RASCH ANALYSIS OF THE EATING ATTITUDES TEST:
AN EXPLORATORY AND CONFIRMATORY STUDY OF RATING SCALE
CATEGORY EFFECTIVENESS, AND CALIBRATION OF A NEW RATING SCALE

by

Amanda R. Cook

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
PhD in Health and Human Performance

Middle Tennessee State University
August 2016

Dissertation Committee:

Dr. Norman L. Weatherby, Chair

Dr. Minsoo Kang

Dr. Janet M. Colson

Dr. Andrew Owusu

This dissertation is dedicated to my father, Edward P. Cook, Sr., and to all of the kids in my life: nephews, nieces, cousins, my godson, and munchkins. I love you all.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my dissertation chair and advisor, Dr. Norman Weatherby, for continually being a source of support and great enthusiasm throughout my educational journey. Additionally, I would like to sincerely thank the members of my dissertation committee. Dr. Andrew Owusu, for your lighthearted nature and for including me in so many research opportunities early in my career, thank you. Dr. Janet Colson, thank you for your valuable time, knowledge, and encouragement throughout this process. Finally, to Dr. Minsoo Kang, who helped spark the inspiration for this dissertation topic in Current Measurement Issues class many years ago, thank you for challenging me and helping me persevere. Your guidance has been paramount.

I am highly indebted to the members of the Health and Human Performance department including faculty, staff, and colleagues. The completion of this effort would not have been possible without the camaraderie and understanding of Dr. Tracy Morris, the laughter and offsite health education excursions with Dr. Poliala Mahoney Dickson, opportunities to learn and grow with Vickie Harden, and the intelligence and friendship of Dr. Youngdeok Kim and Dr. Yasir Alaboodi. There are so many others. If you are reading this, you helped me get here. Thank you.

For my mother, Nancy Proctor, my best cheerleader, thank you for your unconditional love, your unwavering belief that I can do anything, and for teaching me how to be good. For my sister, Mary, I could not have done this without you, your help, and your encouragement. To my brother, Eddie, for being a great example of success and a motivating force, thank you. For my father, thank you for teaching me to not be afraid to take risks and about the immense joy found in small moments of clarity. I am grateful

to Duane, Ron, and Tracy for loving them as much as I do and keeping the faith that I would finish one day. For my extended family and friends, for missed birthdays, dinners, baseball and football games, thank you for understanding and always being interested in my journey. I love you all.

Lastly, to Eli. You have been there for every hurdle and every small celebration with a relentless certainty that I would reach this goal, and the next. I am especially grateful to you for our home and your daily love and support which has assuredly guided me to this point, finished. I love you.

ABSTRACT

Proper screening for the risk of having an eating disorder is critical to provide necessary treatment and avoid negative health outcomes. The Eating Attitudes Test (EAT) is one of the most prominent assessment tools for eating disorder risk and is administered in both clinical and non-clinical settings. The EAT uses a Likert scale and relies partially on a total summed score of answers to determine whether a clinical referral for diagnosis and/or treatment is required. Historically, measurement methodologies have neglected to consider that the ordinal nature of total summed scores, or counts, does not provide adequate statistical analysis. In this study, a more advanced measurement model called Rasch, specifically the Rasch Rating Scale Model (RSM) and Rasch calibration, was utilized to analyze the properties of the EAT and to further investigate this valuable tool. The EAT was administered to a sample from a southern college community and an exploratory study was performed to evaluate its category function. A confirmatory study was then performed to substantiate findings from study one. Finally, a calibration was performed to study item difficulty and person ability.

Results show that the EAT with 26 items (EAT-26) may not function as intended. Post-hoc analysis suggested a four or three category Likert scale may be more optimal than the original with six. A confirmatory study provides more evidence including ordered thresholds and satisfactory person separation index and reliabilities. This implies that the participants are better able to distinguish between the categories of the adjusted Likert scale. The calibration of this four category scale provides additional insight into the hierarchy of EAT items and recommendations for future study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Body Mass Index (BMI)

Category (CAT) used in two ways, such as:

      A three category scale (3-CAT)

      The third category of a scale (CAT-3)

Classical Test Theory (CTT)

Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)

Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5)

Eating Attitudes Test (EAT)

      40 item version (EAT-40)

      26 item version (EAT-26)

      23 item version (EAT-23)

Eating Disorder Not Otherwise Specified (EDNOS)

Item Response Theory (IRT)

National College Health Assessment (NCHA)

Other Specified Feeding and Eating Disorders (OSFED)

Rating Scale Model (RSM)

Structural Equation Model (SEM)

CHAPTER I: INTRODUCTION

Eating disorders continue to affect up to 30 million people in the United States (Wade, Keski-Rahkonen, & Hudson, 2011). The three main eating disorders are anorexia nervosa, bulimia nervosa, and binge eating disorder ("Understanding Eating Disorders," 2008).

Anorexia nervosa has consistently maintained the highest mortality rate among all psychiatric disorders (Arcelus, Mitchell, Wales, & Nielsen, 2011; Berkman, Lohr, & Bulik, 2007; Harris & Barraclough, 1998). One meta-analysis found one in five of those with anorexia nervosa who died had committed suicide (Arcelus et al., 2011). Individuals with anorexia nervosa have significantly low weight and a fear of gaining weight or exhibit behaviors that prevent weight gain. There are two subtypes of anorexia nervosa. They are Restricting Type (weight loss through dieting, excessive exercise, and/or fasting) and Binge-Eating/Purging Type (binge eating or purging behavior). Individuals with bulimia nervosa may or may not be underweight, but they engage in periods of binge eating where one feels out of control, followed by compensatory activities such as excessive exercise, laxative and diuretic abuse, purging, and fasting. Those with bulimia nervosa can experience an electrolyte imbalance which can result in an irregular heartbeat and death (National Eating Disorders Association, 2016b). While the language between the two disorders is similar, they exist separately as "two 'specific psychopathologies' essentially as variants of a similar way of thinking occurring within different contexts and histories" (Palmer, 2014, p. 13).

Binge eating disorder is similar to bulimia nervosa but individuals do not express compensatory behaviors to avoid weight gain. Binge eating episodes happen when the

individual eats more rapidly than normal, eats until they are uncomfortable, eats when they are not hungry, eats alone due to embarrassment about the amount of food, and feels depressed and/or guilty following the binge (*Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.)*, 2013). Binge eating disorder is the most prevalent eating disorder as it affects 2-3% of the United States population (Supina, Herman, Frye, & Shillington, 2016) and recently was given its own designation in the latest version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Previously in the DSM-IV, binge eating disorder fell under the residual designation of eating disorders not otherwise specified (EDNOS).

As mental health disorders, anorexia nervosa, bulimia nervosa, and binge eating disorder mentioned above are familiar. Also worth mentioning are lesser known disorders, formerly referred to as EDNOS in DSM-IV and currently classified as other specified feeding and eating disorders (OSFED) subtypes. The EDNOS category of disorders was the most common diagnosis for those with disordered eating who did not meet the criteria for anorexia nervosa or bulimia nervosa (Palmer, 2014). This catch-all category included disorders such as night eating syndrome, binge eating disorder with less frequent occurrence than what is required for clinical diagnosis (sub-threshold), and purging without binge eating. Eating disorder subtypes are prevalent throughout our society yet only a small number of those who suffer receive care (Hoek & van Hoeken, 2003). Non-treatment can result in a variety of negative outcomes, both physical and psychological such as overweight status, reduced self-esteem, depression, suicidal ideation, and substance use (Fletcher, Kupshik, Uprichard, Shah, & Nash, 2008; Loth, Mond, Wall, & Neumark-Sztainer, 2011; Neumark-Sztainer & Hannan, 2000). While

EDNOS are sometimes viewed as less severe than anorexia nervosa or bulimia nervosa, the crude mortality rate for EDNOS (collectively) was found to be 5.2% while rates for anorexia nervosa and bulimia nervosa were 4.0% and 3.9% respectively (Crow et al., 2009). The EDNOS category has been removed and is now closely related to the new category, OSFED, in the new DSM-5 (Attia et al., 2013).

OSFED describes "A feeding or eating disorder that causes significant distress or impairment, but does not meet the criteria for another feeding or eating disorder" (National Eating Disorders Association, 2016a, para. 1). The diagnostic and categorical changes made from the DSM-IV to the DSM-5 in 2013, primarily removing binge eating disorder from EDNOS and creating the new classification of OSFED, are expected to reduce the amount of individuals placed in a broad residual category.

Some objectives included in the Healthy People 2020 initiatives include to "Increase the proportion of adults with mental health disorders who receive treatment" and to "Reduce the proportion of adolescents who engage in disordered eating behaviors in an attempt to control their weight" (U.S. Department of Health and Human Services, 2016). With these important Healthy People objectives, it is imperative that the measurement tools used to gauge potential for having an eating disorder are effective to ensure that those who need treatment are properly referred. The gold standard of the structured clinical interview is time-intensive and expensive to administer. For screening purposes, self-report instruments are common (Tury, Guelec, & Kohls, 2010).

Four general groups of eating disorder assessment methods exist. They are general measures of eating disorder symptoms, DSM diagnostic questionnaires, screening questionnaires, and measures of specific eating disorder symptoms (Tury et al., 2010).

General measures of eating disorder symptoms include instruments such as the Eating Disorders Inventory (EDI) with 64 items, and the Eating Disorder Examination Questionnaire (EDE-Q) with 33 items. Diagnostic measures include the Eating Disorder Diagnostic Scale (EDDS) which is 22 items, the Questionnaire for Eating Disorder Diagnoses (Q-EDD) which contains 50 item, and the Questionnaire on Eating and Weight Patterns –Revised (QEWP-R) with 28 items. The screening questionnaires include the Eating Attitudes Tests (EAT-40, EAT-26), Bulimia Test (BULIT) with 36 items, and the Bulimia Investigation Test (BITE) with 33 items (Tury et al., 2010). One of the more popular measures for specific eating disorder symptoms is the Binge Eating Scale (BES). As various diagnostic changes have taken place over the years, research and clinicians have attempted to keep up with a variety of instruments to target various needs.

Several issues in addition to changes to the diagnostic criteria make screening for eating disorders difficult. These include secrecy around the behaviors and symptoms, unwillingness to cooperate for treatment (Vandereycken & Van Humbeeck, 2008), and many blurred lines existing between healthy habits and eating disorder symptomology in the general population. Thoughts such as "I am occupied with a desire to be thinner" are now considered the norm in society for many (Tury et al., 2010). In light of these issues, it is important to continually assess the ways in which eating disorders are screened for, assessed, and treated.

The Eating Attitudes Test 26 (EAT-26)

The Eating Attitudes Test (EAT) was the first survey to evaluate symptoms of eating disorders (Mintz & O'Halloran, 2000). The self-report instrument was developed as a result of concerns raised at a National Institutes of Mental Health (NIMH) consensus

panel (Garfinkel & Newman, 2001). Originally designed to screen for anorexia nervosa in 1979, the EAT utilized the Feighner criteria (Feighner et al., 1972) for anorexia nervosa and initially consisted of 40 items (EAT-40). The survey was later refined to 26 items and is currently known as EAT-26. The Feighner criteria included having symptoms of the disease prior to age 25, losing at least 25% of one's original body weight, and distorted, unrealistic attitudes about weight, eating, and food regardless of hunger or logic. The criteria also stated that bulimia or self-induced vomiting could be characteristics of anorexia nervosa but bulimia nervosa was not designated as its own disorder. In the EAT-26, the anorexia nervosa restrictor type and the anorexia nervosa bulimia type were distinguished with questions targeting both subtypes (Garner, Olmsted, Bohr, & Garfinkel, 1982). For instance, item 2, "I avoid eating when I am hungry," and item 9, "I vomit after I have eaten," are two items describing potentially different subtypes of anorexia nervosa as it was understood at the time, yet it was not until the DSM-IV came out in 1994 that anorexia nervosa and bulimia nervosa were considered separate diseases with their own criteria for diagnosis. Additionally, when the EAT-26 was first created, there was no concept of atypical or sub threshold eating disorders (Mintz & O'Halloran, 2000). The numerous changes in diagnostic criteria for eating disorders over the years provide cause to question whether the EAT-26 performs well by current standards.

The EAT-26 has a three-factor structure. Thirteen items load onto Factor 1 called "Dieting." These items are related to carefully watching the diet for such things as calorie content, carbohydrate, and sugar content and having a desire to be thinner. Six items load onto Factor 2 which is referred to as "Bulimia and food preoccupation." These items

relate to vomiting after meals and thinking about food excessively. The final factor

includes seven items and is labeled "Oral control." Oral control items are associated with

self-control of eating and feeling pressure to gain weight from others (Garner et al., 1982,

p. 876-877). Factors are described as being correlated to the total EAT-26 scores with

Factor I having the highest correlations ($r = 0.93$) and Factors II ($r = 0.64$) and II ($r =$

0.60) having a weaker positive relationship (Garner et al., 1982). EAT-26 developers

suggested that in some cases, scores on the individual factors may be better predictors of

outcomes for specific groups, such as utilizing Factor 1 questions with only anorexia

nervosa subjects since it was not found to be related to bulimia. However, it appears most

studies to date have utilized the total summed score method when using the EAT-26

(Garfinkel & Newman, 2001; Mintz & O'Halloran, 2000).

Development and use of the EAT-26.

The design of the EAT-26 includes 26 statements with a six category (6-CAT)

Likert scale of *Always, Usually, Often, Sometimes, Rarely,* and *Never.* Statement

examples include "I eat diet foods" and "I am preoccupied with the thought of having fat

on my body." Scoring is based on a 4-point scale with *Always* receiving three points,

Usually receiving two points, *Often* receiving one point and *Sometimes, Rarely,* and

Never receiving zero points. One item, number 26, is reverse scored. The overall EAT-26

instrument is based on a referral index measured by 1) the total score based on the

answers to the EAT-26 items, 2) answers to several behavioral questions related to eating

symptoms and weight loss, and 3) the individual's body mass index (BMI) calculated

from self-reported height and weight (Garner et al., 1982). Any respondent who meets

the cut-off criteria for any of the three items is considered at risk for having an eating disorder and should be referred to a qualified health practitioner.

While the original version of the EAT was designed among clinical samples, the EAT-26 is described as a screening tool to assess the risk for eating disorders which can be used in both clinical and non-clinical settings such as schools, camps, and sports programs. It is important to note that the EAT-26 is not a diagnostic tool but is typically used as a screening or case-finding instrument (Garner, 1993). The inclusion of the behavioral health questions which were added during the National Eating Disorders Screening Program in 1998 helped to improve the overall instrument. However "neither the EAT, nor any other screening instrument, has been established as highly efficient as the sole means for identifying eating disorders. This is attributable in large part to the relatively low prevalence of eating disorders in most populations of interest" (Garner, 2016, para. 11)

The EAT-26 has been used in studies across a variety of populations including German children (Berger et al., 2011), ballerinas (Girao, Antunes, & Silva, 2009), and bipolar patients (Alexandre, Ribeiro, & Cardoso, 2008). According to the EAT-26 website, "it became a Current Contents Citation Classic in 1993 and since that time papers describing the test's development and validation have been some of the most cited papers in the scientific literature on eating disorders" (EAT-26, 2011, para. 1).

The application of the EAT-26 among disordered eating studies has expanded since its initial development. The EAT-26 was recently used to provide both convergent and incremental validity to the preliminary validation of the Yale Food Addiction Scale (YFAS), a popular new measurement tool for screening for food addiction (Gearhardt,

Corbin, & Brownell, 2009). The validation of the Chinese version of the YFAS also utilized the EAT-26 (Chen, Tang, Guo, Liu, & Xiao, 2015).

Criticisms of the EAT-26.

Not all studies provide evidence for the EAT-26 instrument's reliability and validity, yet it has almost become second-nature to cite the EAT-26 as an effective screener for eating disorders without due diligence in assessing its current status. One study (Merlo, Klingman, Malasanos, & Silverstein, 2009) cited Ocker, Lam, Jensen and Zhang (2007) as evidence for the good psychometric properties of the EAT-26. The Ocker et al. study noted that the EAT-26 had been used extensively, however, it also reported "The EAT-26 model…had the least acceptable model fit" and highlighted its shortcomings (p. 35). Through confirmatory factor analysis (CFA), the study found that "the fit indexes from the three-factor EAT-26 represented unacceptable model fit (RMSEA = .11, SRMR = .11, CFI = .73, AGFI = .74)" (Ocker et al., 2007, p. 1) and a 16 item version was preferred. Maiano et al. (2013) found 60% of EAT-26 studies among clinical samples of adolescents were unable to replicate the three-factor structure of the EAT-26. Their exploratory Structural Equation Modeling (SEM) of the EAT-26 found a better model fit using an EAT with only 18 items. The original EAT had 40 items and seven factors which most studies have also failed to replicate (Maiano et al., 2013). Vetrone et al. (2006) showed that the introduction of the EDNOS diagnoses lowered the sensitivity of the EAT-40. The authors concluded "that using a two stage screening approach leads to a very high rate of false negatives with a significant underestimation of the prevalence of eating disorders, particularly of EDNOS" (p. 1).

Rasch Model

One criticism of the types of studies that have helped to provide validity and reliability evidence for the EAT-26 is that social scientists tend to assume that collected data are interval when they are, in fact, ordinal (Bond & Fox, 2001). As explained by Bond & Fox, "...the scales to which we routinely ascribe that measurement status in the human sciences are often merely presumed to have interval-level measurement properties; those measurement properties are almost never tested empirically. It is not good enough to allocate numbers to human behaviors and then merely to assert that this is measurement in the social sciences" (p. 5). The confusion results from different interpretations of ordinal and interval data definitions in Stevens' *On the theory of scales of measurement* (Stevens, 1946). However, Stevens made it clear that interval data is required for statistical interpretations such as correlations. The importance of understanding that counting observations may not be sufficient and that ordinal counts need linear and ratio measures to understand them is not a new concept (Wright & Linacre, 1989). Zhu (1996) stated "if total scores are obtained from a scale without meaningful order or equal measurement units, they will not meet acceptable standards for reliability or validity. The problem is that when this happens, we usually cannot determine whether the failure was caused by inappropriate numeric properties of the scale or inadequate validity or reliability of the instrument" (p. 3).

Georg Rasch began to develop the Rasch measurement model after working alongside Fisher and introduced his model in his 1960 book, "Probabilistic models for some intelligence and attainment tests." One of the main tenets of his measurement philosophy is that "a measure must retain its quantitative status, within reason, regardless

of the context in which it occurs" (Wright & Linacre, 1989, para. 11). A common

analogy that has been used to describe Rasch is the idea of a ruler built to a constant

length regardless of what is being measured or a thermometer only measuring

temperature (Boone, Staver, & Yale, 2013). An assessment, a survey, a test, or other

measurement tool must also maintain its calibration regardless of the person answering

the questions or the latent trait being measured (Bond & Fox, 2015; Rasch, 1960; Wright

& Linacre, 1989). This first principle of the Rasch model is referred to as invariance. The

second principle of the Rasch philosophy refers to the interaction between the person

taking an assessment and the items themselves. As a probabilistic model, Rasch changes

the outcome from how a person will definitively answer an item to how a person is most

likely to answer an item. As a person answers a more difficult question correctly, the

likelihood that they will answer an easier question correctly increases.

Rating scale model.

Wright and Masters (1982) were the first to apply the Rasch model to category

functioning of rating scales. It has since been determined that Rasch analysis is an

effective method to establish the function of categories of rating scales, or the extent to

which the categories are working as they were intended to work (Andrich, 2011; Bond &

Fox, 2015; Zhu, Updyke, & Lewandowski, 1997).

As a polytomous model for rating scale data, the Rasch Rating Scale Model

(RSM) explains the probability that a person (n), encountering an EAT-item (i) will

respond to a category choice of *Always* through *Never* (k) based on their ability (θ_n) and

the difficulty of the item (δ_i) resulting in the RSM log-odds formula of $\ln (P_k / P_{k-1}) =$

$\theta_n - \delta_i - \tau_k$ (Bond & Fox, 2015; Linacre, 2002a; Myers, Feltz, & Wolfe, 2008). The RSM

builds on the Rasch dichotomous model by adding the concept of the category threshold (τ_k) (Linacre, 2002a; Myers et al., 2008). A threshold is described as the point where two categories intersect. At this point, there is a 50% likelihood of a response in either adjacent category (Andrich, 1996, 1998).

The Rasch model is an advanced measurement, item response theory that creates interval data from ordinal data using logits (Rasch, 1960, 1980). The model allows researchers to take the more vigorous linear measures and apply statistical analysis with which Rasch provides several parameters to consider (Zhu et al., 1997). These include model-data fit statistics such as Infit and Outfit Mean-Square (Infit and Outfit) (Linacre, 2002b; Wright & Masters, 1982) and person and item separation and reliabilities.

Infit and Outfit statistics.

Infit and Outfit statistics help to describe the fit of the items to the Rasch model. Infit is the standardized weighted mean square and Outfit is the standardized unweighted mean square (Bond & Fox, 2015; Zhu et al., 1997). Ideal Infit and Outfit statistics have an expected value of 1. Values which are above 1 may have more variation between the model (expected scores) and the observed scores. For example, a fit statistic of 1.50 indicates 50% more variation than what was predicted by the model. This suggests underfit with the model or that answers are more unpredictable than expected. On the other hand, a lower fit statistic of .50 indicates 50% less variation. This is also referred to as overfit or when a test-taker gets harder items correct even though their ability is lower (Bond & Fox, 2015; Kang, Zhu, Ragan, & Frogley, 2007; Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). There are several acceptable fit ranges from which

investigators can adopt based on the type of test such as multiple choice test (0.7-1.3), clinical observation (0.5-1.7), or rating scale (0.6-1.4) (Bond & Fox, 2015).

Person and item separation and reliability.

The person separation statistic expresses how well people (person ability) are separated to show individual differences. Item separation explains how the items function along the scale or how much differentiation there is among items (Wright & Masters, 1982). Person separation is helpful to sort people into different groups. An instrument with a low person separation (< 2, with a person reliability of < 0.8) suggests that the tool may not be able to separate high performers from low performers. Including additional items may result in groups which are easier to differentiate from each other. Item separation is useful to distinguish items that are easier to endorse from items that are more difficult to endorse. Low item separation (< 3, with an item reliability < 0.9) suggests that there may not be enough people in the sample to determine the item separation (Linacre, 2016). Adding a wider range of people may help differentiate the items effectively.

Unidimensionality.

One assumption of Rasch analysis is that all of the items of a survey, assessment, or test must measure one latent trait or variable. This is referred to as unidimensionality. No test can ever perfectly achieve complete unidimensionality, yet that is ideal for proper and useful measurement (Wright & Linacre, 1989). Researchers and practitioners often have to find a middle ground between unidimensionality and purposeful application of an instrument in a clinical setting (Kean, Malec, Altman, & Swick, 2011). In Rasch analysis, data are fit to the unidimensional model and aspects of the data that do not fit are

highlighted for further investigation (i.e., removing an item, removing a person,

rewording an item, collapsing categories, etc.). Rasch "analyzes the data as though they

are unidimensional, and then the fit statistics report how well the data match the

mathematically unidimensional framework that the Rasch analysis has constructed"

(Linacre, 2011, para. 6). An Outfit mean-square statistic > 2.0 associated with a specific

category can "suggest that there is more unexplained noise than explained noise, so

indicating there is more misinformation than information in the observations" and that

portions of the data may not be useful for measurement (Linacre, 2002a, p. 96).

Rasch methodology has previously been used in the field of eating disorder

measurement (Cloyd, 2005; Stein, Riley, Hoyland-Domenico, & Lee, 2015;

Waaddegaard, Thoning, & Petersson, 2003). Most recently, Stein et al. (Stein et al.,

2015) used Rasch measurement to examine one subscale of the Eating Disorders

Inventory III (EDI-III). After the removal of two items due to "significant misfit," the

authors found the subscale had good item fit and good person and item reliability (Stein

et al., 2015, p. 6). The six category Likert scale also functioned well.

One can never say that a test or measurement is ever perfect or complete. Every

time a survey is assessed against a different measurement model, new information is

obtained to guide the improvement of the test (Wright & Linacre, 1989). Since the EAT-

26 is purported to be a unidimensional instrument using summative scoring of Likert

scale responses, it is acceptable to evaluate using this method. There are no known

studies to date which have analyzed the English version of the EAT-26 utilizing the

Rasch model. The Spanish version of the EAT-40 was assessed with the Rasch model

and was found to discriminate between a sample of women in Costa Rica who had

exhibited risk of having an eating disorder and those who did not (Salazar Mora & Prado-Calderón, 2015). Category function was not an aspect of the study.

Validation of measurement instruments using a variety of methods is important to ensure the tools are acceptable. One factor of validity is analyzing whether the rating scales are functioning as intended and measuring the desired latent trait (Myers et al., 2008). On face value, it seems possible that a respondent to the EAT-26 may have trouble differentiating between some of the category options such as *Usually/Often*, *Often/Sometimes*, and *Sometimes/Rarely*.

Purpose of the Study

The broad goal of this study is to assess the Eating Attitudes Test (EAT-26) utilizing the Rasch model. Specifically, the first study aims to 1) evaluate category function of the EAT-26, and 2) perform an exploratory analysis of post-hoc combined categories. Data from this initial study led to additional inquiries to perform a confirmatory analysis of the categorical function of the rating scale for both three and four category response options with the EAT-26. The final study presents a calibration of the 4-CAT option of the instrument. The results of these studies allow for potential improvement of this often-used measurement instrument.

CHAPTER II: AN EXPLORATORY STUDY OF THE PSYCHOMETRIC

PROPERTIES AND CATEGORY FUNCTIONS OF THE EATING ATTITUDES

TEST (EAT-26) UTILIZING RASCH ANALYSIS

All eating disorders, including anorexia nervosa, bulimia nervosa, and binge eating disorder, have an elevated mortality risk (Smink, van Hoeken, & Hoek, 2012) and continue to represent a serious public health concern (Le Grange, Swanson, Crow, & Merikangas, 2012) .

The Eating Attitudes Test (EAT-26) is described as a screening tool to assess eating disorder risk which can be utilized in both clinical and non-clinical settings such as schools, camps, and sports programs. According to its creators, the EAT-26 is "probably the most widely used standardized self-report measure of symptoms and concerns characteristic of eating disorders" (EAT-26, 2011, para. 1). It has been used in a wide variety of cultures and locations with versions in multiple languages (Al-Subaie, 1998; Anderson-Fye & Lin, 2009; Anthi, Maria, Nicole, Evel, & Kalia, 2011; Choudry & Mumford, 1992; Pereira et al., 2008; Szabo & Allwood, 2004) and also with various samples such as ballet students (Neumarker, Bettle, Bettle, Dudeck, & Neumarker, 1998), college athletes (Doninger, Enders, & Burnett, 2005), and exercisers (Lane, Lane, & Matheson, 2004).

Originally designed as a 40-item measurement to assess the symptomology of anorexia nervosa (Garner & Garfinkel, 1979), the EAT was later reduced to 26 items (Garner, Olmsted, Bohr, & Garfinkel, 1982) and has since been used to identify not only anorexia nervosa but also bulimia nervosa (Gross, Rosen, Leitenberg, & Willmuth, 1986), disordered eating symptoms, and binge eating disorder (Orbitello et al., 2006).

The 26 items include statements such as "Other people think that I am too thin," "I like my stomach to be empty," and "I have gone on eating binges where I feel that I may not be able to stop" to which respondents choose from six categories ranging from *Always* to *Never* for a total score measurement of disordered eating behaviors (Table 1).

Rating scale categories should be "well defined, mutually exclusive, univocal and exhaustive" (Guilford, 1954; Linacre, 2002a, p. 86). They also have to be very clear to the respondents to ensure the intent of their answer is captured effectively (Stone & Wright, 1994). It is unclear whether respondents of the EAT-26 successfully distinguish between adjacent categories such as *Usually / Often* and *Often / Sometimes*. Six categories also may be considered burdensome for the participant when three, four, or five categories are sufficient. No published study to date has evaluated the function of the categories with the English version of the EAT-26.

Purpose of the Study

The purpose of this study is to 1) evaluate the function of the six category (6-CAT) version of the EAT-26 and to 2) perform an exploratory analysis of post-hoc recombined categories to determine if any may be more effective in measuring eating disorder risk than the original 6-CAT structure. This study will provide the model-data fit (Infit and Outfit statistics), an analysis of the function of the rating scale of the EAT-26, and potential new rating scale formats.

Methods

Institutional Review Board (IRB) approval was obtained prior to conducting the study. As part of IRB approval, participants were notified that their inclusion in the study

Table 1

Eating Attitudes Test (EAT-26) Items

1	I am terrified about being overweight.
2	I avoid eating when I am hungry.
3	I find myself preoccupied with food.
4	I have gone on eating binges where I feel that I may not be able to stop.
5	I cut my food into small pieces.
6	I aware of the calorie content of foods that I eat.
7	I particularly avoid food with a high carbohydrate content (i.e. bread, rice, potatoes, etc.)
8	I feel that others would prefer if I ate more.
9	I vomit after I have eaten.
10	I feel extremely guilty after eating.
11	I am occupied with a desire to be thinner.
12	I think about burning up calories when I exercise.
13	I other people think that I am too thin.
14	I am preoccupied with the thought of having fat on my body.
15	I take longer than others to eat my meals.
16	I avoid foods with sugar in them.
17	I eat diet foods.
18	I feel that food controls my life.
19	I display self-control around food.
20	I feel that others pressure me to eat.
21	I give too much time and thought to food.
22	I feel uncomfortable after eating sweets.
23	I engage in dieting behavior.
24	I like my stomach to be empty.
25	I have the impulse to vomit after meals.
26	I enjoy trying new rich foods.

was completely voluntary and that they could cease participation at any time without consequence.

A convenience sample of 368 males and females, primarily students from a public university in the Southeastern United States, volunteered to participate in the study. Fifty three participants were excluded due to missing or obvious patterned data resulting in a total sample size of 315 which meets the typical requirement of $n = 200$ for a Rasch analysis (Kang, Zhu, Ragan, & Frogley, 2007). As an incentive, some students were provided with extra credit points from their professors for participating. In an effort to maximize the range of participants, the researchers attempted to reach a wide variety of class-years as well as class-types such as beginning health, nursing, exercise science, nutrition, social work, and sociology. Non-students who participated did not receive any incentive. The only inclusion criterion was that each participant had to be at least 18 years of age. The researchers informed the participants through flyers and announcements that the purpose of the study was to evaluate measurement tools for eating behaviors. Data were collected in large groups in a classroom setting, and cover sheets were provided to ensure confidentiality in responses. An informed consent document was given to all students which detailed potential harms and benefits of participation, the nature of the data to be collected, and how they could cease to participate at any time. No personally identifying information such as names or birth dates was collected and all surveys were identified with only a numerical code for anonymity. Demographic data including gender, age, and body mass index (BMI) calculated from self-reported height and weight was collected and is reported in Table 2.

Table 2

Demographic Information of Participants (n = 315)

Characteristic	*n*	%
Gender		
Women	220	69.84
Men	95	30.16
Race/Ethnicity *(n = 312)*		
White	162	51.92
Black	121	38.78
Hispanic	10	3.21
Asian	9	2.88
Other	10	3.21

	M	*SD*
Age	21.66	5.45
BMI		
Women	25.66	6.93
Men	26.30	5.96

Measure.

Each item on the EAT-26 receives three points for *Always* responses, two points for *Usually*, and one point for *Often* responses. Other responses receive zero points. Item 26 is the only item that is reverse scored. A total score of 20 or above indicates a referral to a health professional capable of diagnosing an eating disorder is necessary. The other criteria for screening on the EAT-26 are underweight BMI, and answers to health questions with six month's recall that point toward problematic eating behaviors. For this

study, only the initial 26 items are assessed. Despite the four category scoring method, a Likert scale with the full 6 categories (*Always* = 1, *Usually* = 2, *Often* = 3, *Sometimes* = 4, *Rarely* = 5, and *Never* = 6) as explained by Tury, Guelec, and Kohls, (2010) is used for this analysis and item 26 is reversed to match other items (Linacre, 2002a).

Data analysis.

The Rasch Rating Scale Model (RSM;Andrich, 1978; Masters, 1980; Wright & Masters, 1982) was used to evaluate the EAT-26 instrument with the computer program WINSTEPS version 3.75 (WINSTEPS, 2012). The RSM is expressed as:

$$\log\left(\frac{\pi_{nix}}{\pi_{ni(x-1)}}\right) = \beta_n - \left(\delta_i + \tau_x\right)$$

Linacre (2004) explains "$\pi_{ni(x-i)}$ is the probability of person *n* being observed in category *x-1* of item *i*. This expression of the Rasch model emphasizes that the probabilistic, log-odds, structure of the data, on the left, is modeled to be the manifestation of an additive combination of latent parameters on the right" (p. 96). The Rasch model provides several tools to assess the effectiveness of a measure and its category functions (Zhu, Updyke, & Lewandowski, 1997). This study's methodology for assessing the rating scale structure of the EAT-26 was derived and influenced by previous works by Linacre (1999b, 2002a), Zhu and colleagues (1997), and Myers and colleagues (2008).

To determine the effectiveness of rating scale categories, each category is assessed for its functioning using guidelines for optimizing rating scale category effectiveness (Linacre, 2002a). While there are eight total guidelines, Linacre provides a summary table to select only those guidelines which are most pertinent to the study (p. 104). This table allows the researcher to select the most essential or helpful guidelines

based on the goals of the study. Six of the guidelines were selected for this study and are highlighted below.

      1) Average measures (average ability of people responding in a category) advance monotonically with each category.

      2) Outfit mean squares < 2.0 which suggest acceptable fit.

      3) Step difficulties (thresholds) which advance. Thresholds "indicate the measures at which adjacent categories are equally probable and thus define the boundaries between the categories" (Andrich, 1996, para. 4). It is important that as the risk of having an eating disorder increases among respondents, those respondents' probabilities of answering categories associated with higher risk of having an eating disorder increase as well.

      4) Not only must thresholds advance, they must do so by required amounts. Linacre (2002a) provided guidelines for thresholds to advance by 1.4 logits for 3-CAT scales and 1.0 logits for 5-CAT scales. Myers, Feltz, and Wolfe (2008) cited evidence from Huynh (1994) to determine the 4-CAT separation amount of 1.1 logits.

      5) Step difficulties (thresholds) should advance by less than 5.0 logits.

      One tool which is useful in the assessment of these guidelines is the graph of category probabilities available in Winsteps. This graph helps to visualize the distinctions between the categories. If a category has too much spread across other categories or does not appear as modal, or if the categories do not increase monotonically, the step calibrations (or thresholds) are considered disordered. This can provide insight into whether the category is too narrow or if too many answer choices were provided for the respondent to be able to discriminate between them effectively (Linacre, 1999a).

Empirical item-category measures plots in Winsteps provide an additional visual tool to assess whether categories are ordered and advance in a 1-2-3-4-5-6 manner.

Several studies have also assessed category effectiveness by selecting the category combination with the highest separation statistics (Royal, Ellis, Ensslen, & Homan, 2010; Zhu et al., 1997). Item separation statistics determine how well the scale differentiates the items of the measurement while the person separation statistics determine how well the scale puts people into different ability groups (Zhu et al., 1997). In this study, the person separation indexes were considered acceptable if they were at least 2.0 as "the person separation index and the person separation reliability should be important indicators of effective rating scale structure" (Myers et al., 2008, p. 303). Item separation statistics were not assessed because the evidence for item separation indexes as useful for indicating effective rating scale structures is unclear (Myers et al., 2008).

Category frequency was also considered as observations across categories which are uneven or irregular could signal problems with category function (Linacre, 2002a; Zhu et al., 1997). An ideal category frequency would have equal amounts per category, (i.e., a 3-CAT scale would have 33% frequencies for each category or a 4-CAT scale would have 25%).

Model-data fit is assessed with two fit statistics, referred to as Infit and Outfit. Infit is the information-weighted fit and Outfit is the standardized unweighted mean square (Bond & Fox, 2015; Zhu et al., 1997). Infit and Outfit statistics have an expected value of 1 and can range from 0 to infinity. Infit statistics are sensitive to irregular responses where items and persons are well-targeted while Outfit statistics are sensitive to items and person across the entire range of items and person locations. Values which

are above 1 may have more variation between the model (expected scores) and the observed scores while values below 1.00 have less variation (Engelhard, 2013). A fit statistic valued at 1.50 would indicate 50% more variation than what was predicted by the model, or what is also referred to as underfit. A fit statistic of .50 would mean 50% less variation also referred to as overfit (Bond & Fox, 2015; Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008). A few acceptable fit ranges include options for multiple choice tests (0.7-1.3), clinical observations (0.5-1.7), or rating scales (0.6-1.4) (Bond & Fox, 2015). For this study, any Infit or Outfit statistic less than 0.5 or greater than 1.5 warranted further consideration (Linacre, 2002b).

Results

Evaluate the category function of the original 6-CAT scale.

The average measures were ordered and the Outfit mean squares for categories were all < 2.0 suggesting adequate fit. One of Linacre's guidelines states that step difficulties (thresholds) should advance in order (2002a). Thresholds did not advance in order or by the adequate amounts with category 1 = none, category 2 = -.53, category 3 = -.16, category 4 = -.41, category 5 = .56, and category 6 = .53. This disordering of thresholds could be a signal that the rating scale may be problematic.

The category probability graph (Figure 1) shows the categories do not advance monotonically. The category representing *Usually* intersects with *Sometimes* before it intersects with *Often*. The *Often* category intersects with *Always* and *Sometimes* at approximately the same point (-.04 logits). The *Sometimes* category then intersects with both *Rarely* and *Never* at .06 logits. Additionally, each category should have a modal peak that is clearly distinct and separate from the other categories. *Rarely* and *Often*
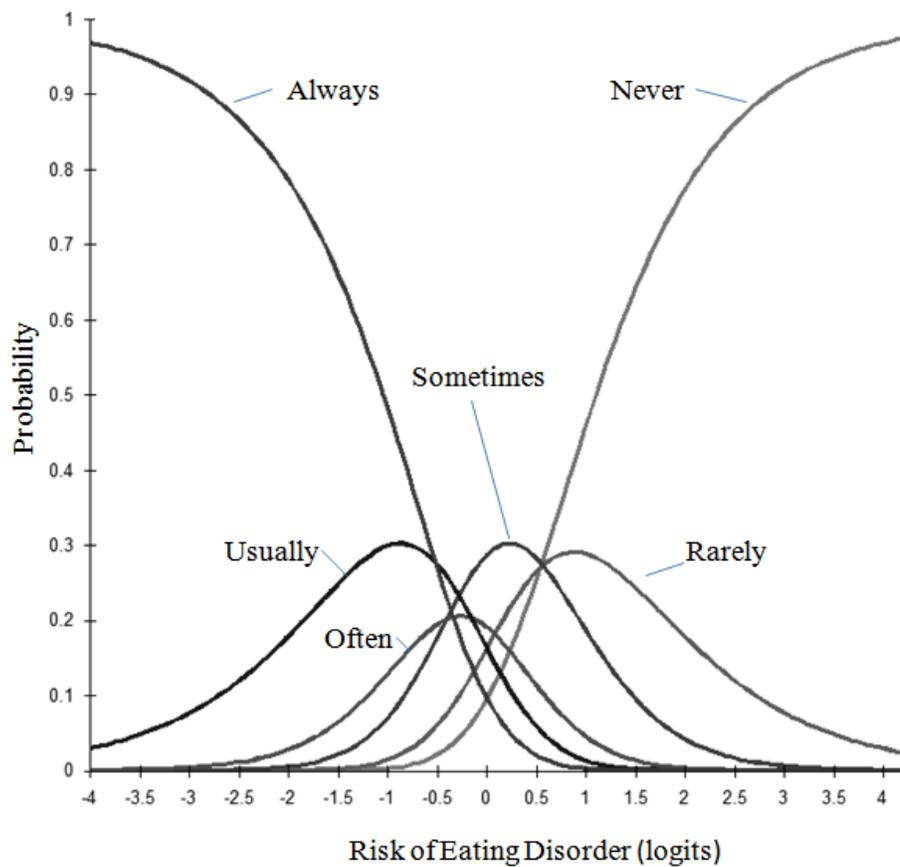
Figure 1. Category probability graph of the 6-CAT original EAT-26 showing disordered thresholds.

never have a likelihood of being selected that is higher than other categories. This is clearly illustrated in the figure by noting there is no independent peak such as what is observed with *Sometimes*. *Usually* has a very small likelihood of being selected over other categories at logit -.04. The empirical item-category measures plots in Winsteps confirm the findings of disordered thresholds. Several items show respondents disordering between *Always, Usually,* and *Often,* and *Sometimes* and *Rarely*.

The person separation statistic was acceptable at 2.60 with 0.87 separation-reliability. The 6-CAT rating scale category frequencies were not regular. In this sample, 39% answered *Never,* 22% answered *Rarely,* 19% answered *Sometimes,* 8% answered *Often*, 7% answered *Usually*, and 6% answered *Always*. This suggests several infrequently used categories (*Always*, *Usually*, and *Often*) and one category used much more than the others (*Never*). Misfit results showed 15% of items were outside of the acceptable Infit and Outfit range.

Exploratory analysis of post-hoc recombined categories.

The results of the Rasch analysis suggested that the categories of the original scale were not functioning as intended. As a result, categories were collapsed for additional analysis. Andrich (2010) states "... when data do not fit the model or when thresholds are reversed, then exploratory analysis which might involve pooling categories in such a way may be instructive in understanding the misfit or the reversed thresholds" (p. 50). The six categories were collapsed into their adjacent categories to create new rating scale combinations. First, all possible combinations were created manually and "in a somewhat mechanical way" as described by Zhu, et al. (1997a, p. 301) and as "mechanically thorough" as described by Myers, et al. (2008, p. 301). For example,

categories 2, and 3 from the original six categories (123456) were combined into category 1 to create a four category scale (111234). Alternatively, a five category scale was created by collapsing categories 5 and 6 (123455) and a three category scale was created by collapsing categories 2, 3, 4, and 5 (122223). This resulted in a total of 25 possible category combinations to compare to the original 6-CAT rating scale. A two category scale was not utilized in this study due to results and guidelines from earlier studies (Myers et al., 2008; Zhu et al., 1997).

While Myers (2008) provided some criticisms of the mechanically thorough method as being too cumbersome and not explicitly following the guidelines of Wright and Linacre (1992a), it was selected for this study based on his suggestions for a somewhat "hybrid" model (p. 309). This hybrid model includes the suggestion to construct all possible rating scale options, remove those that do not meet one of the guidelines for collapsing categories, and evaluate the remaining collapsings. The mechanical method of collapsing categories may seem excessive, yet the current Rasch computing software makes the task very manageable and ensures every possible category combination is considered (Zhu et al., 1997).

Recombined category combinations were assessed with the same Linacre's guidelines as the original 6-CAT but were subsequently removed from further analysis if a guideline was not satisfied. Qualitative judgment was then applied to the remaining collapsings that met the guidelines to determine practicality of certain category combinations and to ensure that they made sense (Wright & Linacre, 1992b).

A total of 25 post-hoc data sets representing the collapsed mechanically created category combinations were analyzed according to Linacre's guidelines. These

collapsings included 3-CAT, 4-CAT, and 5-CAT possibilities. All collapsings met the first two guidelines which were ordered average measures, and Outfit mean-squares less than 2.0, which were the same as the original 6-CAT scale. Table 3 shows the comparisons of post-hoc category combinations for the next guidelines, step difficulties which advance and by the required amounts.

A total of five category combinations (122334, 122223, 111223, 112223, and 122233), which are marked by asterisks in Table 3, met the requirements for thresholds to advance by designated amounts based on the number of categories compared to the original structure (123456) which showed disordered thresholds. These post-hoc category collapsings with ordered thresholds also showed higher person separation (2.76 - 2.81) and person reliability (.88 - .89) statistics than the original 6-CAT structure (2.60, 0.87), except 122233 which was slightly lower at 2.32 and 0.84.

These five collapsings were selected for continued analysis based on the hybrid model. Category frequency along with function (average measures and threshold detail) for these items is provided in Table 4. Several of the collapsings showed improvement in balanced category frequencies over the original 6-CAT scale. One potentially problematic collapsing, 122223, had 55% of the responses for the middle category.

Table 3

Comparison of Threshold Guidelines for Rating Scale Effectiveness Across Post-hoc Category Combinations

Collapsing		Step Difficulties Advance	SD Advance by ≥ 1.4(3) ≥ 1.1(4) ≥ 1.0(5)	Person Separation	Person Reliability
123456		disordered	No	2.60	0.87
112345		disordered	No	2.67	0.88
122345		ordered	No	2.85	0.89
123345		disordered	No	2.73	0.88
123445		disordered	No	2.46	0.86
123455		disordered	No	2.11	0.82
111234		ordered	No	2.66	0.88
122234		disordered	No	2.96	0.90
123334		disordered	No	2.53	0.86
123444		disordered	No	1.56	0.71
112234		disordered	No	2.86	0.89
112334		disordered	No	2.61	0.87
112344		disordered	No	2.19	0.83
122334	*	ordered	Yes	2.76	0.88
122344		disordered	No	2.33	0.84
123344		disordered	No	2.11	0.82
123333		disordered	No	1.18	0.58
122223	*	ordered	Yes	2.80	0.89
111123		disordered	No	2.57	0.87
111223	*	ordered	Yes	2.76	0.88
112223	*	ordered	Yes	2.81	0.89
112333		disordered	No	1.64	0.73
112233		ordered	No	2.30	0.84
111233		disordered	No	2.23	0.83
122233	*	ordered	Yes	2.32	0.84
122333		ordered	No	1.65	0.73

Note. All collapsings met guidelines 1 and 2.

* indicate combinations with ordered thresholds that also advance

Table 4

Summary of Category Frequency and Function for Collapsings Meeting Guidelines

Collapsing	Category frequency (%)						Average measure						Threshold estimate				
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5
123456	6	7	8	19	22	39	-0.58	-0.03	0.2	0.53	0.78	1.43	-0.53	-0.16	-0.41	0.56	0.53
122334	6	15	40	39			-0.92	-0.15	1.06	2.36			-1.38	-0.33	-1.71		
122223	5	55	39				-1.17	0.98	3.07				-2.34	2.34			
111223	20	40	40				-1.03	0.27	1.77				-1.03	1.03			
112223	12	49	39				-1.1	0.54	2.32				-1.61	1.61			
122233	6	33	61				-0.59	1.19	2.84				-1.41	1.41			

Note. Frequency percentages as reported in Rasch output may not equal 100% due to rounding. Category combinations represent collapsings that met the selected Linacre's guidelines.

The summary of misfit items is displayed in Table 5. The original 6-CAT rating scale had 85% satisfactory fit statistics (85% Infit and 85% Outfit) with an Infit mean of 1.04 and an Outfit mean of 1.12. Collapsings that featured ordered thresholds which also advanced according to Linacre's guidelines (2002a) had 90% satisfactory fit statistics (93% Infit and 86% Outfit). These collapsings had an Infit mean of 1.00 and an Outfit mean of 1.05, closer to 1.00 than the original 6-CAT rating scale structure. The ordered post-hoc category combinations displayed fewer misfit items than the original 6-CAT rating scale. It would be expected for these statistics to improve after collapsing categories, particularly the Outfit statistics which are most influenced by extremes (Linacre, 2016). This was observed for collapsings 122223 and 122233.

Table 5

Summary of Misfit Items Across Post-hoc Category Combinations

Collapsings	Infit				Outfit				Infit		Outfit	
	Fit		Not fit		Fit		Not fit					
	n	%	n	%	n	%	n	%	Mean	SD	Mean	SD
123456	22	85%	4	15%	22	85%	4	15%	1.04	0.34	1.12	0.57
122334	23	88%	3	12%	22	85%	4	15%	1.01	0.31	1.06	0.47
122223	25	96%	1	4%	23	88%	3	12%	0.99	0.26	1.02	0.36
111223	24	92%	2	8%	22	85%	4	15%	1.00	0.30	1.08	0.52
112223	24	92%	2	8%	22	85%	4	15%	0.99	0.29	1.05	0.43
122233	25	96%	1	4%	23	88%	3	12%	1.00	0.22	1.02	0.43

Note. Category combinations represent collapsings that met the selected Linacre's guidelines. Adequate fit determine by the range of .5-1.5.

The next step in making the determination as to which category combinations might be most effective was to take a more subjective approach with evaluating collapsed categories to ensure each new category combination was sensible (Wright & Linacre, 1992a). The decision was based not only the statistical results, but also the consideration of the category descriptions. For instance, combination 122223 combines *Usually*, *Often*, *Sometimes,* and *Rarely* into one category. *Usually* and *Rarely* seem to be distinct categories and would most likely not be confusing among respondents, therefore combining them does not make logical sense. For this reason, this collapsing was not considered further. The remaining category options (122334, 111223, 112223, and 122233) could practically be combined and may provide improvements on the original 6-CAT scale.

Discussion

The purpose of this study was to 1) evaluate the function of the six category (6-CAT) version of the EAT-26 and to 2) perform an exploratory analysis of post-hoc recombined categories to determine if any were more effective in measuring eating disorder risk than the original 6-CAT structure. The current EAT-26 does not fit the Rasch model with 6 categories. Linacre cited a summarization of Guilford (1954) by Nunnally (1967) that "in terms of psychometric theory, the advantage is always with using more rather than fewer steps" except "in instances where a large number of steps confused subjects or irritated them" (Linacre, 2002a, p. 86). This may well be the issue with the rating scale of the EAT-26, but the disordered thresholds "degrades the interpretability of the resulting measures. It can indicate that a category represents too narrow a segment of the latent variable or corresponds to a concept that is poorly defined

in the minds of the respondents" (Linacre, 2002a, p. 98). The disordered thresholds provide some evidence that respondents may have had difficulty clarifying their responses on the screening tool and that the EAT-26 may be easier to administer and provide improved measure with fewer categories.

It was observed that instead of the six category scale, a 4-CAT scale (122334), combining categories 2 and 3, and 3 and 4, or the 3-CAT options of 111223, 112223, or 122233 may be more effective. These combinations resulted in ordered thresholds, higher person separation index and reliability, and improved model-data fit, except for 122233 which did not have a higher person separation index and reliability. Typically, reducing the number of categories results in reduced separation and reliability statistics (Myers et al., 2008; Zhu et al., 1997). While the person separation index was reduced in this 122233 collapsing, it is important to note that the value of the person separation index across all of the various structures should not be the only determining factor when deciding category effectiveness. Instead, the "relative decrease in separation indexes in relation to the change in the other indicators" should be considered (Myers et al., 2008, p. 303) or in the case of this study, the relative increase should be considered in collapsings where they were higher.

One argument against the 3-CAT formats is that odd-numbered answer choices give the respondent an option to not have to make a decision (Fisher, 2006). This can result in larger than normal responses in the middle or neutral category. The post-hoc category frequency results suggest the middle category would have 40-49% of the total observations in the 3-CAT format for 111223 and 112223. Ideally this amount would be similar across all of the categories at 33% to have more balanced responses (Zhu et al.,

1997). While the middle category frequency for the 122233 collapsing shows 33%, the third category is 61% making this option skewed and perhaps less than ideal.

It would be appropriate to perform a confirmatory study to assess the frequency distributions across the categories of the remaining, more promising collapsings using a similar sample. The 111223, 112223 options could linguistically be reconfigured into a 3-CAT *Always, Sometimes,* or *Never* rating scale. Those with experience in screening for and diagnosing eating disorders may find other qualitative labels more appropriate. This should also be assessed empirically.

Rating scales with a larger number of categories provide more information if the categories function as intended. The data suggest that the second and third category choices (*Often* and *Sometimes*) in the 4-CAT format were distinguishable for the respondent providing further evidence that the 4-CAT format may be more effective than the 3-CAT option. The post-hoc frequency distributions of 6%, 15%, 40%, and 39% across categories are again not similar, however they present some improvement on the 6-CAT format (6%, 7%, 8%, 19%, 22%, and 39%). For category labels, one possibility is an *Always, Often, Seldom* (or *Sometimes*), or *Never* rating scale which seems to fit linguistically with the new category combinations. A future confirmatory study should take these findings into account and administer the various options to a similar sample to determine the most effective scale.

Strengths of this study include a reasonable sample size to determine effective Rasch statistics and a rigorous analysis of the EAT-26 using modern measurement theory. This study has identified problematic issues with disordered rating scale thresholds. This issue, once rectified, may improve the ability of the EAT to effectively

screen for eating disorders. One limitation of this study is the absence of confirmed cases of anorexia nervosa, bulimia nervosa, or binge eating disorder that could substantiate the effectiveness of the EAT-26 at detecting eating disorder risk. However, because the EAT-26 has been used and is recommended for non-clinical samples, the determination was made to use it among a college student, adult sample. The prevalence of eating disorders is low in this population. One study determined that "in a sample of 100 college-age women, about 25 could be classified as high risk. Of these 25, eight or nine would be at very high risk, and of these, two or three would develop an eating disorder for an incidence rate of 2–3%" (Jacobi et al., 2011, p. 1946). The National College Health Assessment (NCHA) for this university sample found that out of 1,274 students, 3 males and 1 female indicated that they had been diagnosed and/or treated in the last 12 months for anorexia nervosa and 4 males and 3 females were treated for bulimia for a total prevalence of .8% (L. Schrader, personal communication, July 7, 2014). These were the only two eating disorders assessed by the NCHA survey. Due to these low amounts, this current study did attempt to oversample for those who may have eating disorder risk by choosing courses for participation with students in nutrition, health, and exercise science. The person separation index statistic of 2.60 does provide evidence that there was enough spread in the sample for effective Rasch analysis along with a strata analysis which shows nearly five (4.8) groups of person ability can be determined. Additionally, nearly 11% of the sample scored above the cut-off score of 20. This is similar to general findings of 10-15% of females scoring high and 2-3% of males scoring high on the EAT-26 previously (Garfinkel & Newman, 2001).

Further study is needed to provide evidence of the EAT-26 scale with a new three or four-category structure perhaps combining a non-clinical sample with clinical subjects for comparison. To confirm the findings of this current study, a confirmatory analysis should be performed with data from a sample similar to what was used for this exploratory study.

REFERENCES

Al-Subaie, A. S. (1998). Eating Attitudes Test in Arabic: Psychometric features and normative data. *Saudi Medical Journal, 19*(6), 769-775.

Anderson-Fye, E. P., & Lin, J. (2009). Belief and behavior aspects of the EAT-26: The case of schoolgirls in Belize. *Culture, Medicine and Psychiatry, 33*(4), 623-638. doi:10.1007/s11013-009-9154-2

Andrich, D. A. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594. doi:10.1177/014662167800200413

Andrich, D. A. (1996). Category ordering and their utility. *Rasch Measurement Transactions, 9*(4), 465-466.

Andrich, D. A. (2010). Understanding the response structure and process in the polytomous Rasch model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models*, (pp. 123-152). New York, NY: Routledge.

Anthi, L. L., Maria, K., Nicole, K. M., Evel, N., & Kalia, L. (2011). Detection of disordered eating in high-risk samples using the Greek version of EAT-26. *Psychology & Health, 26*(2), 164-167.

Bond, T., & Fox, C. (2015). *Applying the Rasch model, fundamental measurement in the human sciences* (Third ed.). New York, NY: Routledge.

Choudry, I. Y., & Mumford, D. B. (1992). A pilot-study of eating disorders in Mirpur (Pakistan) using an Urdu version of the Eating Attitudes Test. *International*

Journal of Eating Disorders, 11(3), 243-251. doi:10.1002/1098-

108x(199204)11:3<243::aid-eat2260110307>3.0.co;2-p

Doninger, G. L., Enders, C. K., & Burnett, K. F. (2005). Validity evidence for eating

attitudes test scores in a sample of female college athletes. *Measurement in*

Physical Education and Exercise Science, 9(1), 35-49.

doi:10.1207/s15327841mpee0901_3

EAT-26. (2011). Retrieved from http://www.eat-26.com/index.php

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social,*

behavioral, and health sciences. New York, NY, US: Routledge.

Fisher, W. P. (2006). Survey Design Recommendations. *Rasch Measurement*

Transactions, 20(3), 1072-1074.

Garfinkel, P. E., & Newman, A. (2001). The Eating Attitudes Test: Twenty-five years

later. *Eating and Weight Disorders, 6*(1), 1-24.

Garner, D. M., & Garfinkel, P. E. (1979). Eating Attitudes Test: Index of the symptoms

of Anorexia-Nervosa. *Psychological Medicine, 9*(2), 273-279.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes

Test: Psychometric features and clinical correlates. *Psychological Medicine,*

12(4), 871-878.

Gross, J., Rosen, J. C., Leitenberg, H., & Willmuth, M. E. (1986). Validity of the Eating

Attitudes Test and the Eating Disorders Inventory in Bulimia Nervosa. *Journal of*

Consulting and Clinical Psychology, 54(6), 875-876. doi:10.1037/0022-

006X.54.6.875

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill.

Jacobi, C., Fittig, E., Bryson, S. W., Wilfley, D., Kraemer, H. C., & Taylor, C. B. (2011). Who is really at risk? Identifying risk factors for subthreshold and full syndrome eating disorders in a high-risk sample. *Psychological Medicine, 41*(9), 1939-1949. doi:10.1017/s0033291710002631

Kang, M., Zhu, W. M., Ragan, B. G., & Frogley, M. (2007). Exercise barrier severity and perseverance of active youth with physical disabilities. *Rehabilitation Psychology, 52*(2), 170-176. doi:10.1037/0090-5550.52.2.170

Lane, H. J., Lane, A. M., & Matheson, H. (2004). Validity of the Eating Attitude Test among exercisers. *Journal of Sports Science and Medicine, 3*(4), 244-253.

Le Grange, D., Swanson, S. A., Crow, S. J., & Merikangas, K. R. (2012). Eating disorder not otherwise specified presentation in the US population. *International Journal of Eating Disorders, 45*(5), 711-718. doi:10.1002/eat.22006

Linacre, J. M. (1999a). Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Measurement Transactions, 13*(1), 675.

Linacre, J. M. (1999b). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*(2), 103-122.

Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 1.

Linacre, J. M. (2004). Rasch model estimation: further topics. *Journal of Applied Measurement, 5*(1), 95-110.

Linacre, J. M. (2016). Rasch measurement forum - Disordered thresholds and misfit

[Comment]. Retrieved from http://raschforum.boards.net/thread/438/disordered-

thresholds-misfit-rsm

Masters, G. N. (1980). *A Rasch model for rating scales* (Order No. T-27575). Available

from ProQuest Dissertations & Theses Global. (303091363). Retrieved from

https://ezproxy.mtsu.edu:3443/login?url=http://search.proquest.com/docview/303

091363?accountid=4886

Myers, N. D., Feltz, D. L., & Wolfe, E. W. (2008). A confirmatory study of rating scale

category effectiveness for the Coaching Efficacy Scale. *Research Quarterly for

Exercise and Sport, 79*(3), 300-311.

Neumarker, K. J., Bettle, N., Bettle, O., Dudeck, U., & Neumarker, U. (1998). The Eating

Attitudes Test: Comparative analysis of female and male students at the Public

Ballet School of Berlin. *European Child and Adolescent Psychiatry, 7*(1), 19-23.

Nunnally, J. C. (1967). *Psychometric theory.* New York, NY: McGraw-Hill.

Orbitello, B., Ciano, R., Corsaro, M., Rocco, P. L., Taboga, C., Tonutti, L., . . .

Balestrieri, M. (2006). The EAT-26 as screening instrument for clinical nutrition

unit attenders. *International Journal of Obesity, 30*(6), 977-981.

doi:10.1038/sj.ijo.0803238

Pereira, A. T., Maia, B., Bos, S., Soares, M. J., Marques, M., Macedo, A., & Azevedo, M.

H. (2008). The Portuguese short form of the Eating Attitudes Test-40. *European

Eating Disorders Review, 16*(4), 319-325. doi:10.1002/erv.846

Royal, K. D., Ellis, A., Ensslen, A., & Homan, A. (2010). Rating scale optimization in survey research: An application of the Rasch Rating Scale Model. *Journal of Applied Quantitative Methods, 5*(4), 586-596.

Smink, F. R. E., van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: Incidence, prevalence and mortality rates. *Current Psychiatry Reports, 14*(4), 406-414. doi:10.1007/s11920-012-0282-y

Smith, A., Rush, R., Fallowfield, L., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*(1), 1. doi:10.1186/1471-2288-8-33

Stone, M., & Wright, B. D. (1994). Maximizing rating scale information. *Rasch Measurement Transactions, 8*(3), 386.

Szabo, C. P., & Allwood, C. W. (2004). Application of the Eating Attitudes Test (EAT-26) in a rural, Zulu speaking, adolescent population in South Africa. *World Psychiatry, 3*(3), 169-171.

Tury, F., Guelec, H., & Kohls, E. (2010). Assessment methods for eating disorders and body image disorders. *Journal of Psychosomatic Research, 69*(6), 601-611. doi:10.1016/j.jpsychores.2009.05.012

WINSTEPS. (2012). Rasch Software. Retrieved from http://www.winsteps.com/winsteps.htm

Wright, B. D., & Linacre, J. M. (1992a). Combining (collapsing) and splitting categories. *Rasch Measurement Transactions, 6*(3), 233-235.

Wright, B. D., & Linacre, J. M. (1992b). Disordered steps? *Rasch Measurement Transactions, 6*(2), 225.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis, Rasch measurement.* Chicago, IL: MESA Press.

Zhu, W. M., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement, 1*(4), 286-304.

CHAPTER III: RATING SCALE CATEGORY EFFECTIVENESS OF THE

EATING ATTITUDES TEST (EAT-26): A CONFIRMATORY STUDY

UTILIZING RASCH ANALYSIS

The Eating Attitudes Test 26 (EAT-26) is a widely utilized measurement tool to assess the risk of having an eating disorder (Garfinkel & Newman, 2001; Garner, 1993; Mintz & O'Halloran, 2000). As a screening tool, there are three portions of the EAT-26 that are evaluated to determine whether a person should be referred to a health care practitioner for possible diagnosis. These include 26 statements with a Likert six category (6-CAT) rating scale, determination of a very underweight body mass index (BMI), and behavioral health questions to assess disordered eating. Screening positively in any of the three areas would warrant a referral. Each of the 26 items on the first portion of the screen include response categories of *Always, Usually, Often, Sometimes, Rarely,* and *Never.* Item 26, "Enjoy trying rich new foods" is the only item that is reverse scored. The Likert rating scale functions properly when the categories are easy to differentiate by the respondents (Linacre, 2002). For instance, does a respondent successfully distinguish between *Usually* and *Often* or *Sometimes* and *Rarely* in the EAT-26?

Studies have previously assessed the EAT-26 items utilizing traditional statistics and models (Garner, Olmsted, Bohr, & Garfinkel, 1982; Gross, Rosen, Leitenberg, & Willmuth, 1986; Mintz & O'Halloran, 2000; Ocker, Lam, Zhang, Jackson, & Pease, 2002). Many of these studies relied on the total scores of the items. This can result in a misinterpretation of the data (Anshel, Weatherby, Kang, & Watson, 2009; Bond & Fox, 2015; Wright & Linacre, 1989). As best said by Andrich, (2011) "The handling of ratings has ranged from the very elementary to the highly sophisticated. In an elementary form,

and assumed in classical test theory, the ratings are scored with successive integers and treated as measurements; in a sophisticated form, and used in modern test theory, the ratings are characterized by probabilistic response models with parameters for persons and the rating categories" (p. 1). When ordinal, raw scores are statistically analyzed as interval measures, the distances between each item are not considered identical. It is critical that eating disorder researchers utilize statistical analyses which are appropriate and sound (Berkman, Lohr, & Bulik, 2007). The Rasch probabilistic model for rating scales (RSM) provides a method for turning ordinal data into interval data and then allowing analysis of the categorical rating scale to determine if it works as intended (Wright & Masters, 1982).

A previous exploratory study of the EAT-26 rating scale utilizing Rasch analysis (Cook, Weatherby, Kang, Colson, & Owusu, 2016) found that a four category (4-CAT) or three category (3-CAT) rating scale may be more beneficial than the original six category (6-CAT) structure. The original 6-CAT EAT-26 did not meet all of Linacre's guidelines for rating scale category effectiveness (Linacre, 2002). Categories did not advance monotonically and disordered thresholds were observed. Disordered thresholds are a sign that there is a problem with the function of the rating scale (Andrich, 2011). One potential resolution to disordered thresholds is to combine or collapse problematic categories that respondents cannot distinguish into one. For example, a 6-CAT (123456) structure can be combined into a 3-CAT (122223) structure by combining categories 2, 3, 4 and 5. Post-hoc analyses of 25 possible datasets with combined (collapsed) categories were assessed. Of the possible combinations, one 4-CAT (122334) and two 3-CAT (112223 and 111223) combinations had better Rasch model fit when compared to the

original 6-CAT. It is not enough to only perform post-hoc adjustments of data. They must be experimentally investigated as well (Andrich, 2011).

Data from the previous exploratory study (Cook et al., 2016) led to an additional inquiry to perform an analysis of the categorical function of the rating scale for both the 4-CAT and 3-CAT EAT-26 rating scales using data from a similar sample. This confirmatory method is used "when previous research has provided evidence for the ineffectiveness of the original rating scale and a more effective post hoc structure, then the post hoc structure identified in the previous study should be tested to provide confirmatory, cross-validation evidence for this scale" (Myers, Feltz, & Wolfe, 2008, p. 301).

Purpose of the Study

The purpose of this study is to analyze the effectiveness of the newly developed 3 and 4-CAT rating scale options. The primary research question is Do the collapsed three and four category rating scales of the EAT-26 suggested in previous research provide better model fit than the original six category scale with data from a similar sample?

Methods

The community of a public university in the southeastern U.S. was invited to participate in this study with the goal of at least 200 participants per survey according to power analysis. In an effort to maximize the range of participants, the researchers attempted to reach a wide variety of class-years as well as class-types such as beginning health, nursing, exercise science, nutrition, social work, and sociology. As an incentive, some professors offered students extra credit points during the semester for completing

the surveys. Non-students who participated did not receive any incentive. The only inclusion criterion was that each participant had to be at least 18 years of age.

The Institutional Review Board provided approval for all study procedures. As part of IRB approval, participants were notified that their inclusion in the study was completely voluntary and that they could cease participation at any time without consequence. Data was collected online via Survey Monkey with no identifiable information included. A unique number identifier was used for each respondent. The inclusion criterion was that each participant was at least 18 years of age. The researchers informed the participants through flyers and announcements that the purpose of the study was to evaluate measurement tools for eating behaviors. Demographic data including gender, race/ethnicity, age, and BMI calculated from height and weight was also collected via self-report.

A convenience sample of 777 males and females volunteered to participate in the study (Table 1). Fifty three participants were excluded due to missing or obvious patterned data (responding 1,4,1,4,1,4 or 1,2,3,4,1,2,3,4) throughout resulting in a total sample size of 632 ($n = 322$ for the 3-CAT survey and $n = 310$ for the 4-CAT survey). These were similar sample sizes as the first exploratory study ($n = 315$).

Table 1

Demographic Information of Participants

	3-CAT *n = 322*		4-CAT *n = 310*	
Characteristic	*n*	%	*n*	%
Gender				
Women	215	66.77	219	70.65
Men	107	33.23	91	29.35
Race/Ethnicity				
White	186	57.76	192	64.00
Black	93	28.88	65	21.67
Hispanic	10	3.11	10	3.33
Asian	9	2.80	10	3.33
Other	16	4.97	23	7.67
	M	*SD*	*M*	*SD*
Age	23.07	7.87	22.98	7.21
BMI				
Women	25.07	5.64	24.87	4.81
Men	25.93	4.71	26.84	5.07

Note. Race/Ethnicity for 3-CAT (*n* = 314) and for 4-CAT
(*n* = 300)

Measure.

The category labels or rating scale labels such as *Always* and *Never* are qualitative representations of a scale that help respondents imply meaning (Royal, Ellis, Ensslen, & Homan, 2010). The step counts from category 1 to category 2 and so on...are always equal regardless of the labels (e.g., *None*, *Some*, *All* or *None*, *A little*, *All*) (Wright & Linacre, 1989). Yet changing the label of a category "...will change its meaning psycholinguistically, and so the amount of the underlying variable it represents..." (Linacre, 2002, p. 88).

The combined category rating scale possibilities from the previous study of the six category (6-CAT) EAT-26 required that new category labels be created. The 4-CAT collapsing (122334) suggested that *Usually* and *Often* be combined and that *Sometimes* and *Rarely* be combined leaving both of the rating scale anchors *Always* and *Never* unchanged. The first 3-CAT collapsing (111223) combined *Always* with *Usually* and *Often,* and then *Sometimes* with *Rarely* while the second 3-CAT collapsing (112223) suggested combining *Always* with *Usually,* and then *Often* with *Sometimes* and *Rarely*. Several new category labels to reflect these category combinations were assessed by clinical experts with experience in screening and diagnosing eating disorders. The determination was made for a 4-CAT rating scale with *Always, Often, Seldom,* and *Never* and a 3-CAT rating scale with the labels *Always, Sometimes,* and *Never*.

Data analysis.

The Rasch computer program WINSTEPS was used to evaluate the EAT-26 instrument with collapsed 3-CAT and 4-CAT rating scale formats (WINSTEPS, 2012).

Several diagnostic statistical tools, also utilized in the first exploratory study, were again used to determine an effective rating scale.

Each rating scale was assessed for its functioning using a selection of Linacre's guidelines for optimizing rating scale category effectiveness (Linacre, 2002). These guidelines were also used in the previous study. Each assessed category format was then compared to the original EAT-26 format. These guidelines included 1) average measures that advance monotonically with category, 2) Outfit mean squares < 2.0, and 3) step difficulties (thresholds between categories) that advance. Andrich (1996) stated that thresholds "should be ordered and increase along the continuum" (p. 2). The final two guidelines are 4) that the step difficulties (thresholds) advance by at least 1.4 logits, and 5) thresholds advance by less than 5.0 logits. Linacre (2002) stated thresholds should advance by 1.4 logits for three category (3-CAT) scales and 1.0 logits for 5-CAT scales. As in the previous study, the 4-CAT separation of 1.1 logits is used based on the Myers et al. (2008, p. 305) study which cited evidence from Huynh (1994).

Category frequencies were considered for even distribution of the categories as suggested by Zhu, Updyke, and Lewandowski (1997) and Linacre (2002). While proper balance of responses across categories is considered ideal, it may be difficult to achieve with this population since eating disorders occur infrequently in non-clinical settings resulting in a skewed response (Garfinkel & Newman, 2001). However, category frequency information can help to guide researchers toward potential problems with categories. The person separation indexes were evaluated and considered acceptable if they were at least 2.0 as "the person separation index and the person separation reliability should be important indicators of effective rating scale structure" (Myers et al., 2008).

The model data fit was evaluated with both Infit and Outfit statistics provided by Rasch analysis. These statistics are mean squares residuals as chi-square values divided by degrees of freedom (Anshel et al., 2009; Linacre & Wright, 1994). With an ideal value of 1.0, this study utilized a recommendation by Wright and Linacre for values between .5 and 1.5 as being acceptable for measurement (Wright & Linacre, 1994).

Results

All rating scale combinations displayed average measures that advanced monotonically with category, and acceptable fit with Outfit mean squares < 2.0. Table 2 provides summary data for the rating scales against Linacre's threshold guidelines. For both the new 3-CAT and 4-CAT options, ordered thresholds advanced by the correct amounts which did not occur with the original 6-CAT scale. This is a sign that the new rating scales are functioning as they should and that respondents understand the qualitative labels provided.

Table 2

Comparison of Threshold Guidelines for Rating Scale Effectiveness Among
6-CAT, 4-CAT, and 3-CAT Rating Scales

| Guideline	6-CAT	4-CAT	3-CAT
Step Difficulties Advance	disordered	ordered	ordered
SD Advance by $\geq 1.4(3), \geq 1.1(4), \geq 1.0(5)$	No	Yes	Yes
Person Separation	2.60	2.60	2.10
Person Reliability	0.87	0.87	0.82

Note. All combinations met guidelines 1 and 2.

CAT = Category, SD = Step Difficulties (thresholds)

Visually this can be observed in Figure 1 which provides a comparison of the category probability curves for the 6-CAT, 4-CAT, and 3-CAT options. The categories in the 4 and 3-CAT formats advance monotonically and have modal peaks which show each category has a likelihood of being selected over any other category.

The 4-CAT rating scale format expressed the same person separation and person reliability (2.60, 0.87) as the original 6-CAT structure (2.60, 0.87). These 4-CAT parameters were lower than what was expected in the post-hoc analysis from the previous study (2.76, 0.88). The 3-CAT format also showed a lower person separation and person reliability (2.10, .82) than what was suggested in the post-hoc analysis of the previous study (2.76-2.81, 0.88-0.89).

Table 3 provides additional detail on category frequency and function (average measures and threshold estimates). While the 4-CAT and 3-CAT rating scales offer an improvement on category frequencies, which should be relatively equal or balanced across categories (Zhu et al., 1997), the first category responses are much lower than the others at 6% and 9%.
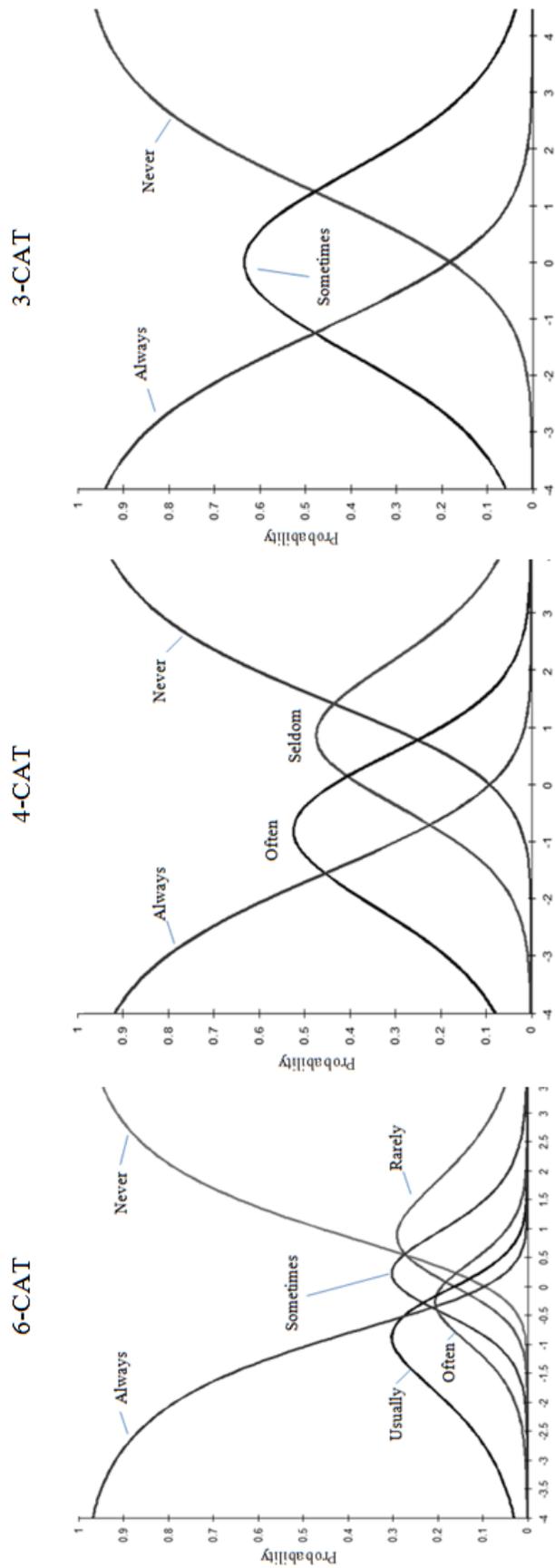
Figure 1. Category function for the original EAT-26 6-CAT rating scale and the new 4-CAT and 3-CAT rating scales.

Table 3

Detail of Category Frequency and Function

Collapsing	Category frequency (%)						Average measure						Threshold estimate				
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5
6-CAT	6	7	8	19	22	39	-0.58	-0.03	0.20	0.53	0.8	1.4	-0.53	-0.16	-0.41	0.56	0.53
4-CAT	6	20	33	41			-0.45	0.15	0.98	2.32			-1.55	0.14	1.41		
3-CAT	9	38	53				-0.19	0.8	2.38				-1.25	1.25			

Note. Frequency percentages as reported in Rasch output may not equal 100% due to rounding. CAT = Category

The new 4-CAT and 3-CAT scales displayed fewer misfit items than the original 6-CAT rating scale. The summary of misfit items is displayed in Table 4. The original 6-CAT rating scale had 85% satisfactory fit statistics (85% Infit and 85% Outfit). The 4-CAT scale had 92% satisfactory fit statistics (92% Infit and 92% Outfit) while the 3-CAT scale had 96% satisfactory fit statistics (100% Infit and 92% Outfit). Both the 4-CAT (1.01, 1.03) and 3-CAT (1.01, 1.01) scales had Infit and Outfit means closer to 1.00 than the original 6-CAT (1.04, 1.12) rating scale structure.

Table 4

Comparison of Misfit Among 6-CAT, 4-CAT, and 3-CAT Rating Scales

CAT	Infit				Outfit				Infit		Outfit	
	Fit		Not fit		Fit		Not fit		Mean	SD	Mean	SD
	n	%	*n*	%	*n*	%	*n*	%				
6-CAT	22	85%	4	15%	22	85%	4	15%	1.04	0.34	1.12	0.57
4-CAT	24	92%	2	8%	24	92%	2	8%	1.01	0.28	1.03	0.39
3-CAT	26	100%	0	0%	24	92%	2	8%	1.01	0.22	1.01	0.39

Note. CAT = categories, SD = standard deviation.

Discussion

      The purpose of this study was to build on the inquiry of the function of the Eating

Attitudes Test (EAT-26) rating scale in a previous study by performing a confirmatory

analysis of the categorical function of the rating scale for both three and four response

options with the EAT-26.

      The four category (4-CAT) option of the rating scale with the labels of *Always,*

Often, Seldom, and *Never* may provide better category function than the original six

category (6-CAT) EAT-26. The greatest improvement was made with ordered thresholds

in the 4-CAT format versus disordered thresholds in the 6-CAT scale. The person

separation and person reliability statistics were not as high as expected from previous

post-hoc analysis, but the 4-CAT rating scale did perform as well as the original EAT-26

(both were 2.60, .87). Zhu et al. (1997) observed in their study that as the number of

categories increased, item and person separation statistics increased as well. For person

separation to remain unchanged when categories decreased may be a good indicator for

the functionality of the 4-CAT scale.

      While the 3-CAT scale displayed better model data fit than both the 4-CAT and 6-

CAT options, the category frequencies were problematic with over 50% selecting the

third category. This finding, in addition to a much lower person separation index (2.10,

.82) than what was predicted with post-hoc analysis (2.76-2.81, 0.88-0.89) may not make

it an ideal option as written. The linguistic challenges of collapsing categories are

difficult. While the eating disorder experts consulted preferred rating scales to have

Always to counteract the opposite *Never* label which was required and unchanged, it is

possible that a label such as *Seldom* may perform better in the 3-CAT format than

Sometimes which was selected to be used for the middle category. A future study may wish to investigate different category labels (i.e., "*Always – Seldom – Never*" versus "*Always – Sometimes – Never*") to reflect very subtle differences between the 111223 and 112223 collapsings.

In this study, the fact that the categories became ordered after the implementation of the new rating scales suggests poor rating scale category effectiveness in the original EAT-26. The strengths of this study are improvement of the 6-CAT disordered threshold rating scale of the EAT-26 to the new 4-CAT ordered rating scale using modern measurement Rasch analysis. The confirmatory study shows that the ordered categories suggested in a previous post-hoc analysis remain when applied to a similar sample.

Perhaps due to the online nature of the data collection in this study versus the original study and the opportunity to obtain extra credit in some courses, more persons required removal from the analysis due to obvious patterned data responses. The Rasch methodology has the capability to handle some guesses in the data but where patterns were distinct, those persons were removed. Despite this limitation, the makeup of the samples was similar in both size and demographically.

Finally, further study of the 4-CAT rating scale is suggested with both clinical (diagnoses) and non-clinical samples to determine if answer choices are more understandable for those persons being screened while still remaining effective at distinguishing between those with and without a risk of having an eating disorder. Additionally, a calibration of the 4-CAT scale using Rasch analysis should be conducted to better understand the item hierarchy of the assessment, where items may be redundant, and where more items may be beneficial.

REFERENCES

Andrich, D. A. (1996). Category ordering and their utility. *Rasch Measurement Transactions, 9*(4), 465-466.

Andrich, D. A. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics and Outcomes Research, 11*(5), 571-585. doi:10.1586/erp.11.59

Anshel, M. H., Weatherby, N. L., Kang, M., & Watson, T. (2009). Rasch calibration of a unidimensional perfectionism inventory for sport. *Psychology of Sport and Exercise, 10*(1), 210-216. doi:10.1016/j.psychsport.2008.07.006

Berkman, N. D., Lohr, K. N., & Bulik, C. M. (2007). Outcomes of eating disorders: A systematic review of the literature. *International Journal of Eating Disorders, 40*(4), 293-309. doi:10.1002/eat.20369

Bond, T., & Fox, C. (2015). *Applying the Rasch model, fundamental measurement in the human sciences* (Third ed.). New York, NY: Routledge.

Cook, A. R., Weatherby, N., Kang, M., Colson, J., & Owusu, A. (2016). An exploratory study of the psychometric properties and category functions of the Eating Attitudes Test (EAT-26) utilizing Rasch analysis (Unpublished doctoral dissertation). Middle Tennessee State University, Murfreesboro, TN.

Garfinkel, P. E., & Newman, A. (2001). The Eating Attitudes Test: Twenty-five years later. *Eating and Weight Disorders, 6*(1), 1-24.

Garner, D. M. (1993). Self-report measures for eating disorders: A citation-classic commentary on the Eating Attitudes Test, an index of the symptoms of Anorexia-

Nervosa by Garner, D.M., and Garfinkel, P.E., and development and validation of a multidimensional eating disorder inventory for Anorexia-Nervosa and Bulimia by Garner, D.M., Olmsted, M.P., and Polivy. *Current Contents/Social and Behavioral Sciences, 8*, 8.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871-878.

Gross, J., Rosen, J. C., Leitenberg, H., & Willmuth, M. E. (1986). Validity of the Eating Attitudes Test and the Eating Disorders Inventory in Bulimia Nervosa. *Journal of Consulting and Clinical Psychology, 54*(6), 875-876. doi:10.1037/0022-006X.54.6.875

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M., & Wright, B. D. (1994). Dichotomous mean-square fit statistics. *Rasch Measurement Transactions, 8*(2), 360.

Mintz, L. B., & O'Halloran, M. S. (2000). The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment, 74*(3), 489-503.

Myers, N. D., Feltz, D. L., & Wolfe, E. W. (2008). A confirmatory study of rating scale category effectiveness for the Coaching Efficacy Scale. *Research Quarterly for Exercise and Sport, 79*(3), 300-311.

Ocker, L. B., Lam, E. T. C., Zhang, J. J., Jackson, A. S., & Pease, D. G. (2002).

    Confirmatory factor analysis of the Eating Attitudes Test. *Research Quarterly for*

    Exercise and Sport, 73(1), A40-A41.

Royal, K. D., Ellis, A., Ensslen, A., & Homan, A. (2010). Rating scale optimization in

    survey research: An application of the Rasch Rating Scale Model. *Journal of*

    Applied Quantitative Methods, 5(4), 586-596.

WINSTEPS. (2012). Rasch Software. Retrieved from

    http://www.winsteps.com/winsteps.htm

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal - measurements,

    however, must be interval. *Archives of Physical Medicine and Rehabilitation,*

    70(12), 857-860.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch*

    Measurement Transactions, 8(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis, Rasch measurement*.

    Chicago, IL: MESA Press.

Zhu, W. M., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of

    optimal categorization of an ordered-response scale. *Journal of Outcome*

    Measurement, 1(4), 286-304.

CHAPTER IV: CALIBRATION OF THE FOUR CATEGORY LIKERT SCALE

EATING ATTITUDES TEST (EAT-26)

Eating disorders among college students continue to be a pressing health issue. One study found "the prevalence of dieting and disordered eating behaviors was high and either remained constant or increased from adolescence to young adulthood. Of particular concern was the large increase in extreme weight control behaviors among youth transitioning from adolescence to young adulthood" (Neumark-Sztainer, Wall, Larson, Eisenberg, & Loth, 2011, p. 1008). Understanding the nature of disordered eating symptomology and being able to effectively measure the risk is critical to identifying those in need and providing treatment. One of the most cited and most widely used screening tools for eating disorder risk is the 26 item Eating Attitudes Test (EAT-26) (Garner, Olmsted, Bohr, & Garfinkel, 1982).

Originally designed to screen for anorexia nervosa in 1979, a longer version of the Eating Attitudes Test with 40 items (EAT-40) included criteria for having symptoms of the disease prior to age 25, losing at least 25% of one's original body weight, and distorted, unrealistic attitudes about weight, eating, and food regardless of hunger or logic (Feighner et al., 1972). During this period, bulimia nervosa was not designated as its own disorder. It was not until the DSM-IV came out in 1994 that anorexia nervosa and bulimia nervosa were considered separate diseases with their own criteria for diagnosis. When the EAT-26 was first created, there was no concept of atypical or sub threshold eating disorders (Mintz & O'Halloran, 2000). The nature of eating disorders falls along a continuum that is changing due to societal norms and culture (Garfinkel & Newman,

2001). Several changes in diagnostic criteria over the years may also alter the way the EAT-26 items function in current populations.

The Rasch model is an advanced measurement, item response theory that creates interval scale data from ordinal data using logits, the unit of measure which calibrates the facets (i.e., items and persons) (Rasch, 1960, 1980). The model allows researchers to take vigorous linear measures from ordinal data and apply statistical analysis with which Rasch provides several parameters to consider (Zhu, Updyke, & Lewandowski, 1997).

The EAT-26 has been assessed using classical test theory (CTT) (Garner et al., 1982; Gross, Rosen, Leitenberg, & Willmuth, 1986). With CTT, a person's risk for having an eating disorder is determined by a total summed score derived from all of the responses to the EAT-26 items. However, there are many criticisms to this approach. One is the inability to control for the difficulty of the items. The other is that ordinal data should not be used as a total summed score for effective measurement (Bond & Fox, 2015; Brinthaupt & Kang, 2014, S. J. Kang & M. Kang, 2006). The Rasch calibration model overcomes these limitations and allows the researcher to investigate the relationship between item "difficulty" and person "ability" on one scale. This allows the research to see the spread of persons and items, areas where there are gaps in content, and areas where there may be redundancy in items (Brinthaupt & Kang, 2014)

Purpose of the Study

Data from the previous Rasch rating scale category effectiveness confirmatory study (Cook, Weatherby, Kang, Colson and Owusu, 2016) led to additional inquiries to perform a calibration of a new four category (4-CAT) version of the Eating Attitudes

Test. The purpose of this study is to calibrate and observe the measurement qualities of the newly developed rating scale option identified in the previous study.

Methods

A sample of 310 respondents from a public university in the south was used for the EAT-26 calibration. This is the same sample used in the previous confirmatory study which suggested a 4-CAT Likert scale may be more optimal than the original 6-CAT scale. Additional information regarding methodology for data collection can be referenced in the previous study. The sample included 219 (70.65%) women with a mean body mass index of 24.87 (± 4.81) and 91 (29.35%) men with a mean body mass index of 26.84 (±5.07). Race/ethnicity of the sample was 64% white, 22% Black, 3% Hispanic, 3% Asian, and 8% other. The average age was 22.98 (±7.21).

Data analysis.

Rasch analysis of the four category (4-CAT) EAT-26 instrument was performed with the computer program WINSTEPS version 3.75 (WINSTEPS, 2012). EAT-26 data were coded with 6= *Always* to 1 = *Never.* A two-facet Rasch model was estimated including EAT-26 item and person parameters. The Rasch model was defined with the formula

$$Ln\ [P_{njk}/1 - P_{nj(k-1)}] = D_n - C_j - F_k$$

where P_{njk} is the probability of an EAT-26 item n being endorsed k category by person being screened j; $P_{nj(k-1)}$ is the probability of an EAT-26 item n being endorsed $k-1$ category by person being screened j; D_n is the level of the EAT-26 item n; C_j is the ability of the person being screened j, and F_k is the difficulty of category step k . An item response is determined by the item difficulty (i.e., "difficulty" or severity level of eating

disorder items) and person ability (i.e., the extent to which an individual may be at risk

for having an eating disorder) and are expressed in logit scores. Several steps are

included in the Rasch calibration process. This study's methodology for calibration of the

EAT-26 with a four category Likert scale was derived and influenced by previous works

by Anshel, Kang, & Jubenville (2012), M. Kang, Zhu, Ragan, & Frogley (2007), and S. J.

Kang & Kang (2006).

Model-data fit.

Model-data fit was evaluated for EAT-26 items using Infit and Outfit statistics.

Two fit statistics, referred to as Infit and Outfit, help to describe the fit of the items to the

Rasch model. Infit is the information-weighted fit and Outfit is the standardized

unweighted mean square (Bond & Fox, 2015; Zhu et al., 1997). Infit and Outfit statistics

have an expected value of 1 and can range from 0 to infinity. Infit statistics are sensitive

to irregular responses where items and persons are well-targeted while Outfit statistics are

sensitive to items and person across the entire range of items and person locations.

Values which are above 1 may have more variation between the model (expected scores)

and the observed scores while values below 1.00 have less variation (Engelhard, 2013). A

fit statistic valued at 1.50 would indicate 50% more variation than what was predicted by

the model, or what is also referred to as underfit. An item with a fit statistic of .50 would

mean 50% less variation also referred to as overfit (Bond & Fox, 2015; Smith, Rush,

Fallowfield, Velikova, & Sharpe, 2008). A few acceptable fit ranges include options for

multiple choice tests (0.7-1.3), clinical observations (0.5-1.7), or rating scales (0.6-1.4)

(Bond & Fox, 2015). For this study, any Infit or Outfit statistic less than 0.5 or greater

than 1.5 warranted further consideration (Linacre, 2002).

Item-person map distribution.

The item-person map distribution allows for item difficulty and person ability to be displayed on a single graph with a shared scale based on logit units of measurement. This allows the positions of item difficulties and those screened to be easily examined visually and to note any gaps in targeting.

Item difficulty.

Item difficulty parameters, expressed in logits, were calibrated. The larger the logit score, the more difficult it was to agree with the item, making it more severe. These could be items related to more extreme behaviors such as binging and purging. Lower logit scores could be interpreted as items pertaining to disordered eating having less difficulty or severity.

Person ability.

The individual's person ability level or their risk of having an eating disorder, was estimated. Screened persons with higher logit scores were most able to agree with EAT items and showed higher risk for disordered eating. Persons with lower (negative) logit scores were most able to disagree with the EAT items and were at a lesser risk for disordered eating.

Item and person separation indexes and separation-reliability indexes.

The item and person separation indexes and separation-reliability indexes were observed. The separation indexes determine if the EAT items and the screened persons are spread out enough along the scale. The separation-reliability indexes determine whether the items would appear in the same order and within the measurement error if screening with another sample, and whether there would be separation of persons along

the measurement scale into similar groups for another sample of items (S. J. Kang & Kang, 2006; Wright & Masters, 1982). A high separation index > 2.0 (Bond & Fox, 2015) and separation-reliability index near 1.00 implies good discrimination for items or persons along the measurement scale (Fisher, 1992).

Differential item functioning.

Differential item functioning (DIF) analyses helped with understanding if items in the EAT functioned differently for each gender. Proper Rasch measurement requires invariance of the item estimates. Items with Mantel-Haenszel (M-H) DIF size > 0.64 logits and statistical significance ($p < .00$) were considered to be biased and showed the item functioned differently between men and women (Zwick, Thayer, & Lewis, 1999); cited from (Anshel et al., 2012; Linacre, 2016).

Results

Model-data fit was evaluated for EAT-26 items using Infit and Outfit statistics. Item 13, "Other people think I am too thin" (Infit = 1.72, Outfit = 2.29), Item 26, "I enjoy trying rich new foods" (Infit = 1.73, Outfit = 2.00), and Item 8, "I feel that others would prefer if I ate more" (Infit = 1.40, Outfit = 1.89) had Infit and/or Outfit statistics above the acceptable range. An item with high Infit and Outfit statistics suggests the item may be erratic or too noisy to meet the strict Rasch requirements (Bond & Fox, 2015). Item 26 in particular has conflicting constructs of "rich foods" and "new foods" which may be confusing for the respondent (Ocker, Lam, Jensen, & Zhang, 2007). The Infit and Outfit statistics for all other items were acceptable once these three items were removed one at a time (*M* Infit *MNSQ* = 1.01 ± 0.22; *M* Outfit *MNSQ* = 0.99 ± 0.25). The resulting Infit range was then 0.62 – 1.43 and the resulting Outfit range was 0.67 -1.48.

Table 1 shows the EAT-26 items having acceptable fit, resulting in a 23 item Eating

Attitudes Test, or the EAT-23.

Overall, 22.58% of screened persons showed Infit and/or Outfit statistics outside

of the acceptable range while mean fit statistics were agreeable to the Rasch model with

mean squares close to 1.0 (*M* Infit *MNSQ* = 1.04 ± 0.45; *M* Outfit *MNSQ* = 0.99 ± 0.46).

Screened persons having Infit statistics outside of the acceptable range were 18.38%

while Outfit were 15.48%. While all fit statistics are important, they are slightly more

important for items than persons. As Wright and Linacre stated "we are usually stricter in

our application of fit rules to items than to persons. A few maverick persons in a dataset

don't worry us - they will have negligible impact on anything else. But a few maverick

items raise questions about test administration, data entry accuracy, the definition of the

latent variable, etc. We will immediately focus our attention on them because they may

be symptomatic of a more pervasive problem, such as the wrong key for a multiple-

choice test, or reversed-coded items on a survey" (1994, "Item Fit and Person Fit," para.

1). Overall, 70 out of 310 persons in the sample with unpredictable responses are not

detrimental to the measure but do suggest more than "a few maverick persons."

Additional inquiry is necessary to understand the nature of the misfit. The vast majority

of these persons were misfit due to their responses being more varied (or noisy) than

what the Rasch model predicted or what is referred to as underfit (Bond & Fox, 2015). To

qualitatively understand the cause of the misfit among these persons, the Person

Keyforms Misfit Order table in Winsteps was utilized. This table provides a list of only

unexpected responses by persons and displays both the expected response and the actual

Table 1

EAT-26 Items After Misfit Removal

1	I am terrified about being overweight.
2	I avoid eating when I am hungry.
3	I find myself preoccupied with food.
4	I have gone on eating binges where I feel that I may not be able to stop.
5	I cut my food into small pieces.
6	I am aware of the calorie content of foods that I eat.
7	I particularly avoid food with a high carbohydrate content (i.e. bread, rice, potatoes, etc.).
9	I vomit after I have eaten.
10	I feel extremely guilty after eating.
11	I am occupied with a desire to be thinner.
12	I think about burning up calories when I exercise.
14	I am preoccupied with the thought of having fat on my body.
15	I take longer than others to eat my meals.
16	I avoid foods with sugar in them.
17	I eat diet foods.
18	I feel that food controls my life.
19	I display self-control around food.
20	I feel that others pressure me to eat.
21	I give too much time and thought to food.
22	I feel uncomfortable after eating sweets.
23	I engage in dieting behavior.
24	I like my stomach to be empty.
25	I have the impulse to vomit after meals.

observed response. For a majority of the responses, the misfit was in responding *more agreeably* to an item than what was expected. For example, Rasch analysis for a person may have an "expected" response of *Seldom* to the item "I display self-control around food" but observe an actual answer of *Always*. In another example, several persons were expected to respond *Never* or *Seldom* to the items "I vomit after I have eaten" and "I have the impulse to vomit after meals" but actually responded *Often* or *Seldom* suggesting that even in the absence of other disordered eating symptomology, these practices can still exist to some degree. Another item which appeared as a misfit for several persons was "I am terrified of being overweight." Rasch modeling predicted an answer of *Seldom* for many but the observed answer was *Always*. It is unknown whether these responses were due to guessing, lack of knowledge, misunderstanding, or another unknown reason. The observed patterns of answering more favorably to items than predicted lends evidence to the idea that some disordered eating or dieting habits are perhaps considered more normal in society than they were previously (Tury, Guelec, & Kohls, 2010) and the EAT-26 can result in many high EAT-26 scorers who do not express all of the symptoms of anorexia nervosa or bulimia nervosa but may be subclinical in nature (Garfinkel & Newman, 2001). Eating disorder symptoms also appear along a continuum from no eating disorder to full eating disorder with partial and subclinical or sub-threshold variations in between (Jacobi et al., 2011). An additional tool which can be helpful in determining the relationship of person risk to item "difficulty" (level of disordered eating) is the item-person map which provides a visual of how persons respond to certain items.

The item-person map distribution is shown in Figure 1. This map allows for EAT-23 item difficulty and person ability, or the extent to which a screened person has a risk of having an eating disorder, to be displayed on a single graph using logits as the common scale. We would usually expect to see the items uniformly distributed along the scale "like marks on a ruler" and a normal or skewed distribution for the persons (Linacre, 2012).

The logit scale is displayed on the left side of the item-person map followed by distribution of persons represented by pound (#) and dot (.) symbols. Items are centered around the mean (M) and show a distribution of one standard deviation (S) and two standard deviations (T). Persons located directly across from items have a 50% probability of endorsing that item correctly based on their ability. Persons near the top of the map were most able to disagree with the EAT-23 items while those at the bottom were most able to agree with the items. The right side of the scale shows the items which are more difficult to endorse towards the top (*I vomit after I have eaten*) and the items which were easier to agree with towards the bottom (*I display self-control around food*). The map shows that the persons were well distributed among the items, though there were fewer people who endorsed items that were more difficult or expressed more extreme levels of disordered eating symptomology.

```
        PERSONS - MAP - ITEMS
            <more>|<rare>
   5            .  +
                   |
                   |
                   |
                   ||
   4               +
                   |
                   |
                   |
                   |
   3               +
                   |  Item 9
               .   |
                   |
                   |T Item 25
   2               +
               .   |
                   |
                   |
               .   |
   1            .  +S Item 20   Item 24
            .## T|   Item 4
               .  |   Item 2
              .#  |   Item 18
              .#  |   Item 10
   0         .##  +M Item 15   Item 16   Item 21   Item 22   Item 7
           #### S|   Item 17   Item 5
            .### |   Item 23
           .#### |   Item 3
        .######### |   Item 14
  -1      .###### +S Item 11   Item 6
          .#### |
      .######### M|
         .###### |   Item 1    Item 12
         .###### |   Item 19
  -2       .#### +
           #### |T
        .######## |
           .# S|
           .# |
  -3       .## +
           .## |
           .# |
            . T|
               |
  -4         .# +
               |
             .# |
               |
               |
  -5           +
             . |
               |
               |
               |
  -6         . +
          <less>|<frequ>
```

Figure 1. A map of the distribution of persons and EAT-23 items.
Note. Items are on the right side of the map and persons are on the left side. "#" equals three people and "." equals one person.

This sample is ordered as would be expected as there is a fairly low frequency of eating disorders in this population (Garfinkel & Newman, 2001; L. Schrader, personal communication, July 7, 2014). There is a gap of items for persons who were less likely to have a risk for an eating disorder (-2 through -6 logits). While this group may not be the primary focus or target of the EAT screen, it may be important to provide additional items to show greater separation between the levels of eating disorder risk and to definitively capture this low to no-risk group.

EAT-23 item difficulty parameters (in logits) were calibrated and reported with standard errors, and Infit and Outfit statistics in Table 2. The greater the logit score, the more difficult it was for a person being screened to agree with the item. The EAT-23 item difficulty ranged from 2.79 to -1.73 logits. The most difficult to endorse item was "I vomit after I have eaten," with a value of 2.79 (SE = 0.18). The easiest to endorse item was "I display self-control around food," with a value of -1.73 (SE = 0.08). Item separation was 10.69, signifying the EAT-23 shows an adequate spread of items across the construct being measured, the risk of having an eating disorder. The item separation-reliability was .99 which indicates a strong probability of the items being placed similarly if they were administered to another similar sample (Bond & Fox, 2015).

Support for construct validity is provided by the item hierarchy which does seem to move from easier to endorse items to items which are more difficult to endorse. However, near the middle of the scale near logit 0, it does become more difficult to assign hierarchy. Items 15, 16, 21, 22 and 7 all have similar difficulty parameters (.04 through - .06) on the logit scale. It may be possible to reduce some of these items to create a shorter scale if they are considered redundant. For instance, one of either item 16

Table 2

Summary of Evaluation of the 23-Item Version of the EAT

Item	Calibration Logit	SE Logit	Infit MNSQ	Outfit MNSQ
I vomit after I have eaten.	2.79	0.18	1.19	0.76
I have the impulse to vomit after meals.	2.29	0.15	1.06	0.67
I feel that others pressure me to eat.	1.09	0.11	1.02	0.90
I like my stomach to be empty.	0.95	0.10	0.88	0.88
I have gone on eating binges where I feel that I may not be able to stop.	0.71	0.10	1.15	1.29
I avoid eating when I am hungry.	0.58	0.09	0.62	0.78
I feel that food controls my life.	0.40	0.09	0.96	0.92
I feel extremely guilty after eating.	0.28	0.09	0.82	0.76
I give too much time and thought to food.	0.04	0.09	0.72	0.73
I avoid foods with sugar in them.	0.00	0.09	1.20	1.19
I feel uncomfortable after eating sweets.	-0.02	0.09	0.91	0.85
I particularly avoid food with a high carbohydrate content (i.e. bread, rice, potatoes, etc.).	-0.05	0.09	0.86	0.91
I take longer than others to eat my meals.	-0.06	0.09	1.34	1.42
I eat diet foods.	-0.23	0.08	0.74	0.81
I cut my food into small pieces.	-0.25	0.08	1.24	1.28
I engage in dieting behavior.	-0.41	0.08	0.71	0.70
I find myself preoccupied with food.	-0.51	0.08	0.92	1.07
I am preoccupied with the thought of having fat on my body.	-0.74	0.08	0.91	0.90
I am occupied with a desire to be thinner.	-0.92	0.08	0.87	0.86
I am aware of the calorie content of foods that I eat.	-1.04	0.08	1.18	1.23
I think about burning up calories when I exercise.	-1.58	0.08	1.07	1.07
I am terrified about being overweight.	-1.59	0.08	1.41	1.38
I display self-control around food.	-1.73	0.08	1.43	1.48

Note. Item separation = 10.63; Item separation reliability = .99; SE = standard error; MNSQ = mean square residual.

(*I avoid foods with sugar in them*) or item 22 (*I feel uncomfortable after eating sweets*) may be sufficient to capture this concept. There is also a gap of content between items 20 and 4 and items 25 and 9. While items 25 and 9 express the more extreme behavior of eating disorders (bulimia), there could be additional items to capture persons just below that threshold.

The level of risk of having an eating disorder for each person screened, or the person ability level, was estimated. Persons with higher (positive) logit scores were more likely to agree with EAT-23 items and were potentially at greater risk for having an eating disorder. The average level of disordered eating risk was -1.39 ($SD = 1.25$). The range for risk across all persons was 6.76 to -6.40 logits, indicating a wide spread of risk. Person separation was 2.74. This is an indication of the extent to which persons are separated along the measured trait. With a person separation-reliability of 0.88, there is an adequate likelihood that person placement would be similar if this sample was administered a similar assessment measuring disordered eating (Bond & Fox, 2015).

Differential item functioning (DIF) analyses were completed for gender groups. None of the items functioned differently across gender with no items having a statistically significant ($p < .001$) and M-H DIF size > 0.64 logits.

Discussion

The purpose of this study was to calibrate and observe the measurement qualities of the newly developed four category (4-CAT) rating scale option of the EAT-26 identified in a previous study. These study results suggest empirical support for the psychometric properties of the 4-CAT Eating Attitudes Test. The Rasch calibration indicates acceptable model–data fit with 23 of the original 26 items having Infit and

Outfit statistics within the acceptable range. Removal of the three misfit items results in a psychometrically sound eating disorder risk assessment with 23 items and a four category Likert scale (EAT-23).

The item-person map shows adequate spread and placement of item difficulties and eating disorder risk levels of screened persons. A clear separation of items and persons is critical to show the tool can distinguish between groups. These results provide support for using the EAT-23 as a shortened and easier to discriminate measure of eating disorder risk.

Calibration of the EAT items helps to understand the hierarchy of the items in the measure. Agreement to EAT items that perhaps once suggested disordered eating was present in an individual are considered much more normal in our current body image and fitness conscious society (Tury et al., 2010). Additionally, eating disorder symptoms occur on a continuum (Fitzgibbon, Sanchez-Johnsen, & Martinovich, 2003; Jacobi et al., 2011) that changes over time due to society and the pressure of social media, culture, film and fashion (Garfinkel & Newman, 2001; Stice, 2002). Since the EAT-26 was first created in 1982, many changes have occurred with diagnostic criteria, knowledge of symptomology, and cultural views of disordered eating, dieting, and body image. The easiest to endorse items within this sample suggest were "I display self-control around food," "I am terrified about being overweight," "I think about burning up calories when I exercise," and "I am aware of the calorie content of the foods I eat." In the only other known study to apply the Rasch model to the Eating Attitudes Test, Salazar Mora & Prado-Calderón (2015) found the easiest to endorse item was "I think about burning up calories when I exercise." These are statements that many in society may feel

comfortable agreeing with without necessarily feeling that agreement indicates

disordered eating, yet some studies suggest disordered eating and dieting behaviors can

progress to a clinical eating disorder over time" (Neumark-Sztainer et al., 2011; Shisslak,

Crago, & Estes, 1995).

While agreement with these items could imply an excess of restriction and control

pertaining to food and weight, they could also be an unintended response to many recent

public health marketing efforts aimed at healthy eating, portion control, and exercise.

These marketing messages may have inadvertently resulted in increased stigma around

overweight and obesity resulting in stronger "agree" responses to these items (Puhl &

Heuer, 2010). Still, while a significant group of people could endorse these items, another

significant portion of the study sample was more able to disagree with the items resulting

in a content gap where no items were available to capture this group effectively.

Additional items should be added to better describe the nature of these respondents and to

ensure that results are not due to "acquiescence response bias" (Revilla, Saris, &

Krosnick, 2014, p. 74). Potential items for this group could include those that identify

eating habits based on a desire for health or wellness rather than a desire for thinness.

Several similarities appear with the hierarchy of items and persons between this

study and the study of the EAT-40 using Rasch with a Spanish sample (Salazar Mora &

Prado-Calderón, 2015). As mentioned, this is the only other known study to date to use

Rasch methodology with an Eating Attitudes Test (EAT). In both of these studies, the

hardest to endorse items were related to bulimia and included "I have the impulse to

vomit after meals" and "I vomit after I have eaten." Both studies also showed a pooling

of items (which may indicate redundancy) near the mean though they were much more

pronounced in the Spanish study with 34 items (out of 40) and just 5 items (out of 23) with the current study. The present study provides much more discrimination among persons and items when compared to Salazar Mora & Prado-Calderón's results which utilized a six category (6-CAT) Likert scale structure. In both studies, most persons were less likely to endorse those items that suggested the greatest disordered eating symptomology. Additionally, items for persons who were most likely to disagree with EAT-26 statements were missing in the Spanish sample as well suggesting a gap of content to target this group. Future research may take potential redundant items into account and create new content items where gaps are apparent.

Calibration and investigation of the psychometric properties of a measure should be ongoing. The results of this study suggest that the EAT-23 has good psychometric qualities and is an improvement upon the original EAT-26 with six Likert categories. Due to the problems associated with summing scores from ordinal items based on classical test theory, researchers who use the EAT-23 instrument should reference these Rasch analysis results. Table 3 offers a conversion of the summed, raw EAT-23 scores into logit scores. These logit scores should be utilized by future EAT-23 researchers. Since there is a change in total categories and items from the original EAT-26, a new cut-off score should be determined for the EAT-23 with a study utilizing samples of known clinical cases.

Additional measurement research is required which may lend more evidence to the improvement of the instrument. This includes applying the tool to various samples such as different cultures, ages, and populations with known clinical eating disorder diagnoses. This study has shown the Rasch model provides a useful method to provide validity evidence for a tool using person ability and item difficulty estimates.

Table 3

Rasch Logits Conversion Values for EAT-23 Total Scores

EAT-23 Total Score	Measures in Logits	EAT-23 Total Score	Measures in Logits
23	-6.41	58	-0.02
24	-5.17	59	0.07
25	-4.43	60	0.16
26	-3.98	61	0.24
27	-3.65	62	0.33
28	-3.38	63	0.42
29	-3.16	64	0.51
30	-2.96	65	0.60
31	-2.78	66	0.69
32	-2.62	67	0.78
33	-2.47	68	0.88
34	-2.34	69	0.97
35	-2.20	70	1.07
36	-2.08	71	1.17
37	-1.96	72	1.27
38	-1.85	73	1.38
39	-1.74	74	1.48
40	-1.64	75	1.60
41	-1.53	76	1.71
42	-1.43	77	1.83
43	-1.34	78	1.95
44	-1.24	79	2.08
45	-1.15	80	2.22
46	-1.06	81	2.36
47	-0.97	82	2.51
48	-0.88	83	2.68
49	-0.79	84	2.85
50	-0.70	85	3.05
51	-0.62	86	3.27
52	-0.53	87	3.52
53	-0.44	88	3.82
54	-0.36	89	4.19
55	-0.27	90	4.69
56	-0.19	91	5.49
57	-0.10	92	6.77

Note. EAT-23 total scores are calculated by summing the ratings for the 23 individual items.

REFERENCES

Anshel, M. H., Kang, M., & Jubenville, C. (2013). Sources of acute sport stress scale for sports officials: Rasch calibration. *Psychology of Sport and Exercise, 14*(3), *362-370*. doi:10.1016/j.psychsport.2012.12.003

Bond, T., & Fox, C. (2015). *Applying the Rasch model, fundamental measurement in the human sciences* (Third ed.). New York, NY: Routledge.

Brinthaupt, T. M., & Kang, M. (2014). Many-faceted Rasch calibration: An example using the Self-Talk Scale. *Assessment, 21*(2), 241-249. doi:10.1177/1073191112446653

Cook, A. R., Weatherby, N., Kang, M., Colson, J., & Owusu, A. (2016). An exploratory study of the psychometric properties and category functions of the Eating Attitudes Test (EAT-26) utilizing Rasch analysis (Unpublished doctoral dissertation). Middle Tennessee State University, Murfreesboro, TN.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York, NY, US: Routledge.

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Jr., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry, 26*(1), 57-63.

Fisher, W. P. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions, 6*(3), 238.

Fitzgibbon, M. L., Sanchez-Johnsen, L. A. P., & Martinovich, Z. (2003). A test of the continuity perspective across bulimic and binge eating pathology. *International Journal of Eating Disorders, 34*(1), 83-97. doi:10.1002/eat.10160

Garfinkel, P. E., & Newman, A. (2001). The Eating Attitudes Test: Twenty-five years later. *Eating and Weight Disorders, 6*(1), 1-24.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871-878.

Gross, J., Rosen, J. C., Leitenberg, H., & Willmuth, M. E. (1986). Validity of the Eating Attitudes Test and the Eating Disorders Inventory in Bulimia Nervosa. *Journal of Consulting and Clinical Psychology, 54*(6), 875-876. doi:10.1037/0022-006X.54.6.875

Jacobi, C., Fittig, E., Bryson, S. W., Wilfley, D., Kraemer, H. C., & Taylor, C. B. (2011). Who is really at risk? Identifying risk factors for subthreshold and full syndrome eating disorders in a high-risk sample. *Psychological Medicine, 41*(9), 1939-1949. doi:10.1017/s0033291710002631

Kang, M., Zhu, W. M., Ragan, B. G., & Frogley, M. (2007). Exercise barrier severity and perseverance of active youth with physical disabilities. *Rehabilitation Psychology, 52*(2), 170-176. doi:10.1037/0090-5550.52.2.170

Kang, S. J., & Kang, M. (2006). Rasch calibration of perceived weights of different sports games. *Measurement in Physical Education and Exercise Science, 10*(1), 51-66. doi:10.1207/s15327841mpee1001_4

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 1.

Linacre, J. M. (2012). Winsteps Rasch Tutorial 2. Retrieved from www.winsteps.com/a/winsteps-tutorial-2.pdf

Linacre, J. M. (2016). DIF - DPF - bias - interactions concepts. *Winsteps Help.* Retrieved from http://www.winsteps.com/winman/difconcepts.htm

Mintz, L. B., & O'Halloran, M. S. (2000). The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment, 74*(3), 489-503.

Neumark-Sztainer, D., Wall, M., Larson, N. I., Eisenberg, M. E., & Loth, K. (2011). Dieting and disordered eating behaviors from adolescence to young adulthood: Findings from a 10-year longitudinal study. *Journal of the American Dietetic Association, 111*(7), 1004-1011. doi:10.1016/j.jada.2011.04.012

Ocker, L. B., Lam, E. T. C., Jensen, B. E., & Zhang, J. J. (2007). Psychometric properties of the Eating Attitudes Test. *Measurement in Physical Education and Exercise Science, 11*(1), 25-48.

Puhl, R. M., & Heuer, C. A. (2010). Obesity stigma: important considerations for public health. *American Journal of Public Health, 100*(6), 1019-1028. doi:10.2105/AJPH.2009.159491

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic models for some intelligence and achievement tests. (Expanded ed.)*. Chicago, IL: University of Chicago Press.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree-disagree scales. *Sociological Methods and Research, 43*(1), 73-97. doi:10.1177/0049124113509605

Salazar Mora, Z., & Prado-Calderón, J. E. (2015). Analysis of the discriminant validity of

the Eating Attitudes Test through the Rasch model. *Revista Reflexiones, 94*(1),

123-135.

Shisslak, C. M., Crago, M., & Estes, L. S. (1995). The spectrum of eating disturbances.

International Journal of Eating Disorders, 18(3), 209-219. doi:10.1002/1098-

108x(199511)18:3<209::aid-eat2260180303>3.0.co;2-e

Smith, A., Rush, R., Fallowfield, L., Velikova, G., & Sharpe, M. (2008). Rasch fit

statistics and sample size considerations for polytomous data. *BMC Medical

Research Methodology, 8*(1). doi:10.1186/1471-2288-8-33

Stice, E. (2002). Risk and maintenance factors for eating pathology: A meta-analytic

review. *Psychological Bulletin, 128*(5), 825-848. doi:10.1037//0033-

2909.128.5.825

Tury, F., Guelec, H., & Kohls, E. (2010). Assessment methods for eating disorders and

body image disorders. *Journal of Psychosomatic Research, 69*(6), 601-611.

doi:10.1016/j.jpsychores.2009.05.012

WINSTEPS. (2012). Rasch Software. Retrieved from

http://www.winsteps.com/winsteps.htm

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch

Measurement Transactions, 8*(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis, Rasch measurement*.

Chicago, IL: MESA Press.

Zhu, W. M., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement, 1*(4), 286-304.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1-28.

CHAPTER V: OVERALL CONCLUSION

Eating disorders remain difficult constructs to measure due to inability to fully define symptoms, secrecy and symptoms kept hidden by those having eating disorders, and co-occurring morbidities that occur alongside eating disorders (Tury et al., 2010). The Eating Attitudes Test 26 (EAT-26) is one tool that was developed to assess eating disordered behavior among those with anorexia nervosa (Garner & Garfinkel, 1979). Since then, it has been used worldwide to assess eating disorder risk in both clinical and non-clinical samples and across a wide range of cultures and populations (Belon et al., 2011; Boyadjieva & Steinhausen, 1996; Garfinkel & Newman, 2001). Despite classical test theory evidence of validity and reliability with the EAT-26 (Garner et al., 1982), several studies doubt the ability of the EAT-26 to effectively capture disordered eating symptomology (Mintz, O'Halloran, Mulholland, & Schneider, 1997; Ocker et al., 2007). One reason is that the definitions and diagnostic criteria for eating disorders have developed significantly since the original EAT-40 was created in 1979 (Mintz & O'Halloran, 2000). Perhaps the most significant changes have been seen in the recent updates to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5, (*Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.)*, 2013). Binge eating disorder, not even referred to in 1979, now has its own category in the DSM-5. Despite the difficulties of screening for eating disorders, it is important to continue to put forth our best efforts to refine measurements. Eating disorders can be deadly, and early treatment can be critical to successful outcomes (Berkman et al., 2007; Eisenberg, Nicklett, Roeder, & Kirz, 2011; Wilksch, 2014). Having the right tools to make decisions is paramount to our ability to create effective change.

The broad goals of this study were to assess the EAT-26 utilizing various applications of Rasch methodology. Specifically, the first study aimed to evaluate the function of the EAT-26 and to perform an exploratory analysis of post-hoc combined categories. Data from this initial study led to additional inquiries to perform a confirmatory analysis of the categorical function of the rating scale for both three and four category response options with the EAT-26. The final study presented a calibration of the 4-CAT option of the instrument.

The Rasch model of analysis provided a valuable method to analyze the EAT-26 for rating scale category effectiveness. Likert-type rating scales are often not the first item of a measurement tool to be assessed for functionality, if not ever (Andrich, 2013). The summative scoring procedure that is most common with the EAT-26 items made the Rasch Rating Scale Method (RSM) (Wright & Masters, 1982) and Linacre's guidelines for optimizing rating scale category effectiveness (Linacre, 2002a) appropriate for the analysis. The most significant finding from the statistics available in the Rasch model was the observance of disordered thresholds in the original EAT-26. Almost without analysis one can look at the category labels (*Always*, *Usually, Often, Sometimes, Rarely*, and *Never*) and wonder if respondents are able to differentiate between some of them. If one determines that they cannot, the next step is to figure out if that makes a difference statistically. The results of the exploratory and confirmatory studies suggested that the 4-CAT rating scale option performed superior to the 6-CAT structure with ordered categories, and the same person separation and reliabilities. Perhaps the most visually significant pieces of evidence are the category probability curves. These figures provide convincing graphs that make the concept of disordered thresholds easier to interpret.

Improvements in the methodology of eating disorder measurement may help to determine which items are most crucial in screening and case-finding. The final study using Rasch calibration found three items outside of the required fit statistics for Rasch measurement resulting in a 23 item, four category (4-CAT) Likert scale Eating Attitudes Test. Item hierarchy, person abilities, and item difficulties provided by Rasch analysis were discussed with suggestions for future studies. The results of these studies provide evidence for continued improvement of this often-used measurement instrument.

REFERENCES

Alexandre, J., Ribeiro, R., & Cardoso, G. (2008). Gender differences in the EAT-26 score of bipolar patients. *Journal of Psychosomatic Research, 64*(6), 636.

Andrich, D. A. (1996). Category ordering and their utility. *Rasch Measurement Transactions, 9*(4), 465-466.

Andrich, D. A. (1998). Thresholds, steps and rating scale conceptualization. *Rasch Measurement Transactions, 12*(3), 648-649.

Andrich, D. A. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics and Outcomes Research, 11*(5), 571-585. doi:10.1586/erp.11.59

Andrich, D. A. (2013). The legacies of R. A. Fisher and K. Pearson in the application of the Polytomous Rasch model for assessing the empirical ordering of categories. *Educational and Psychological Measurement, 73*(4), 553-580. doi:10.1177/0013164413477107

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with Anorexia Nervosa and other eating disorders: A meta-analysis of 36 studies. *Archives of General Psychiatry, 68*(7), 724-731.

Attia, E., Becker, A. E., Bryant-Waugh, R., Hoek, H. W., Kreipe, R. E., Marcus, M. D., . . . Wonderlich, S. (2013). Feeding and eating disorders in DSM-5. *American Journal of Psychiatry, 170*(11), 1237-1239.

Belon, K. E., Smith, J. E., Bryan, A. D., Lash, D. N., Winn, J. L., & Gianini, L. M. (2011). Measurement invariance of the Eating Attitudes Test-26 in Caucasian and

Hispanic women. *Eating Behaviors, 12*(4), 317-320.
doi:10.1016/j.eatbeh.2011.07.007

Berger, U., Wick, K., Holling, H., Schlack, R., Bormann, B., Brix, C., . . . Strauss, B.
(2011). Screening of disordered eating in 12 year old girls and boys:
Psychometric analysis of the german versions of SCOFF and EAT-26. *European
Psychiatry, 61*(7), 311-318. 1. doi: 10.1055/s-0031-1271786

Berkman, N. D., Lohr, K. N., & Bulik, C. M. (2007). Outcomes of eating disorders: A
systematic review of the literature. *International Journal of Eating Disorders,
40*(4), 293-309. doi:10.1002/eat.20369

Bond, T., & Fox, C. (2001). *Applying the Rasch model, fundamental measurement in the
human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bond, T., & Fox, C. (2015). *Applying the Rasch model, fundamental measurement in the
human sciences* (Third ed.). New York, NY: Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*.
New York, NY: Springer.

Boyadjieva, S., & Steinhausen, H. C. (1996). The Eating Attitudes Test and the eating
disorders inventory in four Bulgarian clinical and nonclinical samples.
International Journal of Eating Disorders, 19(1), 93-98.

Chen, G., Tang, Z. L., Guo, G. P., Liu, X. Q., & Xiao, S. Y. (2015). The Chinese version
of the Yale Food Addiction Scale: An examination of its validation in a sample of
female adolescents. *Eating Behaviors, 18*, 97-102.
doi:10.1016/j.eatbeh.2015.05.002

Cloyd, J. H. Jr., (2005). The eating survey: Psychometric characteristics and validity of a

    self-report eating disorders instrument for adolescents. *Dissertation Abstracts*

    International: Section B: The Sciences and Engineering. Brigham Young

    University, US.

Crow, S. J., Peterson, C. B., Swanson, S. A., Raymond, N. C., Specker, S., Eckert, E. D.,

    & Mitchell, J. E. (2009). Increased mortality in Bulimia Nervosa and other eating

    disorders. *American Journal of Psychiatry*, *166*(12), 1342-1346.

    doi:10.1176/appi.ajp.2009.09020247

Diagnostic and statistical manual of mental disorders: DSM-5™ (5th ed.). (2013).

    Arlington, VA, US: American Psychiatric Publishing, Inc.

EAT-26. (2011). Retrieved from http://www.eat-26.com/index.php

Eisenberg, D., Nicklett, E. J., Roeder, K., & Kirz, N. E. (2011). Eating disorder

    symptoms among college students: Prevalence, persistence, correlates, and

    treatment-seeking. *Journal of American College Health, 59*(8), 700-707.

    doi:10.1080/07448481.2010.546461

Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Jr., Winokur, G., & Munoz, R.

    (1972). Diagnostic criteria for use in psychiatric research. *Archives of General*

    Psychiatry, 26(1), 57-63.

Fletcher, B., Kupshik, G. A., Uprichard, S., Shah, S., & Nash, A. S. (2008). Eating

    disorders and concurrent psychopathology: A reconceptualisation of clinical need

    through Rasch analysis. *European Eating Disorders Review, 16*(3), 191-198.

    doi:10.1002/erv.833

Garfinkel, P. E., & Newman, A. (2001). The Eating Attitudes Test: Twenty-five years later. *Eating and Weight Disorders, 6*(1), 1-24.

Garner, D. M. (1993). Self-report measures for eating disorders: A citation-classic commentary on the Eating Attitudes Test, an index of the symptoms of Anorexia-Nervosa by Garner, D.M., and Garfinkel, P.E., and development and validation of a multidimensional eating disorder inventory for Anorexia-Nervosa and Bulimia by Garner, D.M., Olmsted, M.P., and Polivy. *Current Contents/Social and Behavioral Sciences, 8*, 8.

Garner, D. M. (2016). EAT-26 Self-Test : Screening. Retrieved from http://www.eat-26.com/screening.php

Garner, D. M., & Garfinkel, P. E. (1979). Eating Attitudes Test: Index of the symptoms of Anorexia-Nervosa. *Psychological Medicine, 9*(2), 273-279.

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871-878.

Gearhardt, A. N., Corbin, W. R., & Brownell, K. D. (2009). Food addiction an examination of the diagnostic criteria for dependence. *Journal of Addiction Medicine, 3*(1), 1-7.

Girao, A., Antunes, M., & Silva, C. (2009). EAT-26 in the screening of alimentary disorders in adolescent ballerinas. *Hormone Research, 72*, 412.

Harris, E. C., & Barraclough, B. (1998). Excess mortality of mental disorder. *The British Journal of Psychiatry: The Journal of Mental Science, 173*, 11-53.

Hoek, H. W., & van Hoeken, D. (2003). Review of the prevalence and incidence of

    eating disorders. *International Journal of Eating Disorders, 34*(4), 383-396.

    doi:10.1002/eat.10222

Kang, M., Zhu, W. M., Ragan, B. G., & Frogley, M. (2007). Exercise barrier severity and

    perseverance of active youth with physical disabilities. *Rehabilitation Psychology,*

    52(2), 170-176. doi:10.1037/0090-5550.52.2.170

Kean, J., Malec, J. F., Altman, I. M., & Swick, S. (2011). Rasch measurement analysis of

    the Mayo-Portland Adaptability Inventory (MPAI-4) in a community-based

    rehabilitation sample. *Journal of Neurotrauma, 28*(5), 745-753.

    doi:10.1089/neu.2010.1573

Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied*

    Measurement, 3(1), 85-106.

Linacre, J. M. (2002b). What do infit and outfit, mean-square and standardized mean?

    Rasch Measurement Transactions, 16(2), 1.

Linacre, J. M. (2011). Rasch measures and unidimensionality. *Rasch Measurement*

    Transactions, 24(4), 1310.

Linacre, J. M. (2016). Winsteps® Rasch measurement computer program User's Guide.

    Beaverton, Oregon: Winsteps.com.

Loth, K. A., Mond, J., Wall, M., & Neumark-Sztainer, D. (2011). Weight status and

    emotional well-being: Longitudinal findings from project EAT. *Journal of*

    Pediatric Psychology, 36(2), 216-225. doi:10.1093/jpepsy/jsq026

Maiano, C., Morin, A. J. S., Lanfranchi, M. C., & Therme, P. (2013). The Eating

    Attitudes Test-26 revisited using exploratory structural equation modeling.

Journal of Abnormal Child Psychology, 41(5), 775-788. doi:10.1007/s10802-013-9718-z

Merlo, L. J., Klingman, C., Malasanos, T. H., & Silverstein, J. H. (2009). Exploration of food addiction in pediatric patients: A preliminary investigation. *Journal of Addiction Medicine, 3*(1), 26-32.

Mintz, L. B., & O'Halloran, M. S. (2000). The Eating Attitudes Test: Validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment, 74*(3), 489-503.

Mintz, L. B., O'Halloran, M. S., Mulholland, A. M., & Schneider, P. A. (1997). Questionnaire for Eating Disorder Diagnoses: Reliability and validity of operationalizing DSM-IV criteria into a self-report format. *Journal of Counseling Psychology, 44*(1), 63-79.

Myers, N. D., Feltz, D. L., & Wolfe, E. W. (2008). A confirmatory study of rating scale category effectiveness for the Coaching Efficacy Scale. *Research Quarterly for Exercise and Sport, 79*(3), 300-311.

National Eating Disorders Association. (2016a). *Other specified feeding or eating disorder*. Retrieved from https://www.nationaleatingdisorders.org/other-specified-feeding-or-eating-disorder

National Eating Disorders Association. (2016b). *Types and symptoms of eating disorders*. Retrieved from https://www.nationaleatingdisorders.org/types-symptoms-eating-disorders

Neumark-Sztainer, D., & Hannan, P. J. (2000). Weight-related behaviors among adolescent girls and boys - Results from a national survey. *Archives of Pediatrics and Adolescent Medicine, 154*(6), 569-577.

Ocker, L. B., Lam, E. T. C., Jensen, B. E., & Zhang, J. J. (2007). Psychometric properties of the Eating Attitudes Test. *Measurement in Physical Education and Exercise Science, 11*(1), 25-48.

Palmer, R. L. (2014). *Helping people with eating disorders : a clinical guide to assessment and treatment*: Chichester, England: Wiley Blackwell.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980). *Probabilistic models for some intelligence and achievement tests. (Expanded ed.)*. Chicago, IL: University of Chicago Press.

Salazar Mora, Z., & Prado-Calderón, J. E. (2015). Analysis of the discriminant validity of the Eating Attitudes Test through the Rasch model. *Revista Reflexiones, 94*(1), 123-135.

Smith, A., Rush, R., Fallowfield, L., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*(1). doi:10.1186/1471-2288-8-33

Stein, K. F., Riley, B. B., Hoyland-Domenico, L., & Lee, C. K. (2015). Measurement of body dissatisfaction in college-enrolled Mexican American women: A Rasch-based examination of the validity and reliability of the EDI-III. *Eating Behaviors, 19*, 5-8. doi:10.1016/j.eatbeh.2015.06.001

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680. doi:10.1126/science.103.2684.677

Supina, D., Herman, B. K., Frye, C. B., & Shillington, A. C. (2016). Knowledge of binge eating disorder: A cross-sectional survey of physicians in the United States. *Postgraduate Medicine, 128*(3), 311-316. doi:10.1080/00325481.2016.1157441

Tury, F., Guelec, H., & Kohls, E. (2010). Assessment methods for eating disorders and body image disorders. *Journal of Psychosomatic Research, 69*(6), 601-611. doi:10.1016/j.jpsychores.2009.05.012

Understanding eating disorders. (2008, Spring). *NIH Medline Plus: the magazine* [Internet], Retrieved from https://www.nlm.nih.gov/medlineplus/magazine/issues/spring08/articles/spring08pg17-19.html

U.S. Department of Health and Human Services. (2016). *Healthy People 2020 Topics and Objectives*. Retrieved from http://www.healthypeople.gov/2020/topics-objectives

Vandereycken, W., & Van Humbeeck, I. (2008). Denial and concealment of eating disorders: A retrospective survey. *European Eating Disorders Review, 16*(2), 109-114. doi:10.1002/erv.857

Vetrone, G., Cuzzolaro, M., Antonozzi, I., & Garfinkel, P. E. (2006). Screening for eating disorders: False negatives and eating disorders not otherwise specified. *European Journal of Psychiatry, 20*(1), 13-20.

Waaddegaard, M., Thoning, H., & Petersson, B. (2003). Validation of a screening instrument for identifying risk behaviour related to eating disorders. *European Eating Disorders Review, 11*(6), 433-455. doi:10.1002/erv.537

Wade, T. D., Keski-Rahkonen, A., & Hudson, J. I. (2011). *Epidemiology of eating disorders* (Third Edition). doi: 10.1002/9780470976739.ch20

Wilksch, S. M. (2014). Where did universal eating disorder prevention go? *Eating Disorders, 22*(2), 184-192. doi:http://dx.doi.org/10.1080/10640266.2013.864889

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal - measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation, 70*(12), 857-860.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis, Rasch measurement*. Chicago, IL: MESA Press.

Zhu, W. M. (1996). Should total scores from a rating scale be used directly? *Research Quarterly for Exercise and Sport, 67*(3), 363-372.

Zhu, W. M., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement, 1*(4), 286-304.

APPENDICES

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL LETTERS

November 23, 2011

Amanda Cole, Minsoo Kang, Youngdoek Kim, Norman Weatherby

Department of Health and Human Performance

arc3a@mtmail.mtsu.edu, nweatherb@mtsu.edu

Protocol Title: "Evaluation of a Food Addiction Measurement Tool"

Protocol Number:**12-124**

Dear Investigator(s),

The MTSU Institutional Review Board, or a representative of the IRB, has reviewed the research proposal identified above. The MTSU IRB or its representative has determined that the study poses minimal risk to participants and qualifies for an expedited review under the 45 CFR 46.110 Category 7.

Approval is granted for one (1) year from the date of this letter for 250 participants.

According to MTSU Policy, a researcher is defined as anyone who works with data or has contact with participants. Anyone meeting this definition needs to be listed on the protocol and needs to provide a certificate of training to the Office of Compliance. **If you add researchers to an approved project, please forward an updated list of researchers and their certificates of training to the Office of Compliance (c/o Emily Born, Box 134) before they begin to work on the project.**

Any change to the protocol must be submitted to the IRB before implementing this change.

Please note that any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918.

You will need to submit an end-of-project form to the Office of Compliance upon completion of your research located on the IRB website. Complete research means that you have finished collecting and analyzing data. **Should you not finish your research within the one (1) year period, you must submit a Progress Report and request a continuation prior to the expiration date.** Please allow time for review and requested revisions. Your study expires **November 23, 2012**.

Also, all research materials must be retained by the PI or faculty advisor (if the PI is a student) for at least three (3) years after study completion. Should you have any questions or need additional information, please do not hesitate to contact me.

Sincerely,

Emily Born

Emily Born
Research Compliance Officer
Middle Tennessee State University
eborn@mtsu.edu

August 1, 2014

Investigators: Amanda Rae Cole, Norman Weatherby

Department: HHP
Protocol Title: "An Evaluation of the Eating Attitudes Test (EAT-26): Meta-Analysis, Rasch Model of Measurement, and Determination of Cut-Off Scores"

Protocol Number: 14-328
Dear Investigator(s):
I have reviewed your research proposal identified above and your requested changes. I approve of the following changes:

· Online administration of the survey and the revised consent for this method.

Please note that any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615)494-8918. Any change to the protocol must be submitted to the IRB before implementing this change.

You will need to submit an end-of-project report to the Office of Compliance upon completion of your research. Complete research means that you have finished collecting data and you are ready to submit your thesis and/or publish your findings. Should you not finish your research within the one (1) year period, you must submit a Progress Report and request a continuation prior to the expiration date. Please allow time for review and requested revisions. Your studies expire **5/12/2015.**

According to MTSU Policy, a researcher is defined as anyone who works with data or has contact with participants. Anyone meeting this definition needs to be listed on the protocol and needs to complete the online training. If you add researchers to an approved project, please forward an updated list of researchers to the Office of Compliance before they begin to work on the project.

Please note: **all research materials must be retained** by the PI or **faculty advisor (if the PI is a student)** for at least **three (3) years after study completion**. Should you have any questions or need additional information, please do not hesitate to contact me.

Sincerely,
Kellie Hilker
Compliance Officer/ Institutional Review Board Member Middle Tennessee State University

6/3/2015

Investigator: Amanda Cole-Cook and Norman Weatherby
Department: Health and Human Performance
Investigator Email: arc3a@mtmail.mtsu.edu; norman.weatherby@mtsu.edu
Protocol Title: An evaluation of eating attitudes test (EAT-26): Meta-analysis
Rasch Model or Measurement and Determination of cut-off scores

Protocol Number: 14-328

Dear Investigator(s),
I have reviewed your research proposal identified above and your request for
continuation.Approval for continuation is granted for one (1) year from the date of
this letter. Any changes to the originally approved protocol must be provided to
and approved by the research compliance office.

You will need to submit an end-of-project report to the Office of Compliance upon
completion of your research. Should the research not be complete by the
expiration date, 5/12/2015, please submit a Progress Report for continued
review prior to the expiration date.

According to MTSU Policies, a researcher is defined as anyone who works with
data or has contact with participants. Therefore, should **any individuals be
added to the protocol that would constitute them as being a researcher,
ensure that they have taken the correct training and inform the Office of
Compliance prior to their involvement on the project**.

Any change to the protocol must be submitted to the IRB before implementing
this change. Please note that any unanticipated harms to subjects or adverse
events must be reported to the Office of Compliance at (615) 494-8918.

Also, all research materials must be retained in a secure location by the PI or
faculty advisor (if the PI is a student) for at least three (3) years after study
completion. Should you have any questions or need additional information,
please do not hesitate to contact me.

Sincerely,
Research Compliance Office
494-8918

Compliance@mtsu.edu

Moses M. Prabu
Wed 6/3/2015 1:16 PM

To:
Amanda R. Cook;
Norman Weatherby;
Research Compliance Office;

Dear Amanda,

Thanks for notifying your name change clearly in your form. It was indeed very useful. I have updated your name as Cole-Cook as we do not have a provision to completely replace a last name in our database.

Your protocol is currently approved till 05.12.2016. Be advised that you have to obtain an approval from our office should you decide to continue this project beyond the new expiration date. I also wish to add that our office will no longer be able to remind investigators on expiration dates and resubmissions.

Thank you and best wishes with your research

Sincerely,

Moses M. Prabu, Ph.D.
Compliance Officer
Office of Research Services
010A Sam Ingram Building (ING)
Middle Tennessee State University (PO BOX 124)
2269 Middle Tennessee Blvd
Murfreesboro, TN 37132

APPENDIX B: INFORMED CONSENT CHAPTER II

Middle Tennessee State University Institutional Review Board
Informed Consent Document for Research

Principal Investigator: Amanda R. Cole
Study of Eating Behavior Measurement Tools
Institution: Middle Tennessee State University

The following information is provided to inform you about the research project and your participation in it. Please read this form carefully and feel free to ask any questions you may have about this study and the information given below. You will be given an opportunity to ask questions, and your questions will be answered. Also, you will be given a copy of this consent form.

Your participation in this research study is voluntary. You are also free to withdraw from this study at any time without consequences. In the event new information becomes available that may affect the risks or benefits associated with this research study or your willingness to participate in it, you will be notified so that you can make an informed decision whether or not to continue your participation in this study.

For additional information about giving consent or your rights as a participant in this study, please feel free to contact the MTSU Office of Compliance at (615) 494-8918.

1. **Purpose of the study:** You are being asked to participate in a research study to gather information regarding eating behaviors.

2. **Description of procedures to be followed and approximate duration of the study:** You will complete seven surveys related to your eating habits and personality. The surveys should not take you more than one hour to complete. In no way are we diagnosing participants with disorders, but simply validating methods.

3. **Expected costs:** There are no costs associated with this study.

4. **Description of the discomforts, inconveniences, and/or risks that can be reasonably expected as a result of participation in this study:** The questions may make you uncomfortable or bring up negative feelings. If you are concerned you may have an eating disorder, please contact your physician or feel free to contact health services or counseling services on campus (615-898-2670).

5. **Compensation in case of study-related injury:** MTSU will not provide compensation in the case of study related injury.

6. **Anticipated benefits from this study:** The potential benefits to science and humankind that may result from this study are increased knowledge regarding eating habits and behaviors for scientific research.

7. **Compensation for participation:** There will be no monetary compensation for participation in this study.

8. **Circumstances under which the Principal Investigator may withdraw you from study participation:**
   Failure to follow instructions or showing a blatant disregard for the integrity of the study (answering "A" for all the questions).

9. **Contact Information.** If you should have any questions about this research study or possible injury, please feel free to contact Amanda Cole at arc3a@mtmail.mtsu.edu or my Faculty Advisor, Dr. Norman Weatherby at weather@mtsu.edu.

10. **Confidentiality.** All efforts, within reason, will be made to keep the personal information in your research record private but total privacy cannot be promised. No identifying personal information will be asked on these surveys such as name, full date of birth, address or Social Security number. Each survey will only be identified with a number not attached in any way to your name. Information from this study may be shared with MTSU or the government, such as the Middle Tennessee State University Institutional Review Board, Federal Government Office for Human Research Protections, if you or someone else is in danger or if we are required to do so by law.

11. **STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY**
   I have read this informed consent document and the material contained in it has been explained to me verbally. I understand each part of the document, all my questions have been answered, and I freely and voluntarily choose to participate in this study.

APPENDIX C: INFORMED CONSENT CHAPTER III

**Middle Tennessee State University Institutional Review Board
Informed Consent Document for Research**

Principal Investigator: Amanda R. Cole
Study of Eating Behavior Measurement Tools
Institution: Middle Tennessee State University
The following information is provided to inform you about the research project and your participation in it. Please read this form carefully and feel free to ask any questions you may have about this study and the information given below. Contact Amanda Cole at arc3a@mtmail.mtsu.edu and your questions will be answered. Also, you may print a copy of this consent form for your records.

Your participation in this research study is voluntary. You are also free to withdraw from this study at any time without consequences. In the event new information becomes available that may affect the risks or benefits associated with this research study or your willingness to participate in it, you will be notified so that you can make an informed decision whether or not to continue your participation in this study. For additional information about giving consent or your rights as a participant in this study, please feel free to contact the MTSU Office of Compliance at (615) 494-8918.

1. Purpose of the study: You are being asked to participate in a research study to gather information regarding eating behaviors.

2. Description of procedures to be followed and approximate duration of the study: You will complete surveys online related to your eating habits and personality. The surveys should not take you more than one hour to complete. In no way are we diagnosing participants with disorders, but simply validating methods. If a student of Middle Tennessee State University, please contact your health services or counseling services on campus (615-898-2670) for information regarding diagnosis. Otherwise, contact your personal physician or counselor.

3. Expected costs: There are no costs associated with this study.

4. Description of the discomforts, inconveniences, and/or risks that can be reasonably expected as a result of participation in this study: The questions may make you uncomfortable or bring up negative feelings. If you are concerned you may have an eating disorder, please contact your physician or, if a Middle Tennessee State University student, feel free to contact health services or counseling services on campus (615-898-2670). For additional resources related to eating disorders, please visit the National Eating Disorders Association.

5. Compensation in case of study-related injury: MTSU will not provide compensation in the case of study related injury.

6. Anticipated benefits from this study: The potential benefits to science and humankind that may result from this study are increased knowledge regarding eating habits and behaviors for scientific research.

7. Compensation for participation: There will be no monetary compensation for participation in this study. Some professors may provide extra credit for students by having students submit the Confirmation Page at the completion of the online surveys.

8. Circumstances under which the Principal Investigator may withdraw you from study participation:
Failure to follow instructions or showing a blatant disregard for the integrity of the study (answering "A" for all the questions).

9. Contact Information. If you should have any questions about this research study or possible injury, please feel free to contact Amanda Cole at arc3a@mtmail.mtsu.edu or Faculty Advisor, Dr. Norman Weatherby at norman.weatherby@mtsu.edu.

    10. Confidentiality. All efforts, within reason, will be made to keep the personal information in your research record private but total privacy cannot be promised. No identifying personal information will be asked on these surveys such as name, full date of birth, address or Social Security number. Each survey will only be identified with a unique number and no IP addresses will be collected. These surveys utilize the SurveyMonkey platform in order to collect answer data. The SurveyMonkey security statement is available for viewing online at https://www.surveymonkey.com/mp/policy/security/. Information from this study may be shared with MTSU or the government, such as the Middle Tennessee State University Institutional Review Board, Federal Government Office for Human Research Protections, if you or someone else is in danger or if we are required to do so by law.

11. STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY
I have read this informed consent document. I understand each part of the document, all my questions have been answered, and I freely and voluntarily choose to participate in this study. By pressing the "I am 18 and I Agree to Participate" button below, you agree that you consent to participate in this study.

APPENDIX D: EAT-26

Eating Attitudes Test (EAT-26)©

Instructions: This is a screening measure to help you determine whether you might have an eating disorder that needs professional attention. This screening measure is not designed to make a diagnosis of an eating disorder or take the place of a professional consultation. Please fill out the below form as accurately, honestly and completely as possible. There are no right or wrong answers. All of your responses are confidential.

Part A: Complete the following questions:

1) Birth Date   Month: _____   Day: _____   Year: _____   2) Gender: Male   Female

3) Height   Feet : _____   Inches: _____   ☐   ☐

4) Current Weight (lbs.): _____   5) Highest Weight (excluding pregnancy): _____

6) Lowest Adult Weight: _____   7: Ideal Weight: _____

Part B: Check a response for each of the following statements:	Always	Usually	Often	Some times	Rarely	Never
1. Am terrified about being overweight.	☐	☐	☐	☐	☐	☐
2. Avoid eating when I am hungry.	☐	☐	☐	☐	☐	☐
3. Find myself preoccupied with food.	☐	☐	☐	☐	☐	☐
4. Have gone on eating binges where I feel that I may not be able to stop.	☐	☐	☐	☐	☐	☐
5. Cut my food into small pieces.	☐	☐	☐	☐	☐	☐
6. Aware of the calorie content of foods that I eat.	☐	☐	☐	☐	☐	☐
7. Particularly avoid food with a high carbohydrate content (i.e. bread, rice, potatoes, etc.)	☐	☐	☐	☐	☐	☐
8. Feel that others would prefer if I ate more.	☐	☐	☐	☐	☐	☐
9. Vomit after I have eaten.	☐	☐	☐	☐	☐	☐
10. Feel extremely guilty after eating.	☐	☐	☐	☐	☐	☐
11. Am preoccupied with a desire to be thinner.	☐	☐	☐	☐	☐	☐
12. Think about burning up calories when I exercise.	☐	☐	☐	☐	☐	☐
13. Other people think that I am too thin.	☐	☐	☐	☐	☐	☐
14. Am preoccupied with the thought of having fat on my body.	☐	☐	☐	☐	☐	☐
15. Take longer than others to eat my meals.	☐	☐	☐	☐	☐	☐
16. Avoid foods with sugar in them.	☐	☐	☐	☐	☐	☐
17. Eat diet foods.	☐	☐	☐	☐	☐	☐
18. Feel that food controls my life.	☐	☐	☐	☐	☐	☐
19. Display self-control around food.	☐	☐	☐	☐	☐	☐
20. Feel that others pressure me to eat.	☐	☐	☐	☐	☐	☐
21. Give too much time and thought to food.	☐	☐	☐	☐	☐	☐
22. Feel uncomfortable after eating sweets.	☐	☐	☐	☐	☐	☐
23. Engage in dieting behavior.	☐	☐	☐	☐	☐	☐
24. Like my stomach to be empty.	☐	☐	☐	☐	☐	☐
25. Have the impulse to vomit after meals.	☐	☐	☐	☐	☐	☐
26. Enjoy trying new rich foods.	☐	☐	☐	☐	☐	☐

Part C: Behavioral Questions: In the past 6 months have you:	Never	Once a month or less	2-3 times a month	Once a week	2-6 times a week	Once a day or more
A Gone on eating binges where you feel that you may not be able to stop? *	☐	☐	☐	☐	☐	☐
B Ever made yourself sick (vomited) to control your weight or shape?	☐	☐	☐	☐	☐	☐
C Ever used laxatives, diet pills or diuretics (water pills) to control your weight or shape?	☐	☐	☐	☐	☐	☐
D Exercised more than 60 minutes a day to lose or to control your weight?	☐	☐	☐	☐	☐	☐
E Lost 20 pounds or more in the past 6 months	Yes ☐		No ☐			

* Defined as eating much more than most people would under the same circumstances and feeling that eating is out of control

© Copyright: EAT-26: (Garner et al. 1982, *Psychological Medicine, 12*, 871-878); adapted by D. Garner with permission.

APPENDIX E: EAT-26 PERMISSION

eat26_reproduce_permission@eat-26.com

Sun 11/20/2011 10:21 PM

To:Amanda R. Cole <arc3a@mtmail.mtsu.edu>;

Thank you for your permission request to reproduce and use the EAT-26. The EAT-26 is protected under copyright; however, all fees and royalties have been waived because it has been our wish for others to have free access to the test.

Please consider this e-mail as granting you permission to reproduce the test for the purpose suggested in your request as long as the EAT-26 is cited properly. The correct citation is: "The EAT-26 has been reproduced with permission. Garner et al. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. Psychological Medicine, 12, 871-878."

You can download a copy of the scoring instructions and the test on the homepage of the EAT-26 website. If you use the written version of the test, it is recommended that you provide respondents with the link to the EAT-26 website (www.eat-26.com) so that they can learn more about the test.

Again, thank you for requesting permission to reproduce and use the EAT-26. If you intend on publishing your work, please send me your results so that they can be included in a research database being developed on the EAT-26 website (www.eat-26.com).

Best wishes,

David M. Garner, Ph.D.
Administrative Director
River Centre Clinic
5465 Main Street
Sylvania, OH 43560
dm.garner@gmail.com