Assessment in student affairs. In Barr, M. J., Desler, M.K., & Associates (Eds.), *The handbook of student affairs administration* (2nd ed.). San Francisco: Jossey-Bass.

Upcraft, M. L., & Schuh, J. H. (2000b). Measuring student satisfaction and needs. In Barr, M. J., Desler, M.K., & Associates (Eds.), *The handbook of student affairs administration* (2nd ed.). San Francisco: Jossey-Bass.

Wiers-Jenssen, J., Stensaker, B., & Grøgaard, J. B. Student satisfaction: towards an empirical deconstruction of the concept. *Quality in Higher Education, 8*(2), 183-195.

Zoski, K. W., & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: the standard error scree. *Educational and Psychological Measurement, 56*, 443-451.

APPENDICES

Appendix A

2008-2009 KUMC Student Services Survey

KU Medical Center 2008-09 Student Services Survey

The University of Kansas Medical Center is interested in learning about your experiences as a KUMC student. Please fill in the ovals that indicate the appropriate answers. Please leave blank any questions about services you have not personally used, as well as any questions that are not applicable to you. All responses are strictly confidential and will be analyzed as grouped data only. Comments about any aspect of your educational experience at KUMC are welcomed; please write them on the back of any of these pages. *Note: Please complete only one questionnaire per person.*

KUMC CAMPUS ENVIRONMENT

For each of the following statements, please fill in TWO answers: ONE answer to the RIGHT, that indicates how strongly you agree or disagree with the statement; and ONE answer to the LEFT, that indicates how important or unimportant it is to you that the statement is TRUE.

How important or unimportant is it to you that the statement is true? Chose ONE answer from this column for each statement.

How strongly do you agree or disagree with each of the following statements? Chose ONE answer from this column for each statement.

Not at all important	Not very important	Neutral	Somewhat important	Very important	STATEMENTS ABOUT KUMC	Disagree Strongly	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree Strongly
0	0	0	0	0	1. The environment at KUMC is friendly and supportive of students.	0	0	0	0	0
0	0	0	0	0	2. Campus security/safety is adequate.	0	0	0	0	0
0	0	0	0	0	3. Students have a voice in decision-making at KUMC.	0	0	0	0	0
0	0	0	0	0	4. KUMC has a positive image in the Kansas City area.	0	0	0	0	0
0	0	0	0	0	5. There is racial/ethnic harmony at KUMC.	0	0	0	0	0
0	0	0	0	0	6. Students at KUMC are free from sexual harassment.	0	0	0	0	0
0	0	0	0	0	7. Male and female students are treated equally at KUMC.	0	0	0	0	0
0	0	0	0	0	8. There is a feeling of community at KUMC.	0	0	0	0	0
0	0	0	0	0	9. Students have a sense of pride at KUMC.	0	0	0	0	0
0	0	0	0	0	10. It is easy to approach faculty at KUMC.	0	0	0	0	0
0	0	0	0	0	11. It is easy to approach staff at KUMC.	0	0	0	0	0
0	0	0	0	0	12. KUMC is committed to resolving diversity-related problems.	0	0	0	0	0
0	0	0	0	0	13. I am comfortable raising diversity related problems to faculty, administration and staff.	0	0	0	0	0
0	0	0	0	0	14. Students are treated the same regardless of race.	0	0	0	0	0
0	0	0	0	0	15. Students are treated the same regardless of color.	0	0	0	0	0
0	0	0	0	0	16. Students are treated the same regardless of creed.					
0	0	0	0	0	17. Students are treated the same regardless of religion.	0	0	0	0	0
0	0	0	0	0	18. Students are treated the same regardless of national origin.	0	0	0	0	0
0	0	0	0	0	19. Students are treated the same regardless of ancestry.	0	0	0	0	0
0	0	0	0	0	20. Students are treated the same regardless of gender.	0	0	0	0	0
0	0	0	0	0	21. Students are treated the same regardless of age.	0	0	0	0	0
0	0	0	0	0	22. Students are treated the same regardless of sexual orientation.	0	0	0	0	0
0	0	0	0	0	23. Students are treated the same regardless of disability.	0	0	0	0	0
0	0	0	0	0	24. Students are treated the same regardless of veteran status.	0	0	0	0	0

	Disagree Strongly	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree Strongly
25. It has been beneficial for me to be a member of a student organization while at KUMC.	0	0	0	0	0

STUDENT WELLNESS

26. The Student Wellness Program provides services and programs that meet my needs.	0	0	0	0	0
27. Drug use is a problem for students in my program at KUMC.	0	0	0	0	0
28. Alcohol use is a problem for students in my program at KUMC.	0	0	0	0	0
29. Drug use is a problem for me.	0	0	0	0	0
30. Alcohol use is a problem for me.	0	0	0	0	0

31. Can you generally find a satisfactory location to study on campus? 0 Yes 0 No 0 I don't predominately study on campus

32. Please comment about study space on campus:

33. Do you use the campus escort system? 0 Yes 0 No

34. If no, why?
    - 0 Unaware of service
    - 0 Don't have need for service
    - 0 Service is inconvenient, please specify:
    - 0 Other, please specify

35. If yes, are you satisfied with the campus escort service? 0 Yes 0 No

36. If no, why?

37. Are there specific places on campus where you don't feel safe?

ABOUT YOURSELF (ALL INFORMATION IS OPTIONAL)

1. Academic Program or Degree Pursuing:

2. What year of study is this toward your current degree?

   0 1st 0 4th 0 7th or more
   0 2nd 0 5th
   0 3rd 0 6th

3. Are you:
   0 Male 0 Female

4. In which ethnic category do you place yourself?
   0 American Indian/ Alaskan Native 0 Hispanic
   0 Asian/Islander 0 White, non-Hispanic
   0 Black, non-Hispanic 0 Multi-racial

5. Please write your age: _____

6. Are you a U.S. Citizen?
   0 Yes 0 No

7. Are you involved in a KUMC student organization?
   0 Yes 0 No

8. This semester, where do you spend most of your "class" time?

   Kansas City 0 Medical Center 0 Off-site location(s)
   Wichita 0 USKM-W 0 Off-site location(s)
   Other city 0 Other (please specify)

   _____

   Internet
   0 Internet

COMMENTS *Please comment below on any aspect of your educational experience at KUMC. Please be as specific as possible so your comments will be fully understood.*

Appendix B

Variable Definitions

DEPENDENT VARIALBES

Student Satisfaction

Astin (1993) defines student satisfaction as including the following components: faculty interest in students; relationships with the administration; degree of institutional priority to issues such as diversity, social change, resource acquisition; and the enhancement of the institutional reputation.

Independent Variables

Student Involvement

Involved Students

Involved students are defined as respondents who indicated they were involved in a student organization at KUMC when they submitted the KUMC Student Services Survey.

Not-Involved Students

Not-involved students are defined as respondents who indicated they were not involved in a student organization at KUMC when they submitted the KUMC Student Services Survey.

Mode of Survey Administration

Online Surveys

Online surveys are defined as the KUMC Student Services Surveys administered via a website constructed with survey software. Students are invited to complete the survey via an e-mail message.

Paper Surveys

Paper surveys are defined as KUMC Student Services Surveys administered via a paper survey mailed to all students at KUMC through the United States Postal Service.

Student Academic School

Student Academic School

School of Allied Students

School of Allied Health students are defined as undergraduate and graduate students enrolled in a School of Allied Health program at KUMC.

School of Medicine Students

School of Medicine students are defined as graduate and medical students enrolled in a School of Medicine program at KUMC.

School of Nursing Students

School of Nursing students are defined as undergraduate and graduate students enrolled in a School of Nursing program at KUMC.

Appendix C

PARSCALE Syntax

Community Syntax

```
COMMUNITY.PSL COMMUNITY ITEMS
8 POLYTOMOUS COMMUNITY ITEMS, FOUR CHARACTER ID FIELD
>FILE    DFNAME='COMMUNITY.DAT', OFNAME='OMIT.DAT' SAVE;
>SAVE    PARM='COMMUNITY.PAR', SCORE='COMMUNITY.SCO';
>INPUT   NIDW=4, NTOTAL=8, NTEST=1, LENGTH=8, SAMPLE=100;
(4A1,8A1)
>TEST   TNAME=MIXED, ITEM=(1(1)8), NBLOCK=8;
>BLOCK01 BNAME=POLY1, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK02 BNAME=POLY2, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK03 BNAME=POLY3, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK04 BNAME=POLY4, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK05 BNAME=POLY5, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK06 BNAME=POLY6, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK07 BNAME=POLY7, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK08 BNAME=POLY8, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>CALIB   GRADED, LOGISTIC, SCALE=1.7, NQPTS=15,
CYCLES=(1000,100,100,100,100),
         NEWTON=100, CRIT=0.001, ITEMFIT=15, RIDGE=(0.01,1);
>SCORE   EAP, NQPTS=15;
```

Diversity Syntax

```
DIVERSITY.PSL DIVRSITY ITEMS
10 POLYTOMOUS DIVERSITY ITEMS, FOUR CHARACTER ID FIELD
>FILE    DFNAME='DIVERSITY.DAT', OFNAME='OMIT.DAT' SAVE;
>SAVE    PARM='DIVERSITY.PAR', SCORE='DIVERSITY.SCO';
>INPUT   NIDW=4, NTOTAL=10, NTEST=1, LENGTH=10, SAMPLE=100;
(4A1,10A1)
>TEST   TNAME=MIXED, ITEM=(1(1)10), NBLOCK=10;
>BLOCK01 BNAME=POLY1, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK02 BNAME=POLY2, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK03 BNAME=POLY3, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK04 BNAME=POLY4, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK05 BNAME=POLY5, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK06 BNAME=POLY6, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK07 BNAME=POLY7, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK08 BNAME=POLY8, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK09 BNAME=POLY9, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK10 BNAME=POLY10, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>CALIB   GRADED, LOGISTIC, SCALE=1, NQPTS=15,
CYCLES=(1000,100,100,100,100),
         NEWTON=100, CRIT=0.001, ITEMFIT=15, RIDGE=(0.01,1);
>SCORE   EAP, NQPTS=15;
```

Self Syntax

```
SELF.PSL SELF ITEMS
4 POLYTOMOUS SELF ITEMS, FOUR CHARACTER ID FIELD
>FILE    DFNAME='SELF.DAT', OFNAME='OMIT.DAT' SAVE;
>SAVE    PARM='SELF.PAR', SCORE='SELF.SCO';
>INPUT   NIDW=4, NTOTAL=4, NTEST=1, LENGTH=4, SAMPLE=100;
(4A1,4A1)
>TEST   TNAME=MIXED, ITEM=(1(1)4), NBLOCK=4;
>BLOCK01 BNAME=POLY1, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK02 BNAME=POLY2, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK03 BNAME=POLY3, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>BLOCK04 BNAME=POLY4, NITEMS=1, NCAT=5, ORIGINAL=(1,2,3,4,5),
MODIFIED=(1,2,3,4,5);
>CALIB   GRADED, LOGISTIC, SCALE=1, NQPTS=15,
CYCLES=(1000,100,100,100,100),
         NEWTON=100, CRIT=0.001, ITEMFIT=15, RIDGE=(0.01,1);
>SCORE   EAP, NQPTS=15;
```

Appendix D

Excel Formula to Locate Community Satisfaction Score

The following formula was used to locate the precise location of the community satisfaction score for a given integer level raw score on the raw score conversion chart. M2 refers to the integer-level raw score that is being converted. A2-A37 refer to the actual raw scores and B2-B37 refer to the actual community satisfactions scores that were used to create the conversion chart.

```
=INDEX($A$2:$B$37,MATCH(M2,$A$2:$A$37,1),2)+(M2-
INDEX($A$2:$B$37,MATCH(M2,$A$2:$A$37,1),1))*(INDEX($A$2:$B$37,
MATCH($M2,$A$2:$A$37,1)+1,2)-INDEX($A$2:$B$37,
MATCH($M2,$A$2:$A$37,1),2))/(INDEX($A$2:$B$37,
MATCH($M2,$A$2:$A$37,1)+1,1)-
INDEX($A$2:$B$37,MATCH($M2,$A$2:$A$37,1),1))
```

Appendix E

Diversity Subscale Item Characteristic, Item Information, and Total Information Curves

Item Characteristic Curves

Due to the difficulties of getting adequate estimations of satisfaction, the actual curves that are provided by the PARSCALE software are presented in this appendix. The X-axis represents the latent trait of student satisfaction and the Y-axis represents the probability of endorsing the response category.

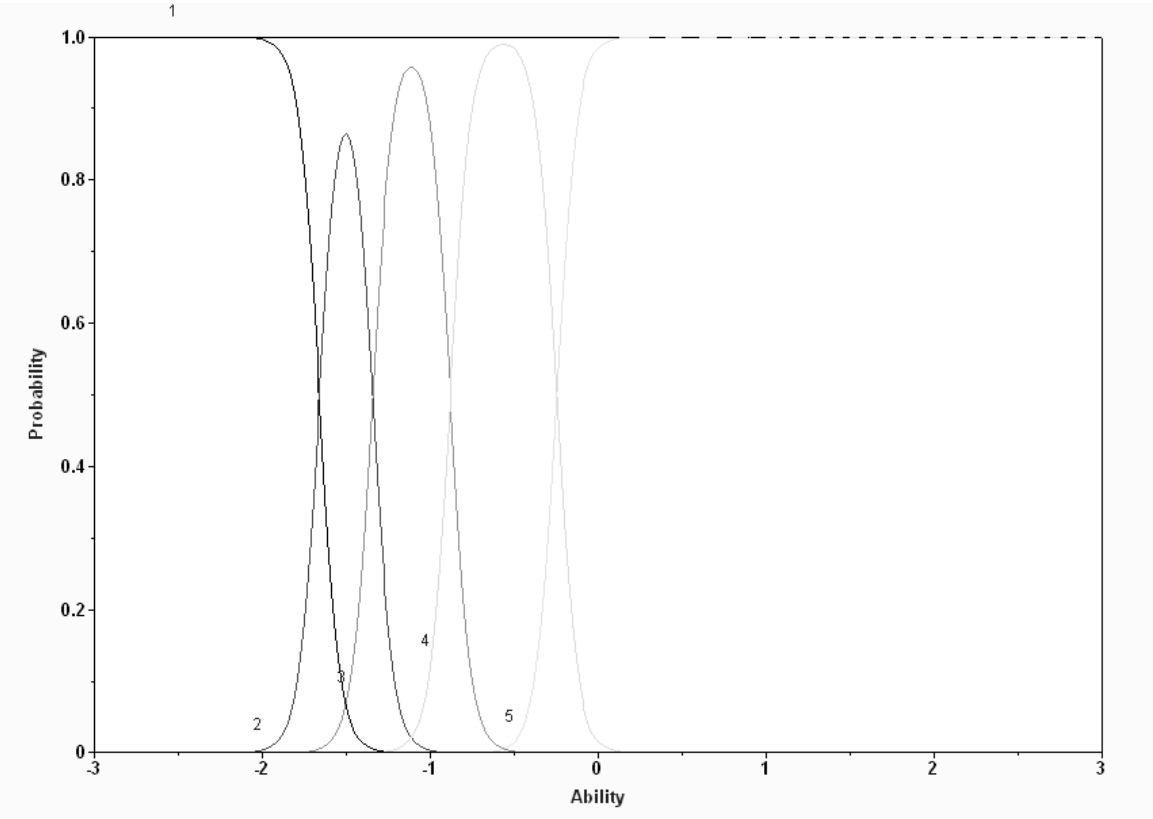

Figure 40: D1 Students are treated the same regardless of race item characteristic curve.

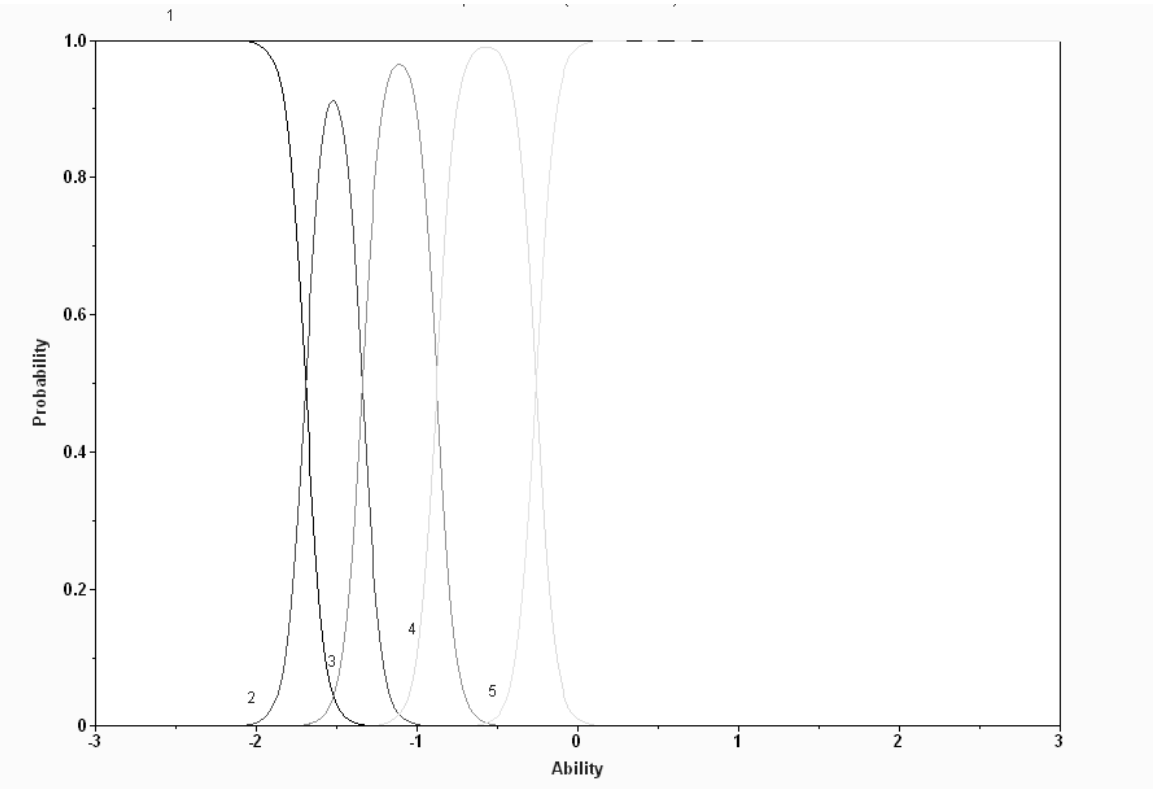

Figure 41: D2 Students are treated the same regardless of color item characteristic curve.

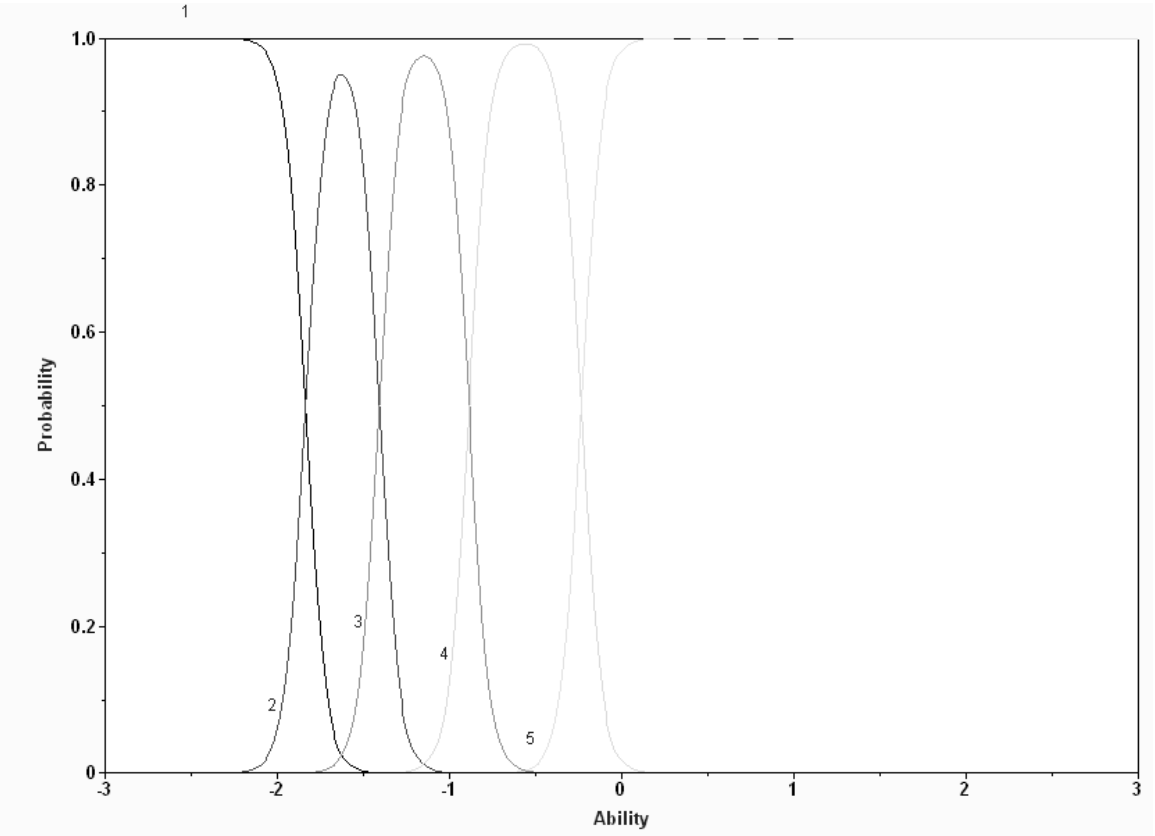

Figure 42: D3 Students are treated the same regardless of creed item characteristic curve.

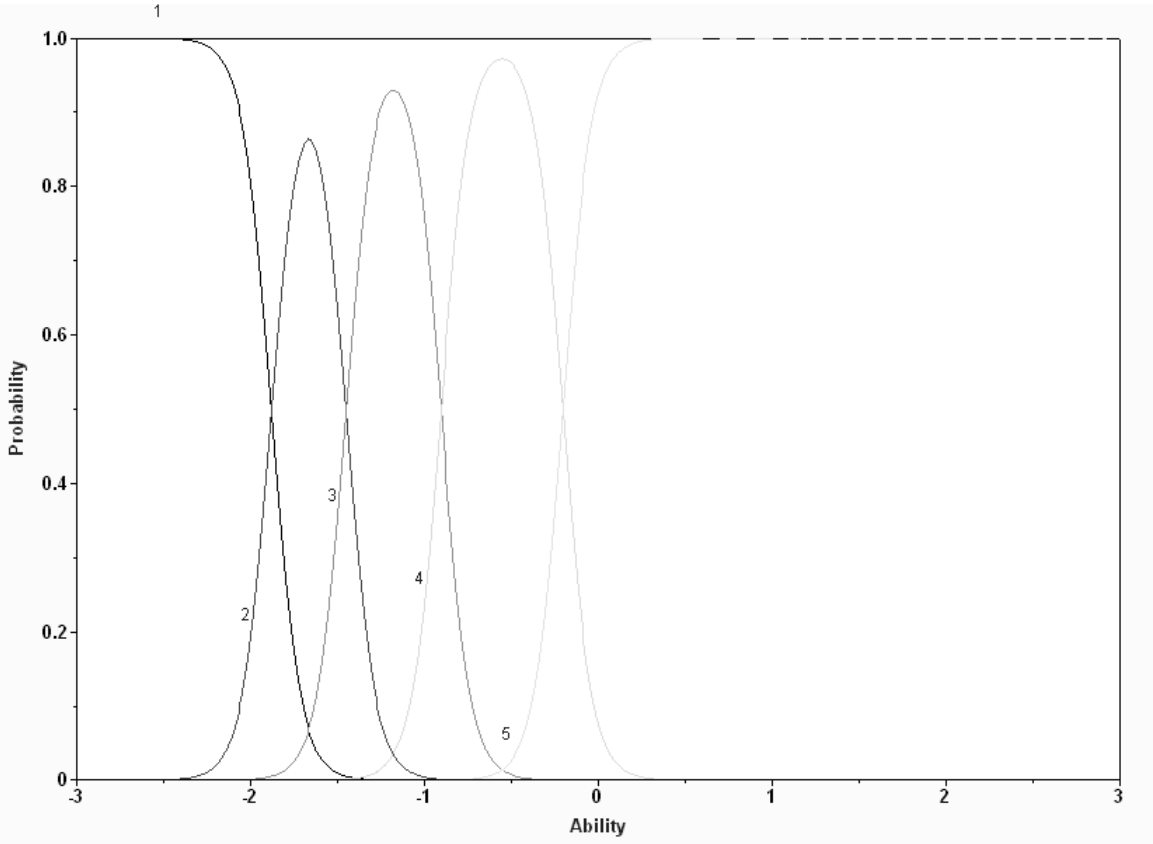

Figure 43: D4 Students are treated the same regardless of religion item characteristic curve.

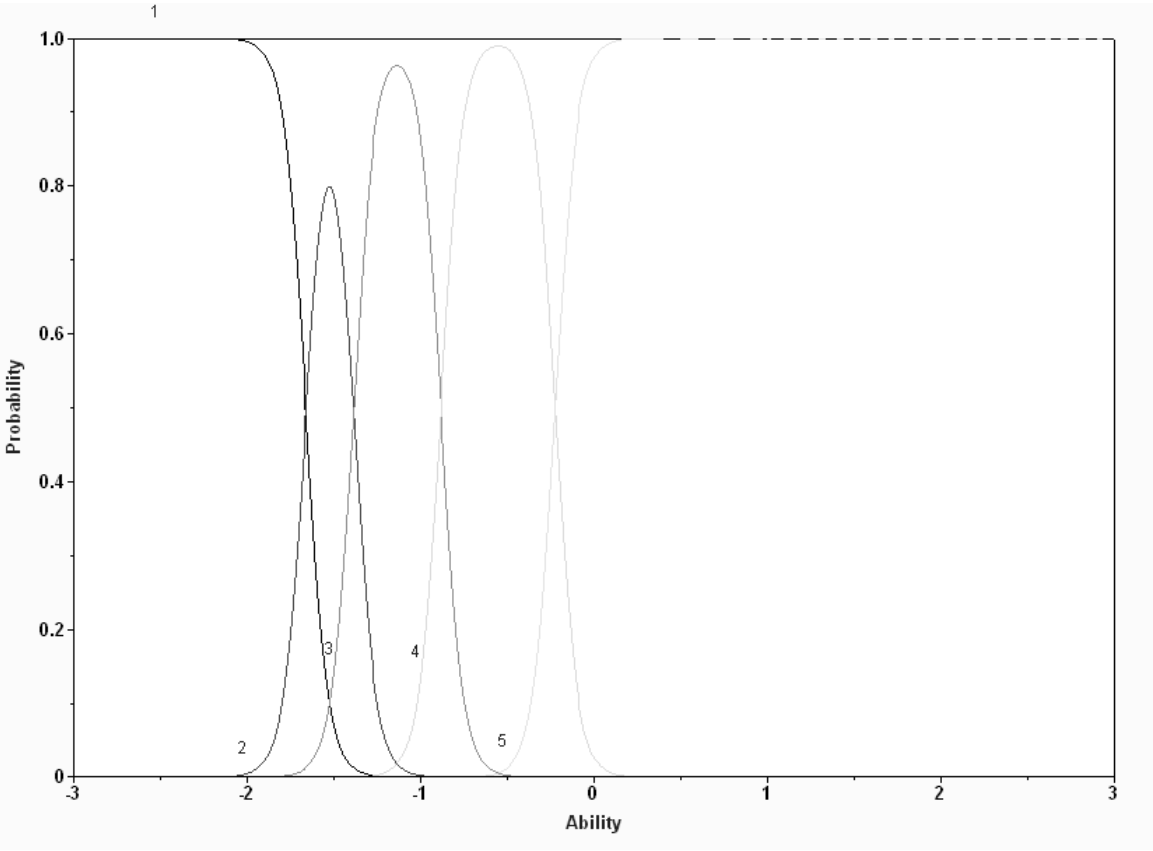

Figure 44: D5 Students are treated the same regardless of national origin item characteristic curve.

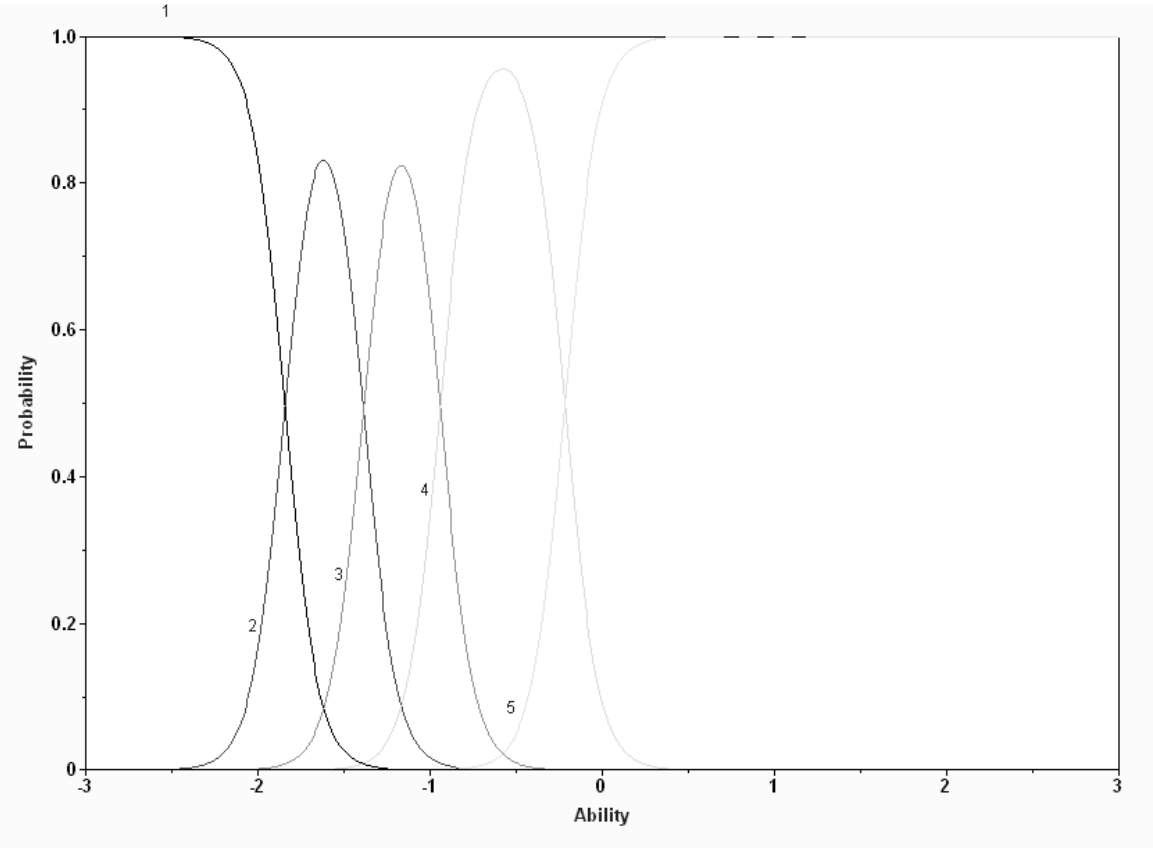

Figure 45: D7 Students are treated the same regardless of gender item characteristic curve.

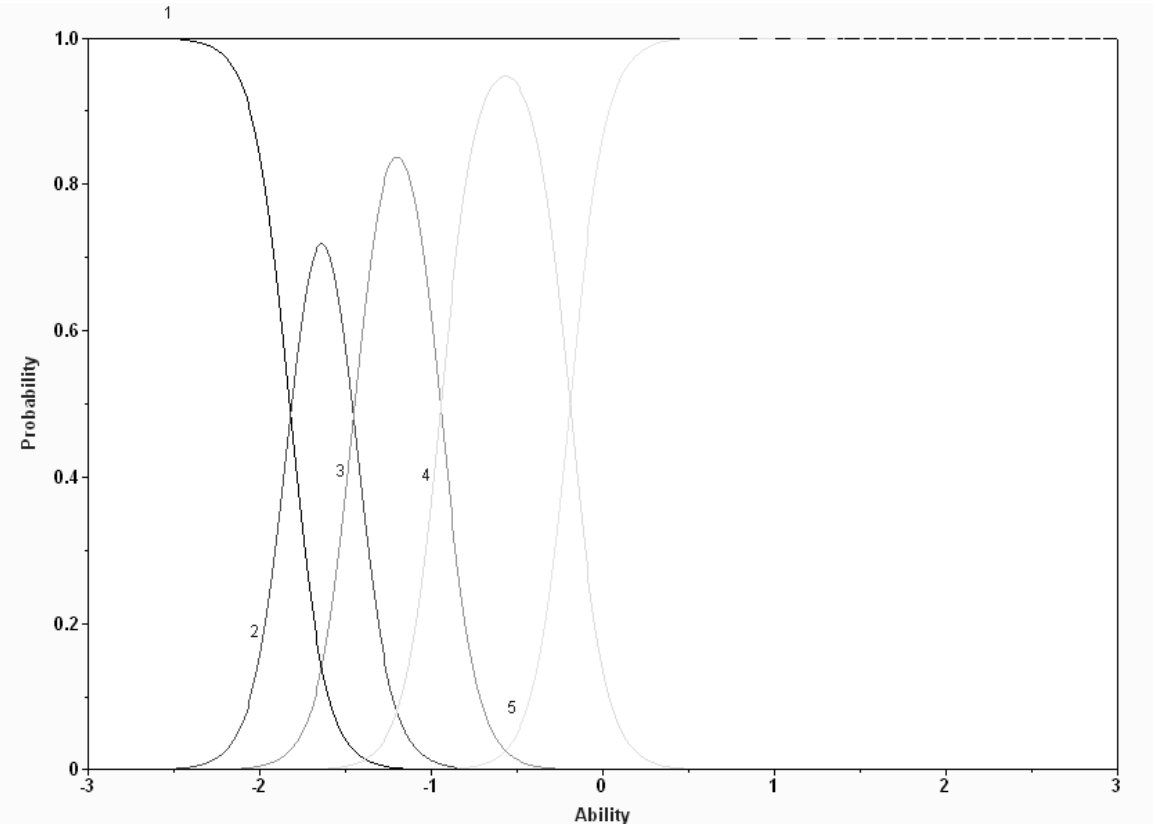

Figure 46: D8 Students are treated the same regardless of age item characteristic curve.

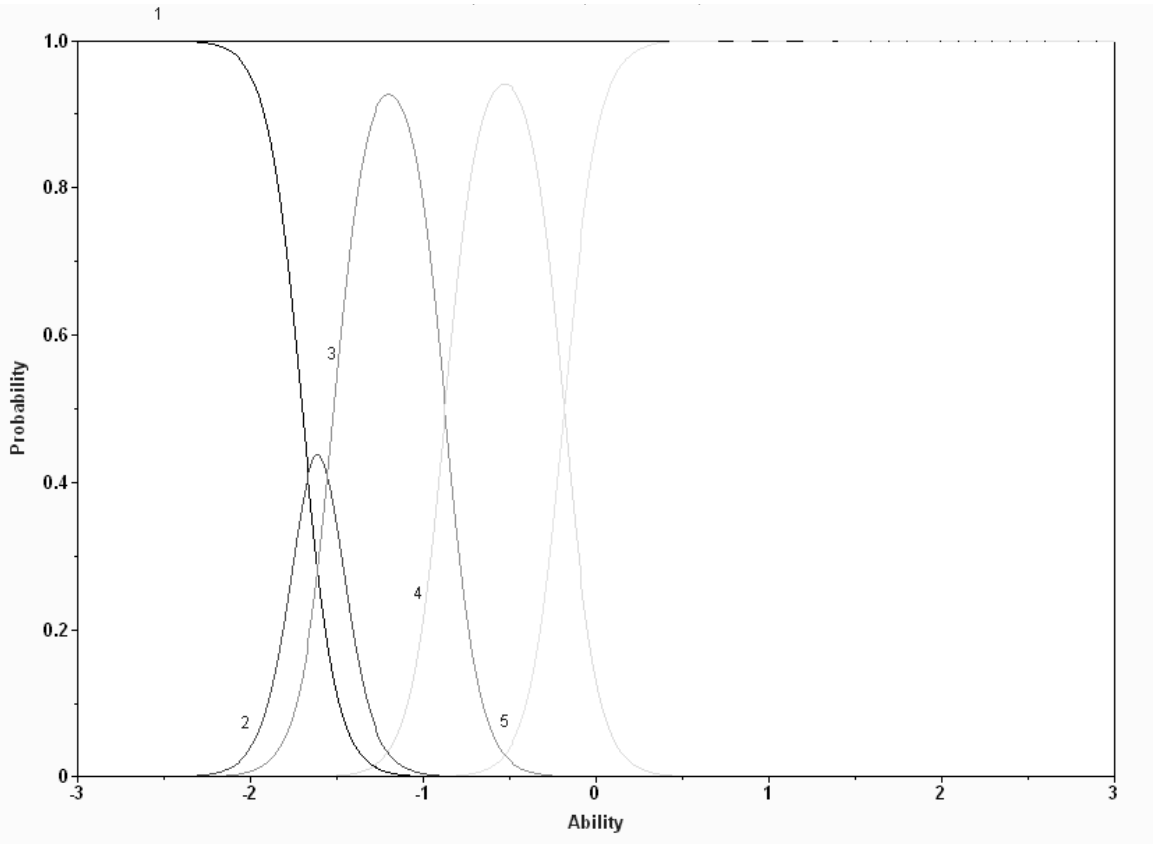

Figure 47: D9 Students are treated the same regardless of sexual orientation item characteristic curve.

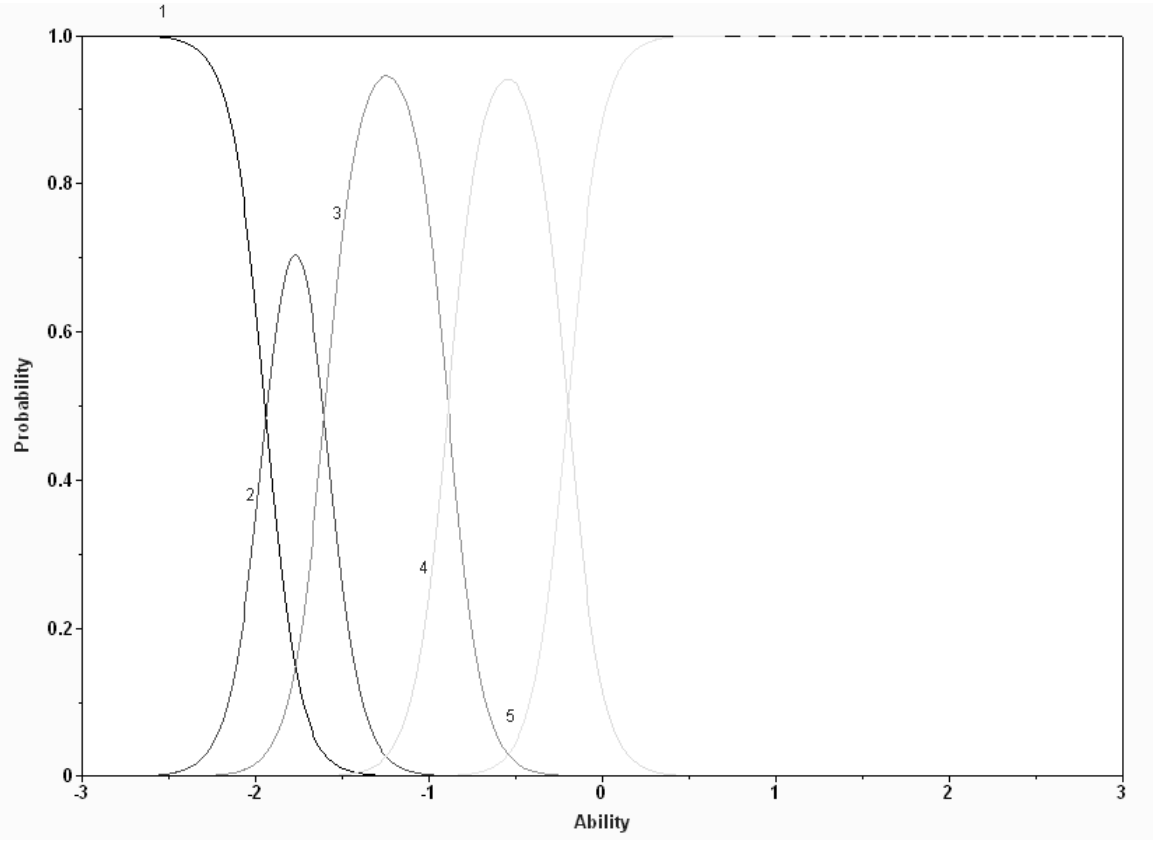

Figure 48: D10 Students are treated the same regardless of disability item characteristic curve.

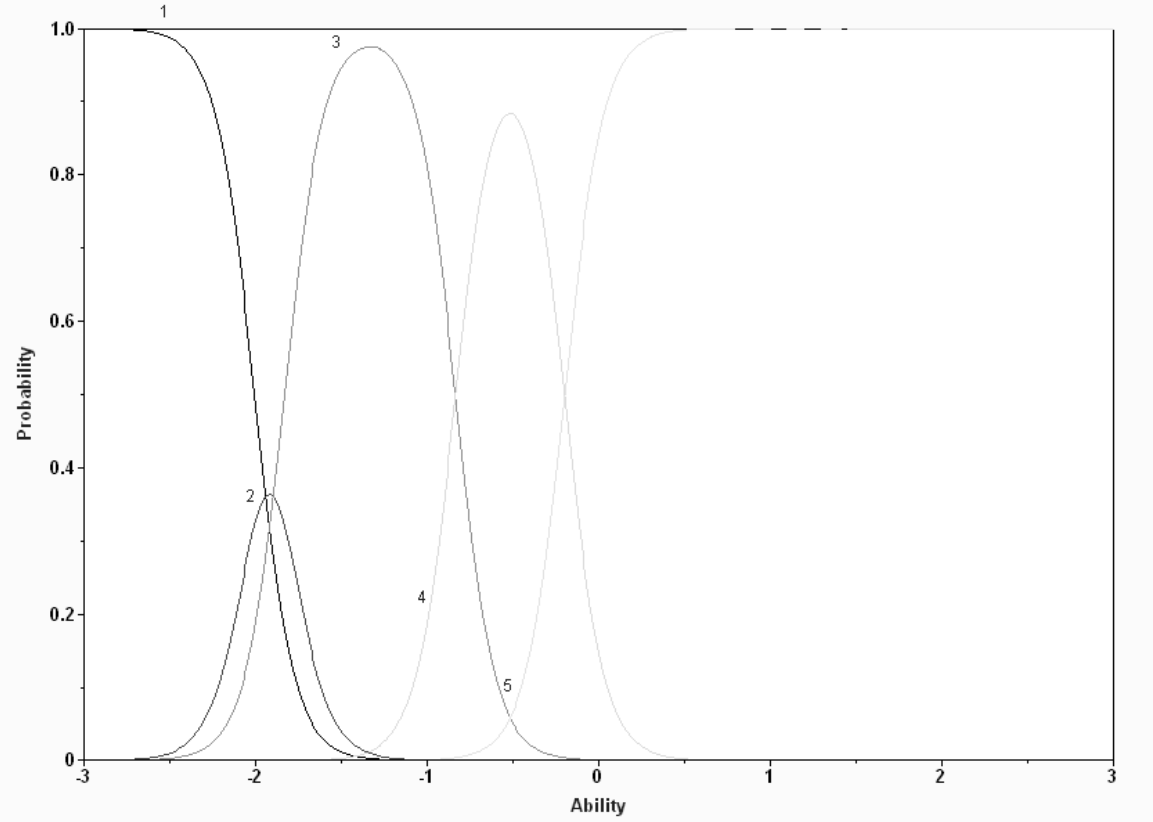

Figure 49: D11 Students are treated the same regardless of veteran status item characteristic curve.

Item Information Curves

Due to the difficulties of getting adequate estimations of satisfaction, the actual curves that are provided by the PARSCALE software are presented in this appendix. In order to produce information curves in PARSCALE, the values that are used to plot the information curve had to be manipulated and changed to 0 where the amount of information that was being estimated had a NaN or Infinity value. The X-axis represents the latent trait of student satisfaction for the theta level and the Y-axis represents the amount of information provided by the item.

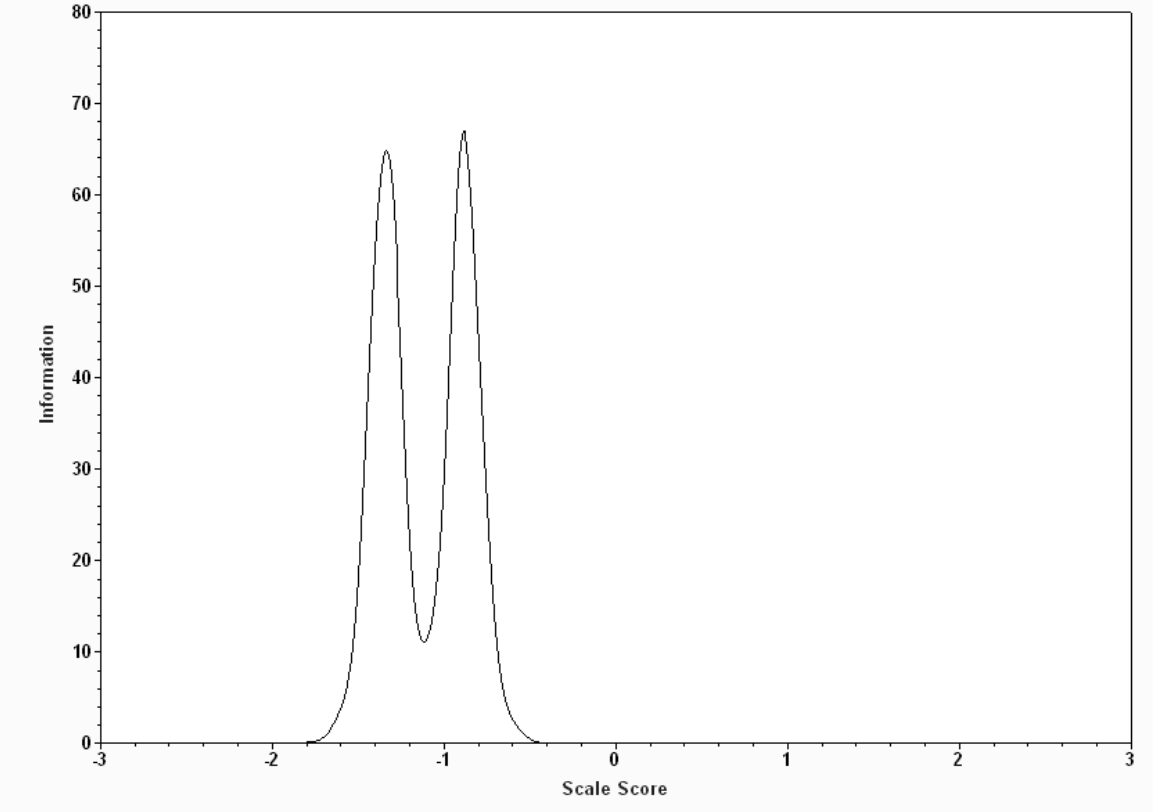

Figure 50: D1 Students are treated the same regardless of race item information curve.

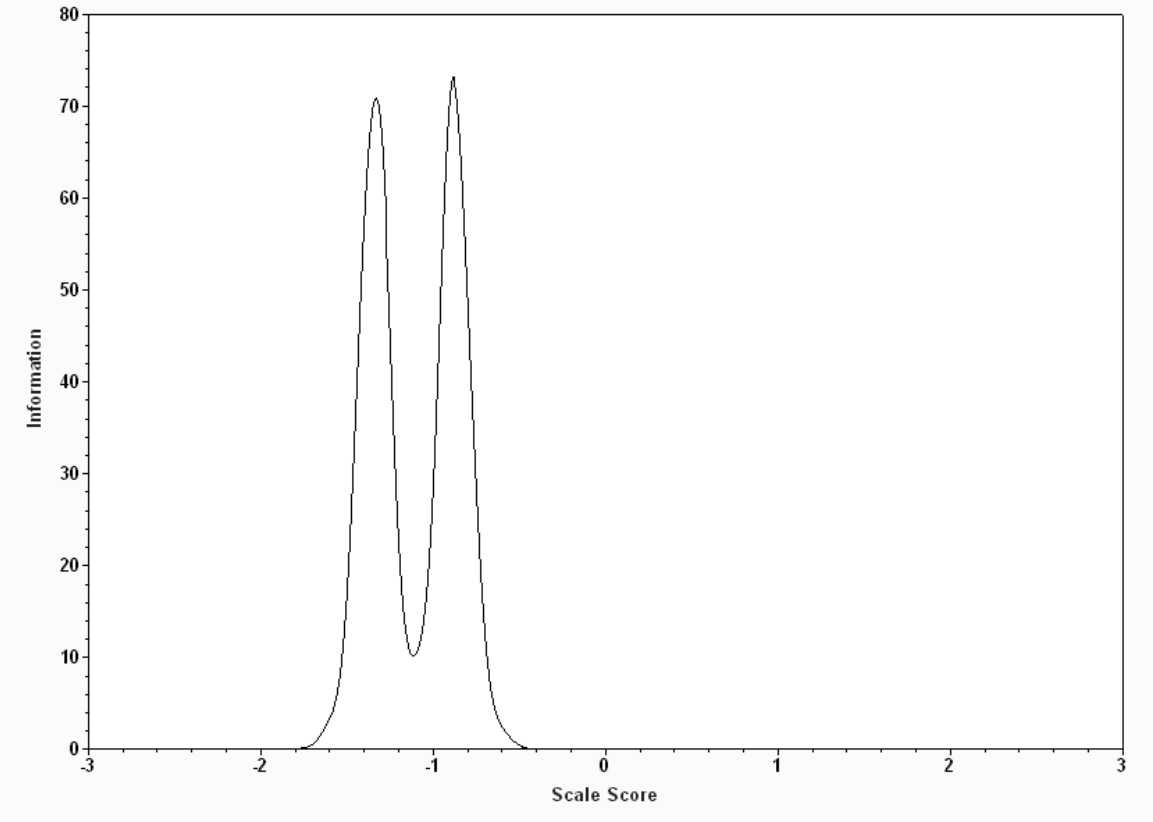

Figure 51: D2 Students are treated the same regardless of color item information curve.

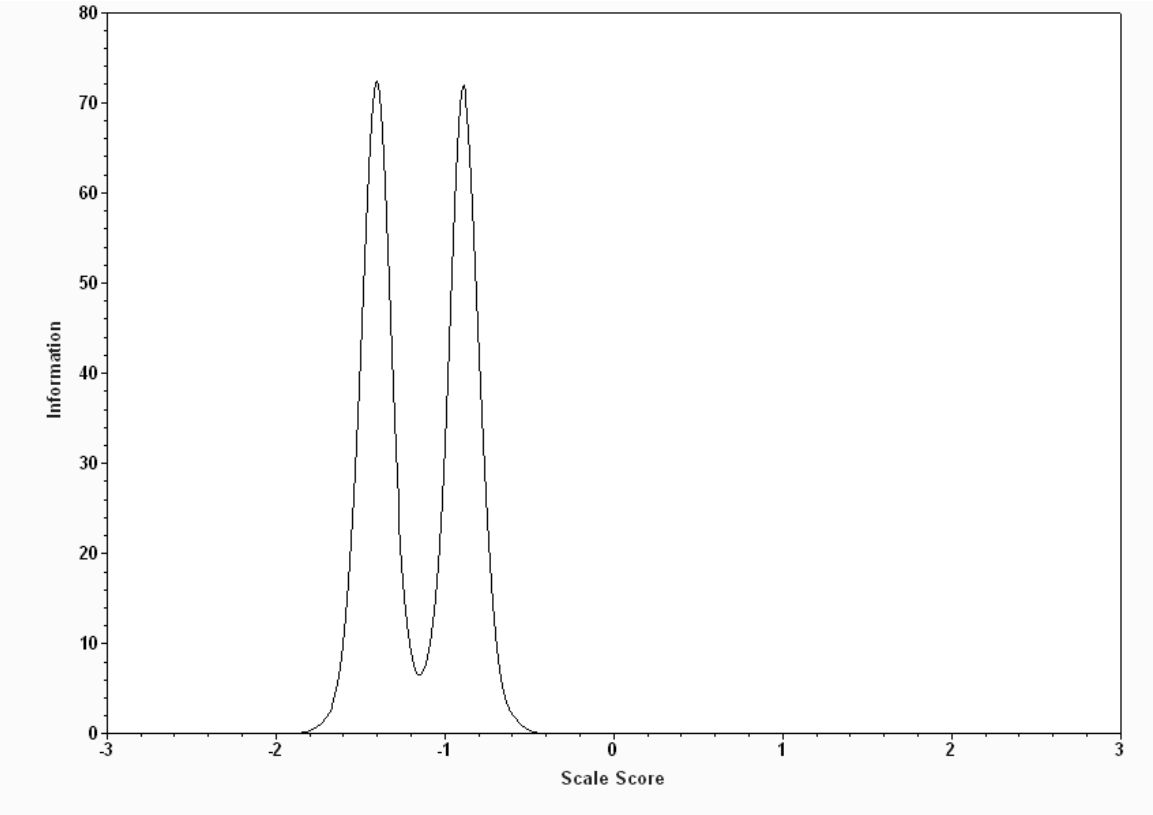

Figure 52: D3 Students are treated the same regardless of creed item information curve.

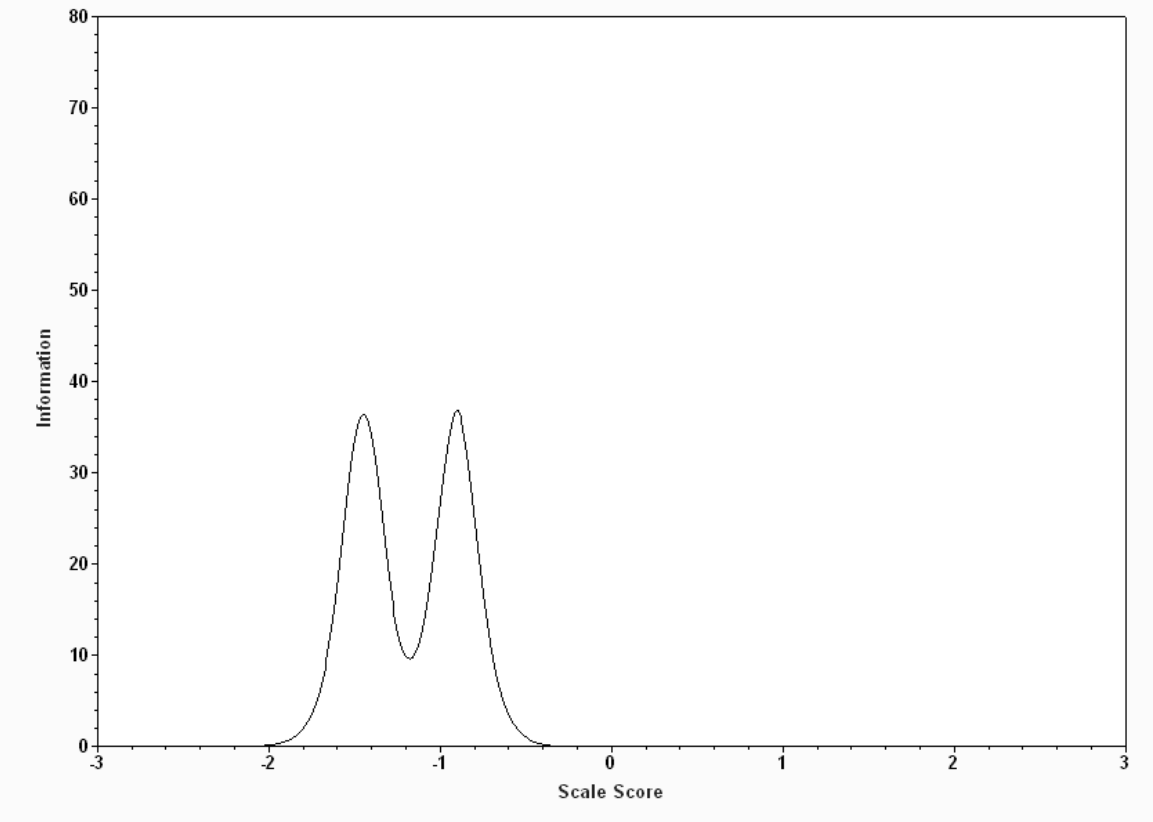

Figure 53: D4 Students are treated the same regardless of religion item information curve.

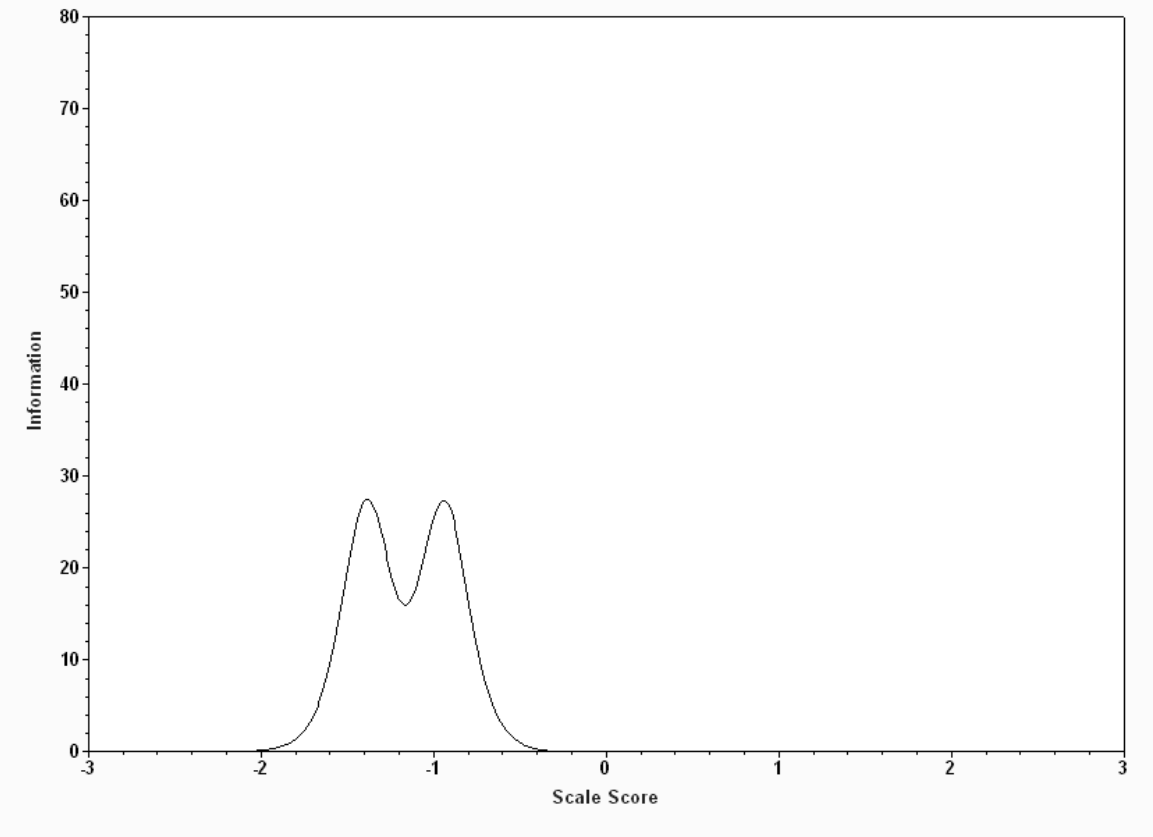

Figure 54: D5 Students are treated the same regardless of national origin item information curve.

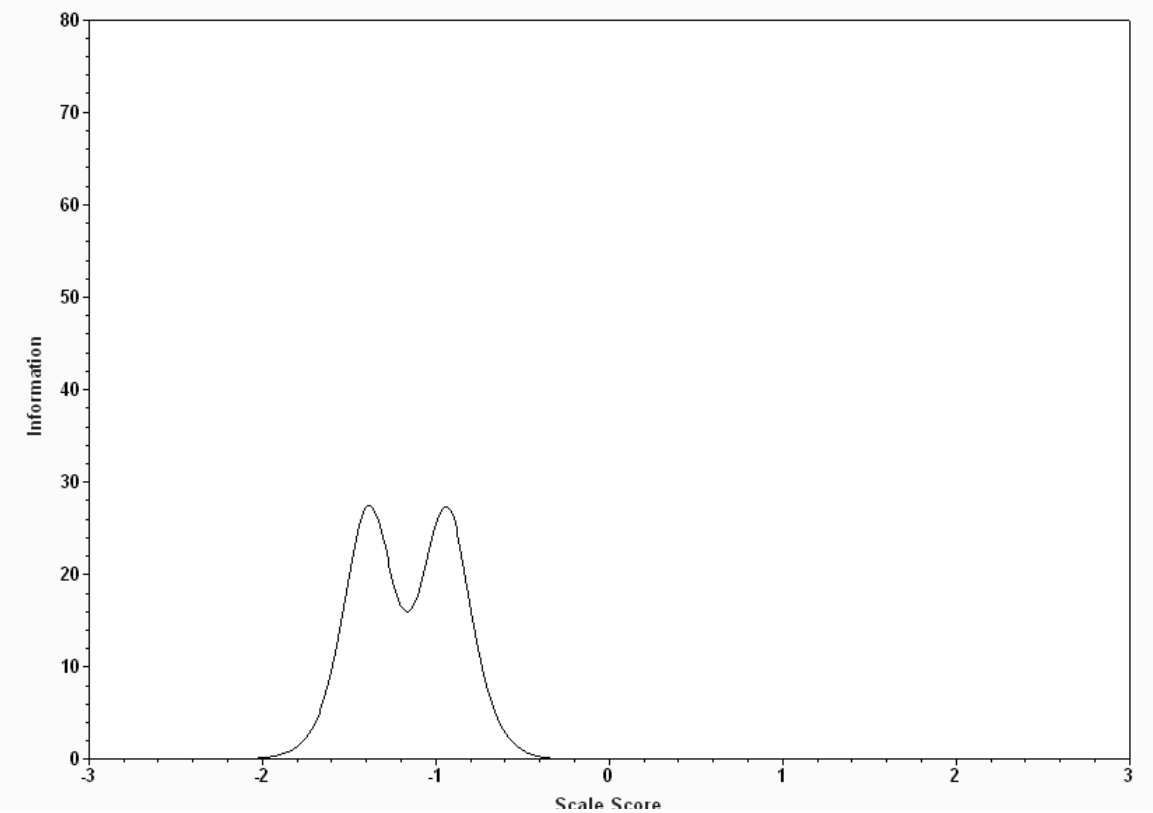

Figure 55: D7 Students are treated the same regardless of gender item information curve.

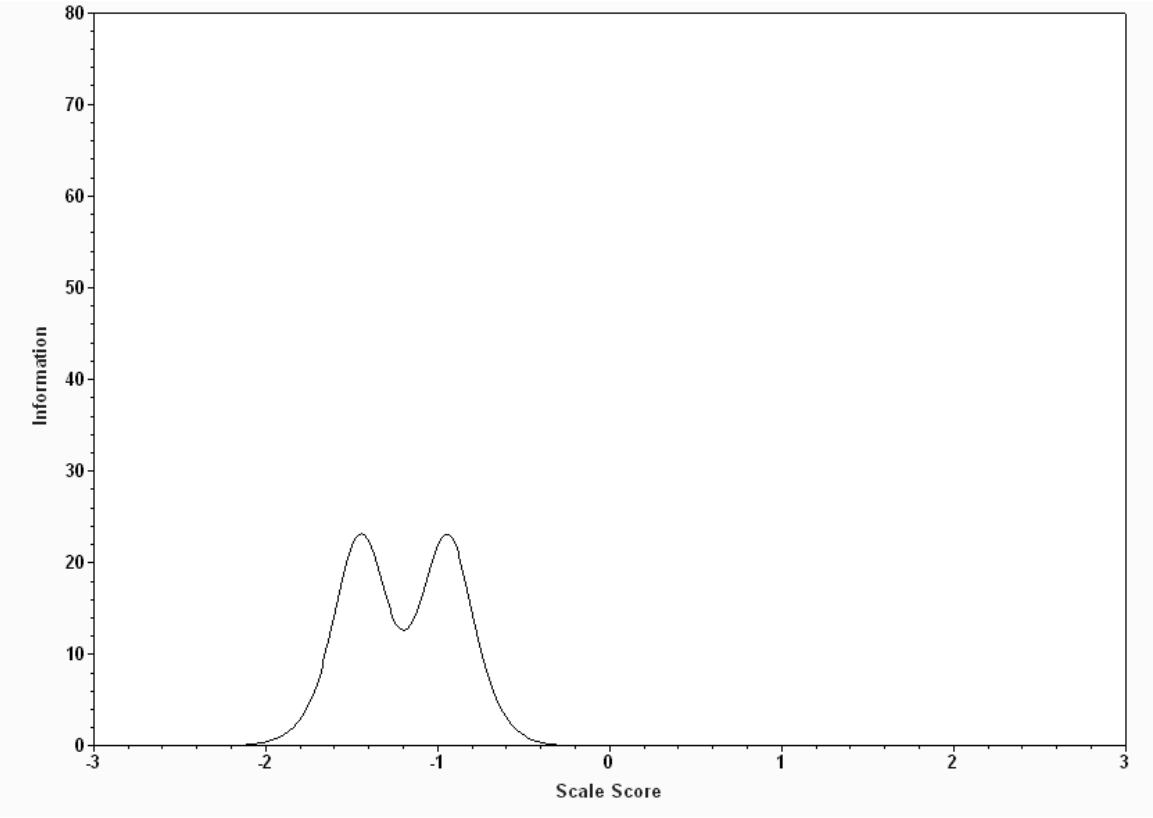

Figure 56: D8 Students are treated the same regardless of age item information curve.

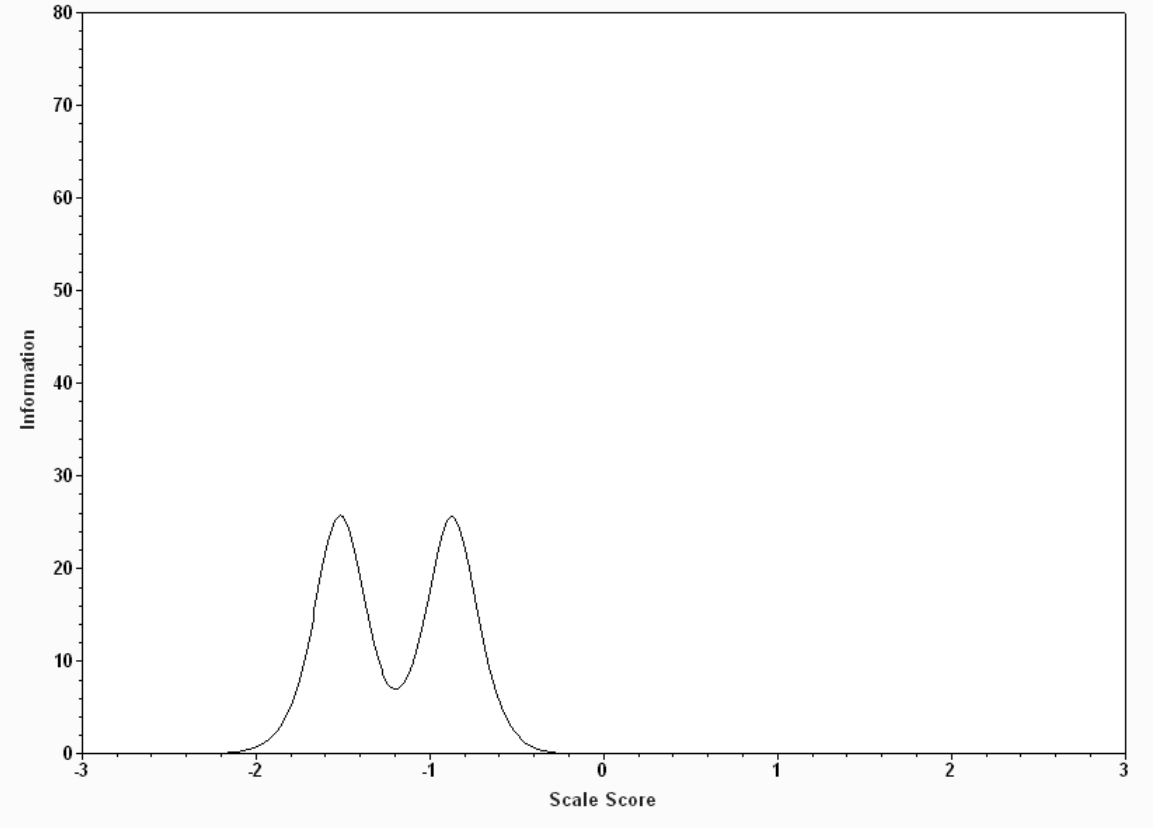

Figure 57: D9 Students are treated the same regardless of sexual orientation item information curve.

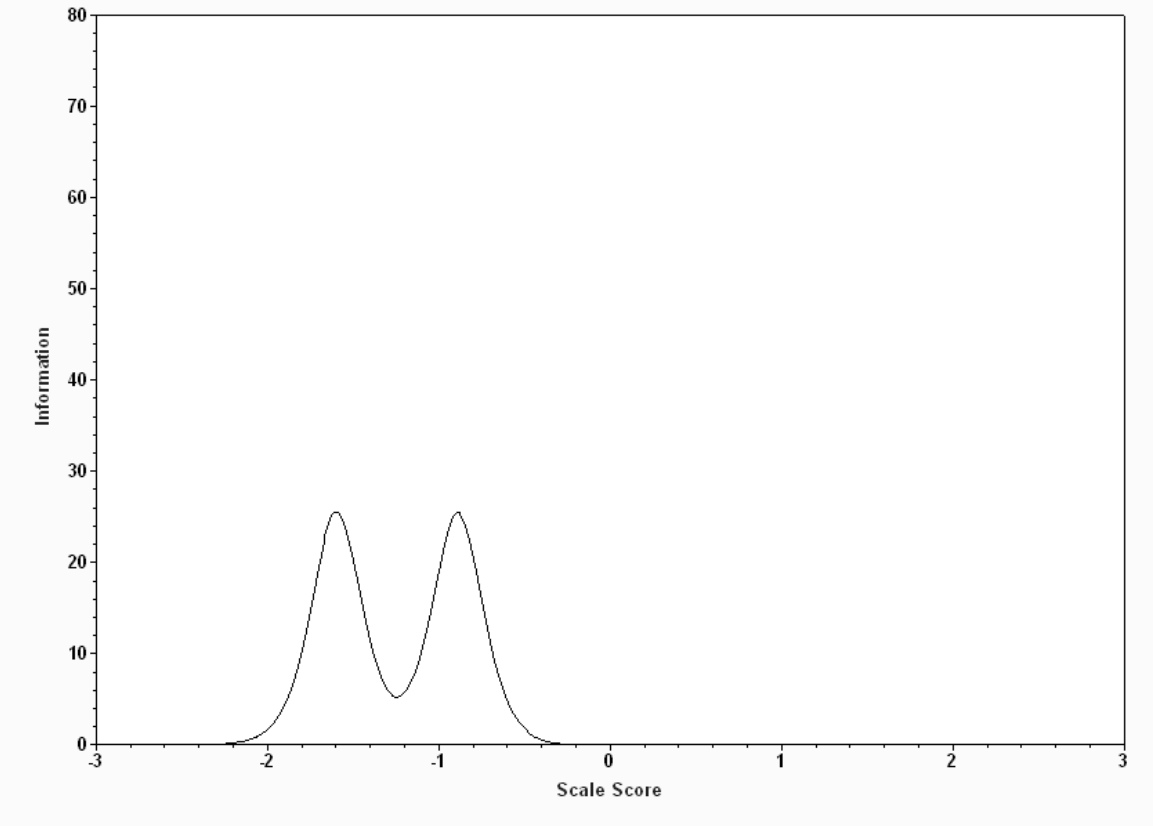

Figure 58: D10 Students are treated the same regardless of disability item information curve.

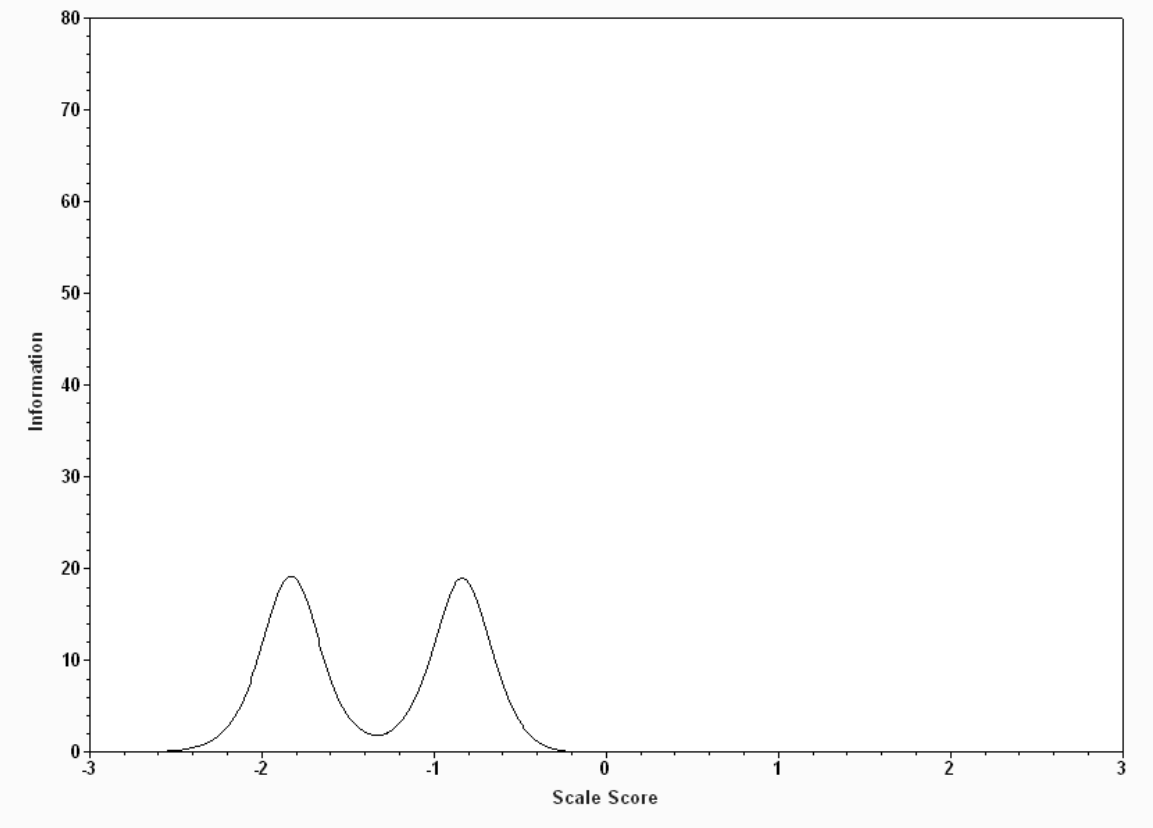

Figure 59: D11 Students are treated the same regardless of veteran status item information curve.

Total Information Curve

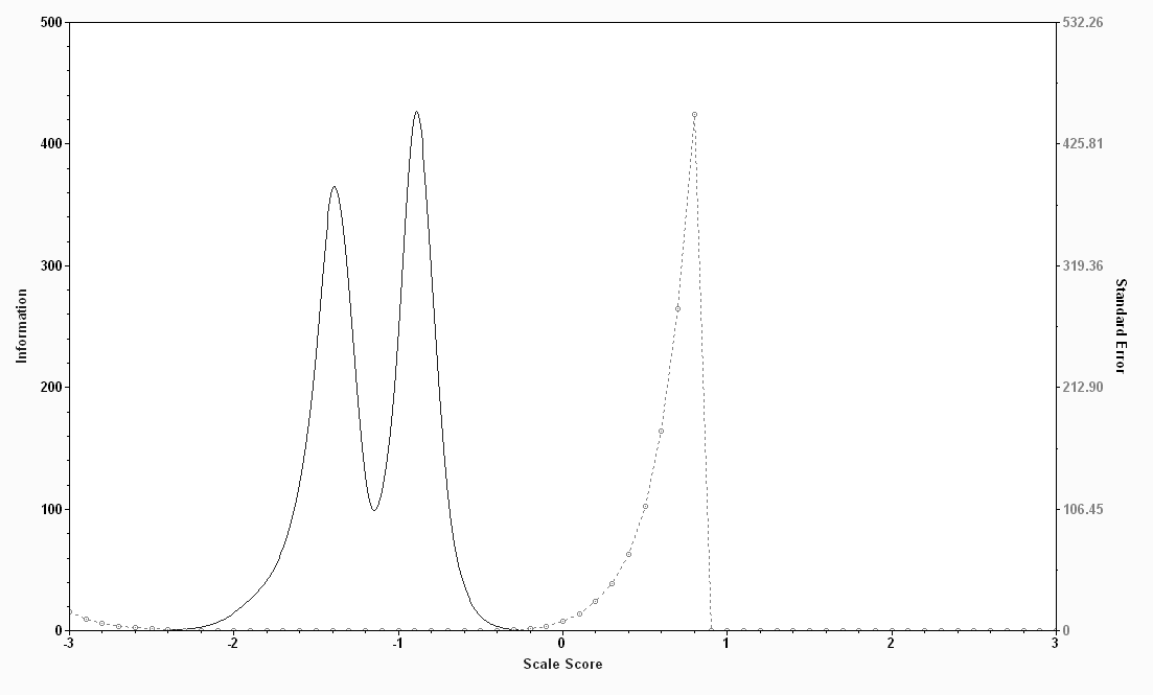

Figure 60: Total information and standard error for the ten diversity items.