THESIS

THE DEVELOPMENT OF A SCALE TO MEASURE ORTHOREXIA NERVOSA

Submitted by

Melanie E. Kramer

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2016

Master's Committee:

Advisor: Kathryn Rickard

Bradley T. Conner
Melissa Wdowik

ABSTRACT

THE DEVELOPMENT OF A SCALE TO MEASURE ORTHOREXIA NERVOSA

Due in part to the obesity epidemic occurring today in the United States, the public discussion of healthy eating has become an emerging social trend (Chaki et al., 2013). Evidence suggests that since the importance of healthy eating has been increasingly emphasized, individuals have become more aware of how their diet impacts their overall health, and many have attempted to adjust their food choices (Bagci Bosi et al., 2007; Nicolosi, 2006). It has been suggested that, for some, healthy eating has been taken to the extreme, resulting in social, physical, and psychological consequences (Chaki et al., 2013). The term Orthorexia Nervosa (ON) was originally developed by Steven Bratman; he conceptualized ON as a way to describe an obsession about proper nutrition and a "fixation" on healthy eating (Bratman & Knight, 2000). Currently, very little is known about ON as a construct, and neither a formal operational definition nor standard diagnostic criteria of ON exist. Although, there are two scales that measure ON (Donini, 2007; Gleaves et al., 2013), neither scale seems to fully capture this phenomenon. This study will attempt to develop a valid and reliable scale, the Orthorexia Nervosa Scale (ONS), to more accurately and fully capture the construct of ON.

The first step was the development of the ONS items, which included a thorough review of the literature and consultation with experts in the field. 103 items were retained to form the initial version of the ONS. Using a development sample (n = 712), I conducted a parallel analysis on the items. An exploratory factor analysis (EFA) was also conducted and items were selected according to their psychometric quality. The EFA revealed a 10-factor structure with 47 items.

The second step was conducting a confirmatory factor analysis on the data from a second development sample (n = 397). Analyses indicated a good model fit of the 47-item, 10-factor scale after making model modifications. This measure will be a useful tool in future research and clinical work related to ON. A new, more comprehensive, measure of ON which captures all of its components could lead eventually to improved treatment for these individuals.

TABLE OF CONTENTS

INTRODUCTION

Origins of ON

The term Orthorexia Nervosa (ON) was originally developed by Steven Bratman; he conceptualized ON as a way to describe an obsession about proper nutrition and a "fixation" on healthy eating (Bratman & Knight, 2000). Within Bratman and Knight's (2000) model, a person may be diagnosed with ON through the presence of the following: 1) spending more than 3 hours per day thinking about and preparing healthy food, 2) feeling superior to those with differing eating habits, 3) rigidly following a particular self-imposed dietary regimen and engaging in compensatory restriction to make up for any dietary indiscretions, 4) attaching self-esteem to adherence to the self-imposed diet, 5) making consumption of a healthy diet the central focus of life, at the expense of other personal values, relationships, previously enjoyed activities, and sometimes physical health, and 6) making nutritional value of a meal more important than the pleasure of eating it. While the concept has been criticized because it originated through Bratman's personal experience, there seems to be evidence that ON exists; for instance, using an already existing measure for orthorexia (e.g. ORTO-15), researchers found a high prevalence in many populations, including athletes, performance artists, individuals in the food sciences field, and resident medical doctors (Segura-Garcia, Papaianni, Caglioti, Procopio, Nistico, Bombardiere, Ammendolia, Rizza, De Fazio & Capranica, 2012; Aksoydan & Camci, 2009; Donini, Marsili, Graziani, Imbriale & Cannella, 2004; Bagci Bosi, Camur & Guler, 2007), and it has been suggested that ON may be embedded within the larger trend towards healthy eating in society (Pilzner, 2002; Kickbusch & Payne, 2003).

Evidence suggests that since the importance of healthy eating has been increasingly emphasized, individuals have become more aware of how their diet impacts their overall health,

and have attempted to adjust their food choices accordingly (Bagci Bosi et al., 2007; Nicolosi, 2006). A LexusNexus search on "healthy eating" indicated that 882 newspaper articles have been written on the subject since 2010, which highlights the heightened focus on healthy eating in recent years. As more food is being produced in ways that are processed, engineered, and modified, individuals are left to rely more heavily on labels and packaging to evaluate their food.

Nicolosi (2006) stated that the combination of the two – the increase in conversations about the importance of healthy eating and the disconnect between food production and consumption – have led to societies with orthorexic tendencies and individuals with ON. He deemed the United States an orthorexic society, and stated that "eating unknown artefacts, without a past and without social roots, as has happened with the advent of modernity, may mean losing the deepest sense of self; and that to eat 'unnatural' foodstuffs means to symbolically de-naturalise oneself" (Nicolosi, 2006, p. 54). In other words, though the "orthorexic society" leads consumers to desire healthy choices, they live amidst a lack of knowledge as a result of their lack of connection with the food preparation industry (Nicolosi, 2006; Rangel et al., 2007).

Currently, no formal operational definition of ON has been empirically derived, and there are no standard diagnostic criteria for ON (Chaki, Pal & Bandyopadhyay, 2013). Chaki, Pal & Bandyopadhyay (2013) describe ON as an unrecognized eating disorder "in which the person becomes obsessed with eating pure, healthy and right kinds of foods to improve health" (p. 1046). Existing research, however, does not present a comprehensive and cohesive picture of what comprises ON. Though some researchers have raised questions about the validity of ON as a unique disorder, others believe that ON is indeed a unique disorder, different from recognized eating disorders, and that it should be included in the DSM-5 due to its distinct features (Chaki et al., 2013; Catalina Zamora, Bonaecha, Sanchez & Rial, 2005; Donini et al., 2005; Brytek-Matera,

2012). While Bratman & Knight (2000) posited that ON is a type of eating disorder where the focus resides in food quality rather than quantity, previous research suggests that this distinction may not hold up; some individuals with eating disorders do demonstrate concern about the specific types of food they will and will not eat (Affenito et al., 2002; Misra et al, 2006). Additionally, there is no empirical research that suggests that ON symptoms cause significant clinical impairment in the individual, though Bratman & Knight (2000) contend that ON can be considered a psychological disorder due to the physical, psychological, and social impact it can have on an individual.

Moroze et al. (2014) have recently proposed diagnostic criteria that fit well with the way in which ON is currently understood (See Table 1). While the diagnostic criteria seem to capture many of the components inherent in ON, they don't appear to address some of the psychological factors that likely comprise ON, such as superiority, downward social comparison, rigidity, disordered eating identity and meaning, loss of control, eating to cope, and outcome expectations. These psychological factors are important to address, not only to aid our conceptualizing and understanding of ON, but to guide our work in potentially developing measures to assess ON.

Proposed Consequences of Orthorexia Nervosa (ON)

For the purposes of this discussion, I will proceed with the assumption that ON exists, and follow with what has been presented in the literature as possible consequences to the individuals who exhibit it. An individual with ON is quite discriminatory about his or her choice of food in the context of "its purity, origin, presence of artificial ingredients or additives (if any), preservatives, etc." (Catalina Zamora et al., 2005). Dependence on such a strict diet may eliminate many essential nutrients and it may lead to several nutritional and mineral deficiencies, which could be harmful to individuals' health and ultimately compromise the quality of life

(Bratman & Knight, 2000; Bagci Bosi et al., 2007). For instance, an individual with ON might choose to starve instead of eating foods that they consider to be impure, unhealthy and harmful to health (Bratman & Knight, 2000; Nymah, 2002; Bagci Bosi et al., 2007). Although long-term empirical studies are missing from the literature, there is some indication that extreme dietary choices can lead to the same medical complications associated with severe anorexia nervosa (AN), such as osteopenia, anemia, hyponatremia, metabolic acidosis, bradycardia and testosterone deficiency (Moroze et al., 2014; Park et al., 2011). There are no existing empirical studies that examine potential long-term physical consequences of ON, however.

In terms of psychological consequences, there are likely many associated with ON. A person with ON develops his or her own highly specified, detailed food rules, which can become very restricting over time (Bartrina, 2007). These rules appear to be obsessive, and individuals who seem to fit the criteria for ON often wind up with disordered thinking. They may punish themselves with increasingly stringent dietary restrictions if they violate a personal food rule by consuming bad or wrong foods (Bratman & Knight, 2000). Outside of meals, extra time is spent researching and cataloging food, weighing and measuring food, and planning future meals, with additional, intrusive, food-related thoughts occurring as well (Bratman & Knight, 2000). Individuals with ON describe their symptoms as an overwhelming obsessive desire to feel pure, natural, and healthy that begins to override other pleasurable aspects of life (Bratman & Knight, 2000; Mathieu, 2005). The behavior becomes restrictive to the degree that it begins to interfere with the person's quality of life (Getz, 2009). An individual with ON experiences frustration when their food-related practices are disrupted, as well as disgust when food purity appears to be compromised. They report feeling guilt and self-loathing when they commit food transgressions (Mathieu, 2005). In fact, ON should only be considered a disorder if the presentation is long-

term, and when an individual's behavior has a significant negative impact on his or her quality of life (Nymah, 2002). While there has been speculation about the above psychological consequences, and while they have been observed by professionals, there is a lack of research examining this area.

An individual with ON's obsession with preparation of their food, as well as their extreme selectiveness about their choice of food, leads to a loss of social relationships and affective dissatisfactions (Chaki et al., 2013). Thus, individuals with ON are at risk for social isolation, due to concerns about being able to maintain healthy eating while in control of one's surroundings. It has been suggested that though an individual with ON may begin by wanting to improve their health or treat a disease, the diet eventually becomes an integral part of their lives (Catalina Zamora et. al., 2005). Their narrow perception of healthy eating may then affect their views about others and cause them to feel morally superior to anyone who does not have the same self-discipline. No empirical studies exist that examine any potential social consequences of ON, however.

Prevalence of ON

In addition to the paucity of research examining ON and its purported physical, psychological, and social consequences, little is known about the prevalence of ON worldwide. More research is needed to advance the understanding of ON, as well as how it is related to other eating disorders.

Kinzl, Hauer, Traweger & Kiefer (2006) examined a sample of 500 female dieticians using the 10-question Bratman test for ON; they found that 12.8 % of the sample met criteria for ON (Kinzl et al., 2006). The 10-question Bratman test has not been shown to be statistically reliable, however. Additionally, the fact that 12.8% of the sample met ON criteria is

disconcerting in that it seems unlikely that such a high percentage would meet criteria for a psychological disorder, even within a vulnerable population.

Bagci Bosi et al. (2007) examined a sample of 318 medical doctors in Turkey (53.1% male), and found that 45.5% of their participants met criteria for ON based on their scores on the ORTO-15. Again, the percentage of their participants that met criteria for ON is high, which indicates that the ORTO-15 may be an invalid measure in that it is not truly capturing the full construct of ON. Additionally, their sample only examined individuals with a high educational level. The participants were all medical doctors, presumably having an increased knowledge of healthy nutrition, which likely skewed the results.

A similar study conducted with medical students (N = 895, male = 464 and female =359) in Turkey selected only 11 items of the ORTO-15 that had factor loadings of .50 or higher for the Turkish version of the scale, and administered a version called the ORTO-11 (Fidan et al., 2010). They found that male students had a higher tendency for ON (a finding also present in Donini et. al., 2007) and that the percentage of the 21 and younger students with an orthorexia tendency was higher than for students older than 21 (Fidan et al., 2010). They used the mean score of the ORTO-11 to establish a cutoff point (score of 27), and found that 43.6% of the students met their criteria for ON. Again, this percentage of the population seems high, and as Fidan et al. (2010) mentioned, it would have been advantageous to have collected information from the students about their health knowledge, as it seems likely that this influences a person's likelihood of having ON. Like the ORTO-15, the ORTO-11 also seems to be an inadequate measure in that it is overestimating the percentage of individuals who meet criteria for ON.

Diagnostic boundaries of ON

Because of the lack of adequate measures or a formal operational definition of ON, it is unknown whether ON is a unique disorder or a subtype of an existing disorder. Initially, Bratman & Knight (2000) thought of ON as its own unique eating disorder. As research has progressed, however, questions have arisen about whether ON is truly a unique disorder or a variant of a current disorder, such as an already existing eating disorder or obsessive-compulsive disorder (Mathieu, 2005; Kummer, Diaz & Teixeira, 2008). Some researchers have argued that ON is "nothing but a psychopathological characteristic" that can be seen on a spectrum from the normal to the pathological (Kummer, Diaz & Teixeira, 2008).

The National Eating Disorder Association (NEDA) recognizes ON: they acknowledge that it is not currently recognized as a clinical diagnosis, but that many people struggle with symptoms associated with this term. Like other eating disorders, NEDA states that ON appears to be motivated by health, but there are underlying motivations, such as compulsion for complete control and escaping fears, that seem to overlap with other eating disorders (NEDA, 2015). To determine whether ON is a variant of an eating disorder or a unique disorder, it is important to address the similarities and differences between ON and established eating disorders.

Bratman & Knight (2000) argued that just as in AN, individuals coping with ON become so focused on controlling their eating habits that their life can become unbalanced and they lose perspective about their eating behaviors. Additionally, ON and AN share the characteristics of high trait anxiety, a need to control the environment and a genetic predisposition for perfection, (Fidan et al., 2010; Mathieu, 2005). Donini et al. (2004) found that a potential overlap between ON and AN exists in relation to a preference to starve over consuming food considered impure, which leads to another similarity between AN and ON, the potential for significant weight loss.

Varga, Dukay-Szabo, Tury & van Furth (2014) found that AN and ON are highly correlated with regard to eating attitudes and behaviors. Additionally, Fidan et al. (2010) found that individuals with a potential eating disorder, determined by their EAT-40 scores, seemed to have an orthorexic tendency.

Gleaves, Graham, & Ambwani (2013) noted that individuals with AN and ON are "overly preoccupied with food, may practice food related rituals, feel a sense of superiority over others based on their eating practices, have rigid or restrictive eating habits, increase restriction following consumption of forbidden foods, link their self-esteem to food-related behaviors, and make their eating-related issues the primary focus of their lives" (p. 2).The authors further noted that both AN and ON are ego-syntonic, meaning that a person's thoughts, attitudes, impulses and behavior associated with either AN or ON are felt to be acceptable and consistent with the rest of the personality (Gleaves, Graham, & Ambwani, 2013; Colman, 2015); this makes it more unlikely that individuals would seek help for their eating problems (Bratman & Knight, 2000). These similarities between AN and ON may indicate that ON is a subset of AN. Moreover, Kinzl et al. (2006) found that nutritionists who met criteria for ON had experienced an eating disorder in the past more frequently than their non-ON colleagues; while this difference was not statistically significant, this presents evidence of some overlap between ON and other eating disorders. Fidan and colleagues (2010) also found that ON was more prevalent in medical students with potential eating disorders based on their scores on the EAT-40. Research also suggests some overlap in personality characteristics; both individuals with AN and ON tend to be, "careful, detailed, and tidy…[and have an] exaggerated need for self care and protection" (Fidan et al., 2010, p. 53).

On the other hand, however, there also seem to be differences between ON and AN which suggest that ON may not be a subset of AN. Chaki et al. (2013) discuss that while individuals with AN and bulimia nervosa (BN) show an obsession about the quantity of the food they eat, individuals with ON show an obsession about the quality of the food they eat. However, some researchers have argued that individuals with AN care about their food quality (Kummer, Diaz, & Teixeira, 2008). Gleaves, Graham, & Ambwani (2013) discussed that while individuals with AN are motivated by a fear of being fat, individuals with ON are motivated to control their eating habits to be healthier. In other words, for individuals with ON, a fear of fatness and/or drive for thinness is not the main motivation. Moreover, research has shown that among Polish female students between 18-25 years old, a strong preoccupation with healthy food was not associated with an unhealthy relationship with their bodies (Brytek-Matera, Donini, Krupa, Poggiogalle, & Hay, 2015). For individuals with ON, the main motivation appears to be a desire to eat a pure and perfectly healthy diet (Chaki et al., 2013; Brytek-Matera et al., 2015). Thus, the biggest differences seem to be between being driven by a goal of a perfect diet and being motivated by weight loss (Fidan et al., 2010).

Another notable point of departure between AN and ON is that anorexic individuals tend to hide their behaviors, whereas individuals with ON are more likely to display their habits (Bratman & Knight, 2000). Additionally, both groups of individuals experience cognitive and perceptual distortions, having magical beliefs about food (Gleaves, Graham, & Ambwani, 2013). Magical beliefs about food can be defined as the tendency "to adopt eating habits and health instructions that many magazines, health care books and food ideologies regard as valid but which obey universal laws of similarity and contagion" (Lindeman, Keskivaara & Roschier, 2000). There is a difference, however, such that individuals with AN experience body image

distortions (Schneider et al., 2009; Streigel- Moore & Bulik, 2007), while individuals with ON may have distorted ideas about food properties, but not necessarily body image (Bratman & Knight, 2000). Yet, as highlighted by Gleaves, Graham, & Ambwani (2013), a study conducted by Eriksonn, Baigi, Marklund, & Lindgren (2008) suggested that internalization of the thin-ideal accounted for most of the variance in ON symptoms.

Research has also shown that ON exhibits overlap with obsessive-compulsive disorder (OCD). Bratman & Knight (2000) describe ON as an "obsessive" adherence to strict dietary requirements, such as feeling compelled to bring personally prepared food to meals, carefully weighing and measuring all foods consumed, detailing and engaging in extreme planning of meals, experiencing accompanying guilt whenever deviating from personal dietary restrictions, and a general preoccupation with food. Similar to individuals with OCD, individuals with ON have a restricted amount of time for other activities, as adherence to a strict eating style hinders normal routines (Donini et al., 2004). Additionally, Mathieu (2005) points out that the anxiety and perfectionism present in ON are also common components of OCD.

Just as ON might by a subset of AN, it may also be a subset of OCD as those with ON appear to spend a lot of time obsessing about food and engaging in compulsions related to food (Gleaves, Graham, & Ambwani, 2013). Where the research is lacking is clarity in terms of whether there is any distress associated with the obsessions, which is diagnostically required to meet criteria for OCD (DSM-V, 2013). The literature has noted that individuals with ON feel spiritually satisfied with eating the correct way, which indicates that ON may be something other than OCD (Gleaves, Graham, & Ambwani, 2013; Bratman & Knight, 2000).

Some researchers have found that ON is more prevalent in males (Fidan et al., 2010; Donini et. al., 2007; Aksoydan & Camci, 2009) On the other hand, Bagci Bosi et al. (2007)

found no gender differences when examining the presence of ON in a sample of medical doctors in Turkey and Brytek-Matera et al. (2015) found no gender difference among Polish students aged 18 to 25 years. In contrast, there is a large amount of research that shows that existing eating disorders (e.g. AN and BN) are more prevalent in females (Lewinsohn, Seeley, Moerk, & Striegel-Moore, 2002; Bulik, Sullivan, Tozzi, Furberg, Lichtenstein, & Pederson, 2006). These contrasting findings point to the fact that more research needs to be done to fully understand the concept of ON in order to accurately examine gender prevalence.

Through consulting with dietitians, psychologists, psychiatrists and graduate students in the eating disorder field, as well as extensively reviewing the literature, the following has been gathered as a basis for fully understanding the components of ON before developing a measure. ON is comprised of the following: 1) Fixation on healthy, biologically pure foods, 2) Strict avoidance of food that has been treated with pesticides, herbicides, and artificial substances, 3) Excessive worry about the food preparation techniques and ingredients used in food preparation, 4) Disproportionate allotment of time to planning, purchasing, preparing and eating meals, 5) Cautious eating style leading to omission of food groups, 6) Cautious eating style leading to nutritional deficiencies, becoming medically compromised, 7) Intense fixation on diet resulting in problems/difficulties in social relationships and low overall quality of life (less time spent on activities and interest), 8) Self-esteem becoming wrapped up in the purity of one's diet, 9) Feeling superior to anyone who indulges in impure dietary habits, 10) Sense of self-righteousness when eating "healthy" or judging others for not eating healthily, 11) Once strict obsession with food becomes a habit, difficulty returning to normal eating, 12) Loss of perspective and balance in life – too much of life's meaning is placed onto food, 13) Identity is comprised mainly of how "healthy" one is eating, 14) Experiencing feelings of guilt over

deviation from one's dietary plan despite a lack of medical or health consequences, and 15) Defensiveness about the way one is eating. Based on the above understanding of ON, a scale will be developed in order to best capture the construct.

I hypothesized that the proposed sub-constructs of the measure, which stem from the operational definition mentioned above, would be: superiority, downward social comparison, rigidity, purity, social avoidance, identity, eating disorder as meaning, loss of control, preoccupation, eating to cope, nutritional deficiencies and relationship problems.

The Measurement of ON

Given the relatively recent defining of orthorexia, measures evaluating the symptoms of this eating pattern are lacking. To date, only two measures have been developed that have been empirically validated to assess orthorexia concerns. The first of these two measures is the ORTO-15, which is based on the dichotomous scale developed by Bratman and Knight (2000). Donini, et al. (2005) expanded Bratman and Knight's (2000) original scale to include 15 items designed to assess symptoms of ON. Responses to each item were based on a 4-point Likert scale, which included, "always," "often," "sometimes," and "never." Scoring for answers that indicate ON is 1 point on the scale, while a score of 4 points on the scale is indicative of normal eating habits. Adding up the scores of each item developed a total score, with total scores below 40 points are defined as meeting criteria for orthorexia. Little validation data exists for this measure, yet it is the most frequently used measure in the small number of existing studies examining orthorexia (Aksoydan & Camci, 2009; Bagci Bosi et al., 2007; Fidan et al., 2010; Varga et al., 2014).

Because no established criteria exist, the ORTO-15 was limited by how their ON groups were formed. Additionally, the ORTO-15 was developed in the Italian language, which likely

changed the scale and validity of the scale when translated into English (Donini et al., 2005). A 4-point Likert-type scale was used, which did not leave the participants a neutral response option; this was problematic in that it may have forced the participants to choose a response that did not apply, skewing the results in various ways. Examples of a few questions are as follows: "Do you allow yourself any eating transgressions?" "Do you think that on the market there is also unhealthy food?" "When eating, do you pay attention to the calories of the food?" There are many concerns that arise from these questions; first, whether people understand the idea of "eating transgressions," as well as whether this question really taps into the construct of ON. Second, it seems likely that the majority of people in the United States would endorse the statement that there is unhealthy food on the market; does this mean that if you endorse this, you are more likely to meet criteria for ON? Third, paying attention to the calories in food does not seem to be related to ON; it seems more likely that the pureness of the food, and whether or not it fits within an individual's strict guidelines, is more pertinent. Thus, it seems that this scale does not truly capture the construct of ON.

Varga and colleagues (2014) examined the psychometric properties of the Hungarian adaptation of the ORTO-15 (ORTO-11-Hu). Their confirmatory factor analysis rejected the original 3-factor structure determined by Donini and colleagues (2014), as well as a single factor structure with the original 15 items. Varga and colleagues (2014) omitted 4 items, and confirmed a single-factor structure for an 11-item version of the scale due to better fit indices (Chi squared =530.8; p <.001). Additionally, using the original cut-off point of 40 on the 15-item version, 74.2% of the sample "had a tendency for ON" (Varga et al., 2014). This seems remarkably high, in that it is highly unlikely that 74.2% of individuals have ON.

The second measure, the Eating Habits Questionnaire (EHQ), was the first ON measure developed in the English language. Gleaves, Graham Ambwani (2013) developed the EHQ to assess the symptoms of ON. The EHQ was found to have 3-factors: a knowledge factor, a problems factor, and a feelings factor. It was also found to have good internal consistency, good test-retest reliability, and evidence for both convergent and discriminant validity. The scale uses a four point Likert scale, which is not recommended when developing scales; 5 point or 7 point Likert scales are preferable (Nunnally & Bernstein, 1994). Additionally, there seems to be some components of ON missing from the measure, which indicates that it is not fully capturing the construct of ON. The scale fails to address an important piece of ON involving the identity and meaning placed on food. Additionally, while self-esteem is alluded to in some of the questions, it may be better if it is asked about directly. Many of the items should be worded differently, including some items that should be reverse scored. Some of the questions are subjective and are not clearly worded; for example, item number 16, "I follow a diet with many rules." Someone with ON who only has one major rule (ex. I will only eat organic fruit) may not endorse this item when, in reality, they would likely meet the criteria for ON. Additionally, there have been no studies empirically validating the efficacy of the EHQ.

Accurately measuring ON is important for furthering the understanding of ON as a distinct construct. The goal of the study presented here is the construction of a new scale to measure ON, and one that addresses the problems with previous scales. This new scale, the Orthorexia Nervosa Scale (ONS), will allow researchers to better understand the construct that is ON, in order to help further determine if differences exist between ON and other disorders.

Knowing more about the differences between ON and other disorders will aid in providing better treatment to these individuals. Nutritionists that I interviewed while gathering

information about ON reported difficulty treating ON in comparison to other eating disorders. Specifically, they mentioned that it hard for their clients to recognize that they are not eating properly because they vehemently believe that they eat healthier than everyone else. This sheds light on the possibility that a different treatment approach may be necessary when working with individuals with ON. If treatment needs to be structured differently for individuals with ON, it is important to know whether ON is its own construct, and therefore its own syndrome, or if it is a subtype of another existing eating disorder. The EHQ (Gleaves, Graham, & Ambwani, 2013) is an empirically sound measure, but it seems to be missing a few major components of ON (e.g. personal identity and meaning) and vaguely words a number of questions that may be misleading. Thus, by aiming to develop a new measure that captures all of the components of ON, it will be more comprehensive, which will eventually lead to improvement of treatment for individuals with ON.

Based on the preliminary and limited ON research that exists, the current study attempted to create a scale that more accurately measures the construct. Thus, the purpose of the study was to construct a scale using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) that would capture the tenants of ON that is more comprehensive than the EHQ (Gleaves, Graham, & Ambwani, 2013).

Initial Scale Construction

The process of developing the ONS items involved multiple steps that began with construction of an initial pool of questions as possible scale items. The first step in this process was a thorough review of relevant research, including reviews of the three other ON scales (Donini et al., Gleaves, Graham, & Ambwani, 2013; Bratman & Knight, 2000). The items in the initial pool for the ONS were constructed using face validity and items from the ORTO-15

(2004), EHQ (2013) and Bratman & Knight's (2000) questionnaire. This process resulted in the construction of 160 items. Of the 160 items originally written, 103 were retained to form the initial version of the ONS. In consultation with faculty members who have expertise in the subject matter and doctoral students who have interest in the subject, the 57 items deleted prior to data collection were removed because they were repetitive, contained colloquial language, or were written in a way that was deemed unanswerable. The goal of this step of scale construction was to complete an EFA and CFA of the 103 items.

METHOD

Participants

Phase 1 data sample. For the first phase of the study, I recruited a sample of psychology doctoral students to complete the initial item review because they are familiar with basic measurement principles and topics related to health related behaviors. Other studies (e.g. Little, Kluemper, Nelson, & Gooty, 2012) have solicited similar types of experts. All doctoral students were asked to rate the representativeness of the items and to comment on general item content and clarity.

Phase 2 data: The development samples. In this phase of the study, two large samples of college students were recruited; one sample was used for an initial EFA and the second sample was used to conduct a CFA. The data collected from both development samples was factor analyzed. There has been in-depth discussion about the ideal number of participants needed to perform factor analysis. In the past, scale development sample sizes have been relatively small (Schmidt & Hunter, 1977). However, Hoelter (1983) recommends a sample size of at least 200. Other researchers have recommended about 5-10 participants per item (Floyd & Widaman, 1995).

I have chosen to recruit samples using Colorado State University Research Subject Pool via Qualtrics. Qualtrics allows researchers to do online data collection and analysis. All surveys were completed online and survey respondents were provided with research credit that is required in introductory psychology classes. While relying on college and university undergraduate pools for behavioral research presents its difficulties, such as lack of diversity, it presents advantages over other online survey resources, such as motivated participants. When

using online survey resources, participants may not be properly motivated to complete some tasks in behavioral research (Holden et al., 2013).

Development sample 1. For the first development sample, 752 total participants were surveyed. When using survey methodology to collect data, it is important to detect lack of motivation and effort that can contribute to random and systematic error in responding (Huang, Curran, Keeney, Poposki, & DeShon, 2012). Per the recommendations made by Huang and colleagues, I observed the response frequency and invariance in responding to exclude cases that did not meet minimum requirements for demonstrating effortful responding.

When response frequency is low (a participant has largely missing data), it may indicate that a participant hurried through the survey at the expense of skipping items to reach the end of the questionnaire and receive research credit. For this sample, only people who answered all of the questions were retained. Another recommendation of Huang et al. (2012) is to exclude participants who demonstrated response invariance. Even when a participant strongly endorses an idea, we can expect that there would be some variation in the subject's response pattern, especially across several different scales and given the reverse-coded items. For this sample, anyone who selected the same response option (e.g. "Agree") for all items was removed from the dataset. A total of 40 participants were dropped due to the above requirements.

Data used in the first development sample (n = 712) included male (29.6%) and female (69.8%) participants from ages 17.5 to 53.4 (M = 19.50, SD = 2.04). Additionally, the sample included college students from a variety of demographic groups, including White (82.6%), American Indian/Native Alaskan (0.6%), Asian/Asian-American (3.7%), Black/African-American (2.7%), and Bi-racial/Multiracial (4.3%). There were a small percentage (2.7%) of participants who chose not to respond. Body Mass Index (BMI) was also collected, with the

mean BMI for this sample being 23.0 (*SD* = 3.8), which indicated that average BMI for this sample fell in the normal range.

Development sample 2. For the second development sample, 397 participants were surveyed. As with development sample 1, the recommendations made by Huang and colleagues (2012) guided the cleaning process. Based on their recommendations, 23 cases were removed from the dataset. Data used in the second development sample (*n* = 374) included male (27%) and female (72.5%) participants from ages 17.7 to 45.4 (*M* = 20.27, *SD* = 3.30). Similar to the first development sample, the participants were from a variety of demographic groups, including White (84.5%), American Indian/Native Alaskan (2.1%), Asian/Asian-American (5.6%), Black/African-American (5.3%), and Bi-racial/Multiracial (4.3%). There were a small percentage of participants (2.9%) that chose not to respond. BMI was also collected, with the mean being 23.11 (*SD* = 4.57), which indicated that average BMI for this sample fell in the normal range. However, BMI was only calculated for 299 participants due to the fact that the height and weight demographic questions were unintentionally left out of the survey for 75 participants.

Procedure

Following the practices and methods for scale development and validation outlined by DeVellis (2012), I created items for my new measure of ON. The guidelines that DeVellis offers are in line with recommendations made by other well-regarded experts in the area of scale development (e.g., Stice, Telch & Rizvi, 2000; Shapiro, Woolson, Hamer, Kalarchian, Marcus & Bulik, 2007; Raycov & Marcoulides, 2010). I used a deductive-type approach to scale development, as it is most appropriate for theory testing, consistent with the purpose of this study, and created a multidimensional scale (Johnson & Christensen, 2008).

When little is known about a construct, exploratory methods like interviews and focus groups are used for initial scale development and testing. However, results from several studies (e.g., Rangel et al., 2007; Donini et al., 2007; Gleaves, Graham, & Ambwani, 2013; Moroze et al., 2014) have led researchers to formulate a conceptualization of ON. These results, in combination with interviewing experts in the field, led to the next logical step in the progression - developing a scale that reflects the theory and verify that it assesses the construct. Then, successive studies can use the measure to determine if the definition and conceptualization of ON is supported with empirical data.

I developed and tested a measure of ON using classical test theory, which assumes that observed scores are a result of a combination of respondents' true scores and error (Nunally & Bernstein, 1994; DeVellis, 2012). Based on this model, error should be random, the amount of error should be equal for all items, and error should be uncorrelated with the true score and other errors.

My scale development followed the recommendations of DeVellis (2012). The first three steps of the scale development process help provide initial content validity evidence for the scale,

as well as help make decisions about which items to retain in the ONS. Steps four, five, and six involve testing and evaluating the scale items, providing initial evidence about the reliability of scores and structure of the scale.

Step 1: Purpose of the Scale. The purpose of this scale is to measure the construct of ON, based on the conceptualization of Bratman & Knight (2000), Moroze et al. (2014) and Gleaves, Graham, & Ambwani, (2013). At this stage, it was also necessary to consider the time frame that the scale will assess (DeVellis, 2012).

A scale that captures current orthorexic behavior tendencies (rather than past behaviors and tendencies) is most useful to initially understand ON. Measuring current orthorexic behaviors and tendencies reduces potential errors due to memory, as past behaviors, especially over a long time period, may be less salient or forgotten (Eisenhower, Mathiowetz, & Morganstein, 2011).

Step 2: Generate the item pool. A good set of items will capture the full content domain of ON (Murphy & Davidshofer, 1991). Moreover, the scale items should fully encompass all behaviors related to ON, exhausting the potential ways to describe what ON is without being redundant. The initial pool of items reflect the conceptualization of ON as described in the literature, pulling primarily from the conceptualization presented in the recent development of the EHQ (Gleaves, Graham, & Ambwani (2013) and Bratman's (2000) original definition. Thinking about potential subconstructs, it is hypothesized that if the full content domain of ON is being assessed, then the likely subconstructs that will result are as follows: superiority, downward social comparison, rigidity, purity, social avoidance, identity, eating disorder as meaning, loss of control, preoccupation, eating to cope, nutritional deficiencies and relationship problems.

DeVellis (2012) recommends generating roughly three or four times the amount of items that the final scale will have. The final number of items is dependent on many factors, so it is best to generate as many items as possible. The properties of the scale will emerge through testing, indicating the ideal number of items to retain for each factor. There is some controversy in the literature regarding the appropriateness of two-item factors. Some researchers have argued that, at the very least, each scale dimension requires at least three items to calculate reliability coefficients (Raykov & Marcoulides, 2011). However, there is emerging literature promoting the appropriateness of two item dimensions (Eisinga, Grotenhuis & Pelzer, 2012).

I began generating items by referencing other ON scales (e.g., Donini et al., 2007; Gleaves, Graham, & Ambwani, 2013), as some of their items are representative of ON. Items were written to be concise and related to the construct, and the original pool contained items with some redundancy to aid in high internal consistency. I also consulted with colleagues (e.g. counseling psychologists, nutritionists, psychiatrists, counseling psychology doctoral students) who are familiar with the content area to generate ideas and write new items that reflects the construct of ON.

Step 3: Determine the format for measurement. During this step, researchers determine the type of response scale that will be used to measure the construct of interest. It is important to consider the format of the response scale because it needs to be compatible with the way items or questions are phrased (DeVellis, 2012), which impacts generating the item pool (Step 2).

A Likert-type response scale seems most appropriate for the ONS because it allows respondents to indicate their levels of agreement or endorsement of items. Other ON scales have measured ON with Likert scales that use response anchors of "False, Not At All True" to "Very

True" (Gleaves, Graham, & Ambwani, 2013) or "Never" to "Always" (Donini et al., 2007). This type of response scale reflects how frequently an individual engages in ON behaviors and the range of response options intentionally or unintentionally assumes that the occurrence of certain behaviors, thoughts and feelings are central to the construct. Thus, the ONS follows a similar type of response scale. ONS items are rated on a 1-5 Likert-type scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree or Disagree, 4 = Agree, 5 = Strongly Agree.

Step 4: Consider including validation items. Before testing the structure of a scale with a development sample, DeVellis (2012) recommends considering the inclusion of other constructs that can be used to obtain validity evidence. Even though testing relationships with other variables is one of the most common ways to obtain validity evidence, it is important to recognize that validity evidence can come from multiple sources. According to the *Standards for Educational and Psychological Testing* (1999), validity evidence can be obtained from sources such as the test content, internal structure of the measure, relationships to other variables, validity generalization, and consequences of the decisions made based on the measure. This project focused on validity evidence obtained from understanding the content and internal structure of the scale, as well from administering the Eating Habits Questionnaire (Gleaves, Graham, & Ambwani, 2013) and the Yale Food Addiction Scale (Gearhart, Corbin & Brownell, 2009). By administering the EHQ (Gleaves, Graham, & Ambwani, 2013), which is the most recent scale developed to measure ON, I measured convergent validity. By administering the YFAS (Gearhart et al.,2009), I examined discriminant validity, because one can posit that ON and food addiction are fairly different constructs. These scales are fully detailed in the *Measures* section below.

Step 5: Test items with a development sample. The survey was administered online to two samples of college students via Colorado State University Qualtrics. The samples needed to be adequately large and representative of the general college population. In addition to rating the ON items, all participants were asked to provide general demographic information such as age, gender, race/ethnicity, and any previous eating disorder history so as to understand and describe the characteristics of the sample.

Because the ONS items will be rated by a single respondent at a single point in time, common method bias may have affected relationships between items or between the scale dimensions (Lindell & Whitney, 2001). To create psychological distance, the survey instructions presented a cover story, making the true purpose of the project (discovering the dimensions of ON) less salient to survey respondents. It can be difficult to create a convincing cover story (Podsakoff, MacKenzie & Podsakoff, 2012), but in the case of this project, it was not difficult to present a broader purpose of the survey rather than revealing its focus of scale development. For example, in the instructions, participants are told that the purpose of the survey is to understand individuals eating habits, but not specifically that the data will be used to create a new scale for ON. Other item characteristic issues, such as social desirability, demand characteristics, and ambiguity, can be addressed during item writing (Podsakoff, Mackenzie, Lee & Podsakoff, 2003). By adhering to these practices, the potential for common method bias should be reduced to a minimum. However, common scale formats and anchors which occur when using Likert-type response scales cannot be eliminated (Podsakoff et al., 2003).

Step 6: Evaluate the items. This step represents the essence of scale development - the analysis of item responses. First, I began by conducting an EFA with the data collected from the first sample, which revealed the initial structure of the scale. Second, I examined basic item

characteristics (e.g. factor loadings) to determine a final set of items to include in the ONS. In subsequent analyses with the first development sample, I pinpointed which items characterize ON best and could be retained to test with the second development sample. With the second development sample data, I conducted a CFA and then examined the reliability estimates for the scale dimensions and confirmed the factor structure of the ONS.

Development sample 1 analysis. There were a total of 104 items included in the development sample 1 survey. A good scale will tap into all components of ON; if items fall on either extreme (being highly endorsed by most people or almost never endorsed), they are only detecting certain levels of ON (DeVellis, 2012). In addition, if all people answer similarly (resulting in low item variance), then the items will not be able to discriminate whether or not a person meets criteria for ON. Ideally, items will have large variances and means close to the central value on the Likert scale, demonstrating that items represent the whole spectrum of ON behaviors, cognitions and emotions. To make sure items are capturing the full range of possible responses, items with means close to the center response (3 on a scale of 1-5), but that also have high variability in responses, are ideal (DeVellis, 2012). This way, the scale can tap into all levels of ON behaviors. I began by evaluating item characteristics, which involved several analyses. First, I calculated the item means and variances. Items were also excluded from exploratory factor analysis if they violated assumptions of normality or if any of the 5 response choices was not endorsed.

Then, I conducted an EFA with the data collected from the first development sample. This method is most appropriate when little is known about a construct (Floyd & Widaman, 1995). Although there is a theory to guide the construction of this scale, when first developing a

measure researchers often conduct an EFA to see if the data reveals a factor structure that is consistent with their expectations. Thus, I followed this often-used approach.

Factor analysis has four primary goals (DeVellis, 2012). Factor analysis provides information that can help a researcher decide how many underlying concepts (also called constructs, dimensions, or factors) are captured in a set of items. The second primary goal of factor analysis is to "condense information" by creating a scale that measures the construct of interest well, but with a smaller subset of items (DeVellis, 2012, p. 117). This allows for the reduction of the full list of generated items to a smaller, more manageable and representative list. The third goal is to determine the meaning of the factors, based both on theoretical justification and the ways in which items covary with each other. Lastly, factor analysis can pinpoint items that are performing well or poorly in the scale dimensions.

For the EFA, I used principal axis factoring (PAF). While principal axis factoring (PAF) and principal components analysis (PCA) may yield similar results about the number and nature of components or factors, PAF treats each X variable as a measurement that may provide information about the same small set of factors as other measured X variables, but at the same time, each X variable may also be influenced by unique sources of error. In PAF, the analysis of data structure focuses on shared variance and not on sources of error that are unique to individual measurements (Wood, Tataryn, & Gorsuch, 1996). Additionally, PAF is more commonly reported in social and behavioral science research reports than PCA, and thus will likely be more familiar to most readers.

The EFA helps determine the number of factors that make up the ONS. Only factors that account for additional variance should be included in the scale (Ford, MacCallum & Tait, 1986). This decision can be made based on eigenvalues, but a scree test and parallel analysis have

shown to be better methods (Ford et al., 1986). Most scales do not include each item as its own factor, but rather, show how items load onto a fewer number of factors. Thus, if any factor has a value greater than one, it is explaining more variance than a single item and can be considered more useful in creating a scale. Kaiser (1960) and Guttman (1954) are often credited with establishing the "eigenvalues greater than one rule," which states that any eigenvalue less than one should be dropped because it does not explain as much variance as the average item. Though this method is widely used, many articles have argued against it, showing that it is unreliable (Bandalos & Boehm-Kaufman, 2008).

Rather than choosing the somewhat arbitrary "greater than one rule," referring to the scree plot or conducting parallel analysis gives more accurate and interpretable information. The scree plot graphically represents the eigenvalues for all the factors, ordering them from greatest to least. There will be a natural breaking point in the plotted line, which indicates where factors are not explaining much additional variance (DeVellis, 2012). When the plot levels off, no additional factors should be retained. This method can be difficult to interpret if there is not a clear breaking point in the data. Thus, parallel analysis is another alternative that is less subjective than a scree plot. With parallel analysis, a random matrix is created of the same size as the actual data. An EFA is performed on the random data and the real data, and then the two matrices are compared. Only factors that have eigenvalues greater than those derived from the random data are retained. Thus, a parallel analysis was used in the current study in order to determine the number of factors that make up the ONS.

Following the results of the parallel analysis, a principal axis factor analysis with oblique rotation was conducted to determine which items loaded onto which factors and which items would be cut from the scale. Factor analysis can use either orthogonal or oblique rotation.

Orthogonal rotation assumes that the factors are uncorrelated. Oblique rotation allows the factors to correlate, modeling a slightly more complex, but possibly more realistic relationship between the dimensions (DeVellis, 2012). This EFA used an oblique rotation, since it is likely that the different dimensions of ON will be related to one another.

Once the number of factors and the type of rotation were determined, I looked at the item factor loadings to decide which items best define the factors. High item loadings are desirable, but item loadings should be at least .30 or .40 indicate that an item fits well with a particular factor (Ford et al., 1986). Low factor loadings can indicate a potential item that may need revised or dropped all together. Additionally, if items cross-load at factors loadings above .35, they were removed from the scale. This process was used to determine the final items to be included in the ONS. Creating the optimal scale length is important because although longer scales tend to be more reliable, they may be unrealistic to administer (DeVellis, 2012). Even though there is no single best way to determine the final items in a scale, the process is guided by the steps of the item evaluation. It is important to consider multiple pieces of information, relying on both the data and theory, to justify decisions to optimize the length of the ONS. Additionally, using multiple pieces of information contributes to the content validity of the ONS.

Reliability. The second sub-step in evaluating the scale items involves calculating the reliability coefficient for each dimension of the scale. Reliability estimates to what extent a scale is measuring the true score of a construct versus error (DeVellis, 2012). There are several methods for obtaining reliability estimates; however, based on a scale's purpose, some are more appropriate than others. For example, test-retest reliability, which assesses to what extent scores remain stable over time (Murphy & Davidshofer, 1991) may be interesting for future studies, but there is not evidence to suggest that ON remains constant and so one would not expect to obtain

this type of reliability evidence. Another common form of reliability is interrater reliability, which refers to consistency across different raters (Murphy & Davidshofer, 1991); however, this stage of the project is not using multiple raters and so this type of reliability evidence is not applicable.

Internal consistency reliability is most relevant for this scale, as it indicates whether all items are related to one another and measure the same construct (DeVellis, 2012). Coefficient alpha is a common measure of internal consistency. This estimate indicates how interrelated items are, which should mean that they are all tapping into the construct similarly (Cronbach, 1951; Raykov & Marcoulides, 2011). Values of about .70 are considered minimally acceptable and values of .90 or above are preferred (Nunnally & Bernstein, 1994). The alpha coefficient specifies how related items are; however, this is not a direct measure of the unidimensionality of the scale (whether a single construct is being measured; Raykov & Marcoulides, 2011). Omega (McDonald, 1999) is a statistic that indicates unidimensionality, and follows similar rules of thumb as the alpha coefficient, with values of .80 - .90 suggesting reliability of scores (Raykov & Marcoulides, 2011). For this study, both the alpha coefficient and omega were calculated.

Thus, to summarize the overall approach for retaining items in the scale, I. I first removed items that operated poorly. Then, as the scale was refined, some items were dropped on the basis of one indicator or another (e.g. factor loadings, cross loadings), but in general, I aimed to use converging evidence and not just a single indicator to retain items for the ONS. Based on the basic item analysis and EFA, I made decisions about which items to retain, revise, or drop from the ONS (see the *Results* section for a full description of these decisions). This revised scale was then tested with a second development sample.

Development sample 2 analysis. The data from the second development sample was used to conduct a CFA. A goal of this study is to confirm (or disconfirm) the structure that has been posited in this proposal based on previous research, which is aligned with a confirmatory factor analysis approach (DeVellis, 2012). The information gained through a CFA determines if items and factors relate in ways that are predicted by existing theory.

For the CFA, I used the maximum likelihood (ML) method for extraction to establish the number of underlying dimensions, also known as factors. The ML method is a preferred method for extraction because it allows for tests of goodness of fit and permits significance testing of factor loadings and correlations among factors (Fabrigar, Wegener, MacCallum, & Straham, 1999). I then specified a 10-factor model that constrained certain items to load on their respective (expected) factors, and then examined fit statistics (discussed in the *Results* section) to see if the model represented the data well. Goodness of fit was evaluated using the chi square (χ^2) statistic, comparative fit index (CFI), and the standardized root mean square residual (SRMR). A nonsignificant χ^2 indicates that the hypothesized model does not significantly deviate from the observed model ($p > .05$). Because χ^2 tests are sensitive to sample size (Ullman & Bentler, 2009), and the current study uses a relatively large sample, it will be beneficial to use fit indices that correct for sample size, such as the CFI. The CFI value ranges from 0 to 1, and indicates the improvement in fit of the hypothesized model compared to a model of complete independence among the measured variables, after controlling for sample size. Values between 0.90 and 0.95 indicate adequate model fit and values of 0.95 and above indicate excellent fit; an excellent fit indicates that at least 95% of the covariation in the data will be reproduced by the hypothesized model (Hu & Bentler, 1999). The SRMR is the square root of the difference between the residuals of the sample covariance matrix and the hypothesized covariance model. SRMR values

range from 0-1, with acceptable models falling between 0.05 and 0.08, and well-fitting models falling below 0.05 (Hu & Bentler, 1999).

As with development sample 1, reliability was also assessed in development sample 2. Internal consistency was calculated for each scale factor using SPSS (Cronbach's alpha) and MPLUS (Omega). Omega captured how homogeneous the items are, demonstrating that they measure a single dimension of ON (McDonald, 1999), and Cronbach's alpha measured how interrelated items are (Raycov & Marcoulides, 2011).

Measures

ON was assessed using the items developed in this study (see the previous section for a description of this process). Appendix A presents the ON scales used with development samples 1 and 2 (removed from document for copyright purposes).

Convergent validity was assessed using the Eating Habits Questionnaire (Gleaves, Graham, & Ambwani, 2013). All 35 items were rated from 1 to 4, *Very True* to *False, Not at all True*. Reliability of this scale has been shown to be good (all alphas for the three subscales are above .82), but the results have yet to be replicated.

Discriminant validity was assessed using the Yale Food Addiction Scale (YFAS) (Gearhart et al.,2009). The content of the YFAS is composed of questions based upon substance dependence criteria in the DSM-IV-TR, and scales used to assess behavioral addictions, such as exercise and sex (Gearhardt et al., 2009). There are 25 items in the scale, and a combination of dichotomous and frequency scoring were used to capture the diagnostic criteria. Questions were adapted to assess the full range of diagnostic criteria related to the consumption of high fat and high sugar foods. This scale has shown good internal reliability (α =.86).

RESULTS

Initial Item Review

Recommendations about item wording and clarity were used to improve the ONS. The initial item review was used to eliminate poorly worded items, double-barreled items (i.e. a single question that asks about more than one issue but only allows for one answer), and items that were perceived to be outside the scope of the orthorexia nervosa construct. This resulted in the scale that was tested with development sample 1. This stage of the process adds content validity evidence for the scale; making sure the scale contains a large number of items that are clearly worded and relevant to the content of interest is an important component of internal consistency reliability, which, in turn, is the foundation of validity (DeVellis, 2012). At this stage of the scale development process, removing too many items could result in a pool of items too small to test with a development sample. Thus, rather than remove all items that all subject matter expert reviewers deemed as a bad item, I primarily used the qualitative comments to revise them.

Development Sample 1 Results

The descriptive analyses and EFA were conducted on the data collected on the 103 items of the ONS using the SPSS statistical software package version 22.0 (SPSS, 2012). Analysis of the descriptive data indicated that, for 101 of the original 103 items, all response choices were indicated and all were normally distributed (see Table 2 for descriptive statistics). Item 78 (I enjoy food) was removed because its kurtosis value was above 3 (3.165) and item 97 (I would never eat an impure food) was removed because it did not capture the full range of item responses. Descriptive analyses indicated that, for all items, all response options were selected and variables were roughly normally distributed (means and standard deviations presented in

Table 2). As such, 101 items were retained for the parallel analyses. Results of the parallel analysis are presented in Table 6. The parallel analysis of the items was conducted to determine the number of factors present in the data using the rawpar syntax implemented in SPSS (O'Conner, 2000). The parameters of the parallel analysis were 1000 random datasets with a confidence interval of 95%. The raw data were used in permutations. Based on the results of this analysis, it was determined that there were 19 factors with significant eigenvalues; a factor is considered significant if the associated eigenvalue is larger than the mean of those obtained from the random, uncorrelated data (Horn, 1965). However, using Kaiser's eigenvalue-greater-than-one rule in conjunction with the parallel analyses, it was determined that there were 10 factors present in the ONS (DeVellis, 2012; Tabachnick & Fidell, 2007; Kaiser, 1960).

Following the parallel analysis, principal axis factoring with oblique, promax rotation was conducted to determine factor loadings for the items that load onto these 10 factors. Items were trimmed if their factor loadings were below .35 or if they had a cross-loading greater than .35 on any other factor. Based on the analysis, it was determined that 47 items would be retained for the final version of the ONS. The 10 factors explained 64.1% of the variance. Factor loadings for the retained items are presented in Table 7 (removed for copyright/publication purposes).

The first factor contains ten items that appear to measure social/interpersonal concerns (see Table 5 for reliabilities). This subscale, Social/Interpersonal Concerns, demonstrated good internal consistency, (Cronbach's $\alpha = 0.92$; $\Omega = .91$), and accounted for 26.89% of the variance. The second factor contains nine items that appear to measure discipline and self control. This subscale, Discipline/Control, demonstrated good internal consistency (Cronbach's $\alpha = .90$; $\Omega = .89$) and accounted for 9.97% of the variance. The third factor contains five items that appear

to measure feelings of superiority and knowledge about healthy eating. This subscale, Superiority/Knowledge, demonstrated good internal consistency (Cronbach's $\alpha = 0.84$; Ω=.86) and accounted for 7.25% of the variance. The fourth factor contains six items that appear to measure the importance of pureness and natural quality of food. This subscale, Pureness/Natural Quality, demonstrated good internal consistency (Cronbach's $\alpha = 0.81$; Ω=.81) and accounted for 3.9% of the variance. The fifth factor contains five items that appear to measure the importance of detoxification and restriction in one's diet. This subscale, Detox/Restricting, demonstrated good internal consistency (Cronbach's $\alpha = 0.75$; $\Omega = .75$) and accounted for 3.7% of the variance. The sixth factor contains two items that appear to measure nutritional deficiencies. This subscale, Nutritional Deficiencies, demonstrated good internal consistency (Cronbach's $\alpha = 0.89$; Ω=.89) and accounted for 3.19% of the variance. The seventh factor contains three items that appear to measure the use of online forums and blogs as a social outlet. This subscale, Online Forums/Blogs, demonstrated good internal consistency (Cronbach's $\alpha = 0.70$; Ω=.68) and accounted for 2.85% of the variance. The eighth factor contains two items that appear to measure defensiveness about one's eating habits. This subscale, Defensiveness, demonstrated good internal consistency (Cronbach's $\alpha = 0.83$; Ω=.83) and accounted for 2.26% of the variance. The ninth factor contains two items that appear to measure feelings of fulfillment and peace when eating healthy. This subscale, Fulfillment/Peace, demonstrated good internal consistency (Cronbach's $\alpha = 0.83$; Ω=.83) and accounted for 2.05% of the variance. The tenth factor contains three items that appear to measure the importance of fasting in one's diet. This subscale, Fasting, demonstrated adequate internal consistency (Cronbach's $\alpha = 0.69$; Ω=.69) and accounted for 2.01% of the variance.

Development Sample 2 Results

Using MPLUS software, I conducted a CFA with maximum likelihood estimation (MPLUS, 2011). Descriptive analyses indicated that, for all items, all response options were selected and variables were roughly normally distributed (means and standard deviations presented in Table 2). Missing data analysis indicated that there were a total of 23 participants who had incomplete data or who answered with the same response for every item (Huang et al., 2012). Table 6 presents the factor loadings for the hypothesized latent factors in the CFA model; all measured variables loaded significantly ($p < 0.001$) onto their hypothesized latent variables. Results of the CFA indicated an acceptable fitting model once model modifications were made, $\chi^2 (980) = 1944.475$, $p < 0.001$, CFI = 0.90, SRMR = 0.06.

In order to obtain an adequate model fit, model modifications had to be performed. The model modification indices indicated that there were certain items that were significantly correlated with each other beyond their shared variance on the factor. In other words, the significance of the correlated error terms indicated that there is something about two specific items within a factor that are more related than all of the items in the factor. The model modifications included correlating 9 pairs of error residuals in order to decrease χ^2 improving overall model fit. In the future, it will important to look at the item pairs that were highly correlated, and make adjustments as necessary, such as removing certain items, in order to achieve a better fitting model (see *Discussion* section for more detail).

Although improved fit is desirable, because this is the first test of the scale, the model fit was deemed acceptable after adding these 9 correlations. Additionally, because the items loaded strongly on their respective dimensions, the evidence presented here shows good initial support

for a 10-factor measure of ON. However, adjustments to the ONS items are necessary and more data must be collected in order to reexamine the ONS factor structure.

Development sample 2 reliability. Internal consistency was calculated for each scale factor (see Table 4) using SPSS (Cronbach's alpha) and MPLUS (Omega). Both of these indicators assess the scale's reliability. Eight of the ten dimensions of the ONS showed good reliability of scores, 1) social/interpersonal concerns: $\alpha = .89$, $\Omega = 86$; 2) discipline/self-control: $\alpha = .91$, $\Omega = .91$, 3) superiority/knowledge: $\alpha = .87$, $\Omega = .87$, 4) pureness/natural quality: $\alpha = .83$, $\Omega = .84$, 5) detox/restricting: $\alpha = .76$, $\Omega = .75$, 6) nutritional deficiencies: $\alpha = .92$, $\Omega = .92$, 7) online forums/blogs: $\alpha = .63$, $\Omega = .62$, 8) defensiveness: $\alpha = .90$, $\Omega = .91$, 9) fulfillment/peace: $\alpha = .82$, $\Omega = .82$, and 10) fasting: $\alpha = .70$, $\Omega = .69$. Factors 7 and 10 have inadequate reliability scores.

Construct validity evidence. Using both development sample 1 and 2 data, the ONS was correlated with the EHQ and YFAS to assess convergent and discriminant validity, respectively. To do this, a total scale score was calculated for the ONS, EHQ, and YFAS in order to calculate a correlation. First, however, in order to determine if a total scale score could be calculated for the ONS, a second order factor structure was run to see if all 10 factors significantly loaded onto an 11^{th} factor. Because all 10 factors significantly loaded onto a second order factor structure, it was determined that a total scale score across all 10 factors could be computed.

The total scale score for the ONS, EHQ and YFAS were 235, 140 and 22, respectively (see Table 8 for means and standard deviations). As I hypothesized, there was a large, significant correlation between the ONS and the EHQ ($r = -.74$, $p < .01$), as the ONS and EHQ are measuring the same construct. The reason the two scales are negatively correlated is because lower scores on the EHQ indicate more ON symptomology whereas higher scores on the ONS

indicate more ON symptomology; in other words, higher scores on the ONS are equivalent to lower scores on the EHQ. Additionally, there was a small, significant correlation between the ONS and the YFAS, (r =.15, p <.01), which provides evidence for discriminant validity; I expected that the YFAS and the ONS would not be highly correlated with each other due to the fact that they are measuring very different eating patterns and behaviors.

DISCUSSION

Through the steps outlined in this paper, a multidimensional measure of ON was developed, tested and partially supported. Exploratory factor analyses revealed that there were 10 significant factors present in the data. The next step was to examine the context of the subscales to estimate the constructs they were measuring. This led to the labeling of 10 subscales and a descriptions of the constructs they were hypothesized to measure: (1) social/interpersonal concerns, 2) discipline/self-control, 3) superiority/knowledge, 4) pureness/natural quality, 5) detox/restricting, 6) nutritional deficiencies, 7) online forums/blogs, 8) defensiveness, 9) fulfillment/peace, and 10) fasting (see Table 6 for specific items). Based on the original hypothesis, the scale reflects components of ON theory that are not included in already existing measures (e.g. superiority, downward social comparison, nutritional deficiencies, rigidity, purity, social avoidance, relationship problems, identity, preoccupation, and relationship problems).

The first subscale contained 10 items that all appear to be related to social and interpersonal concerns. Many of the items were related to social avoidance, going out less in order to avoid eating certain foods, and spending free time making sure all foods eaten are healthy. This subscale also included items that reflect relationship problems and concerns. Existing diagnostic criteria support this factor, positing that a person with ON is so selective about their food choices it will lead to a loss/impairment in social relationships (Moroze et al., 2014; Chaki et al., 2013). The second subscale consisted of items related discipline, self-control and self-esteem. Items that assessed feeling a sense of control when eating healthily and feeling a higher sense of self-esteem when eating a pure diet are included in this subscale. Additionally, items that relate to feeling distressed when not eating healthy, as well as feeling as though eating healthy is something the individual *must* do. Based on the research into the proposed

psychological consequences of ON that elucidate the frustration that a person feels when their food practices are disrupted, as well as disgust when the pureness of one's diet is compromised, this factor fits well as a sub-construct (Mathieu, 2005). The third subscale is comprised of items related to superiority and food knowledge. The five items in this subscale point to feelings of superiority about ones diet, as well as having a larger knowledge base about healthy foods than most people. Bratman & Knight (2000) discuss how a person with ON will spend hours researching food, which is a large component of this sub-contruct. This second, but related piece is connecting the greater knowledge one feels they have about healthy food with a sense of feeling superior and downward social comparison. Existing scales and the current literature have left out this idea; however, it appears to be a component of the construct. The fourth subscale captures a large component of ON – the focus on the pureness and natural quality of food. The items in this factor address the importance of pureness in the foods eaten, as well as a strict avoidance of artificial substances, pesticides and herbicides. Previous research supports the inclusion of this construct, highlighting that a main component of ON is a preoccupation with food purity (Bratman & Knight, 2000; Moroze et al., 2014; Catalina Zamora et al., 2005). The fifth subscale is comprised of items that relate to the importance of detoxification and restriction of certain foods/calories as a way to cleanse the body. As mentioned above, the idea of feeling pure is paramount in ON, and achieving this sense of purity comes through detoxification and restriction of certain foods (Moroze et al., 2014; Bagci Bosi et al., 2007; Bratman & Knight, 2000).

The sixth subscale focuses on nutritional deficiencies, which are discussed in the literature as criteria for ON. Individuals with ON will eventually exhibit signs of impaired physical health due to nutritional imbalances (Moroze et al., 2014). Restricting certain nutritional

elements from one's diet in order to feel "pure" will eventually lead to nutritional deficiencies that may be harmful to one's health. The items in this subscale indicate that nutritional deficiency comprises its own subconstruct of ON. The seventh subscale is related to the social media component that appears to be a component of ON. Research on other eating disorders (e.g. AN, BN) have discussed the importance of an online community (Brotsky & Giles, 2007;Wilson, Peebles, Hardy & Litt, 2006). This appears to be similar for individuals with ON. The items in this subscale are related to online forums and blogs being an individual's main social outlet, as well as an assessment of how much time is spent online reading about food.

The eighth subscale includes items related to defensiveness about the way one is eating. When food rules evolve that center around eating food that the individual considers to be clean and pure, that individual's behavior becomes increasingly rigid under the oppression of the many food rules (Bratman & Knight, 2000). Eating becomes a controversial topic, and creates conflict between the individual with ON and others who view their lifestyle as unhealthy. This may lead to defensiveness about the way a person with ON is eating. The ninth subscale is focused on feelings of fulfillment and peace an individual with ON feels when eating pure and healthily. These feelings of fulfillment help drive the individual to continue striving for a perfectly pure diet (Moroze et al., 2007). Feelings of fulfillment and peace seem to be related to an identity component associated with eating a certain way, although there are no items in the subscale that specifically address identity. The tenth and final subscale is focused on fasting. It appears as though many individuals with ON will choose to not eat if certain foods available to them are not considered pure by their standards (Nymah, 2002; Bagci Bosi et al., 2007). There also seems to be the belief in this subscale that fasting is an effective way to lose weight – this idea seems to be more related to other eating disorders, as the focus of ON is not to lose weight, but to maintain a

"pure" diet (Habermas, 1996; Bratman & Knight, 2000). Thus, a deeper examination into this subscale must be conducted in order to determine the validity of this overlap.

Using this factor solution led to a theoretically plausible scale with 10 subscales that contained items that loaded on their subscales at greater than 0.35. Additionally, the items contained in the ONS loaded significantly on their specified factor only, with crossloadings of less than 0.30 onto the other subscales.

The overall results of the EFA provide strong evidence for the value of the ONS. The scale explained a large amount of variance, and reliability analysis indicated that the subscales were relatively unidimensional. Additionally, the ONS captures many of the hypothesized components of ON that previous scales have failed to capture, such as feelings of superiority and fulfillment one feels with one's pure diet. The next step in the scale development process was to conduct a CFA to replicate these results with a new sample of participants.

Results from the CFA indicate that the hypothesized factor structure that was discovered in development sample 1 was replicated in development sample 2 once model modifications were made. Model modifications needed to be made to the ONS based on the fit indices. There were a number of items that correlated highly with one another, and upon further examination, the items were largely similar to one another. For example, two of the items "I strictly avoid food that has been treated with artificial substances" and "I strictly avoid food that has been treated with pesticides and herbicides" appear to be very similar items. For the purposes of this analysis, the items that were highly correlated with one another were added to the model, thus accounting for the correlation and improving the model fit.

Once these model modifications were made, the model had good fit and all the variables loaded significantly onto their factors. There was no indication of significant cross-loadings.

Eight of the ten subscales showed good internal consistency reliability. As further evidence in support of the ONS, ON was positively related to the EHQ, two scales which are conceptually measuring ON. Additionally, the ONS was not related to the YFAS, a measure of binge eating, which is evidence of discriminant validity. This provides initial construct validity evidence for the ONS and contributes to our understanding of ONS as it relates to other important eating behaviors.

Conclusions about the Scale

While numerous studies indicate a high prevalence rate of ON symptoms (Aksoydan & Camci, 2009; Bagci Bosi, Camur & Guler, 2007; Donini et al., 2004; Segura-Garcia et al., 2012). there are concerns regarding the accuracy of these incident rates (Gleaves, Graham, & Ambwani, 2013). The development of the ONS provides directions for further assessment of this construct and its psychological parallels which addresses some of the measurement issues of existing scales previously discussed. The current study also expands upon the definition of the ON construct, suggesting that problems with healthy eating involve multiple components, such as feelings of superiority, social and interpersonal concerns, and a focus on pureness. These subconstructs reflect components of ON that have been discussed in the literature but have not been included in any scales thus far. More research needs to be done that explores the relationship between ON and personality functioning, social desirability and general psychopathology.

The ONS can be used in future research, particularly aiding in the understanding of the discriminability between ON and AN, as well as ON and OCD, as well as establishing the exclusive factors of ON.

Limitations of the Study

This study is not without its limitations. First, two of the factors did not have adequate reliability estimates – factors 7 and 10 had estimates falling just below .70. As previously discussed, values of about .70 are considered minimally acceptable and values of .90 or above are preferred (Nunnally & Bernstein, 1994).

Secondly, the CFA was deemed to have an acceptable model fit only after model modifications were made. The modifications that were made were largely due to certain items being highly correlated with each other, which indicated that they are extremely similar. This means that there are items in the scale that are asking very similar questions, which makes the overall model less of a good fit if certain items are more correlated with each other than each item and the scale itself. In order to achieve a good fitting model, it will be important to examine these items and make adjustments (e.g. removing and revising items). Then, more data will be collected and a new CFA will be run with the goal of improving overall model fit without making any model modifications.

Because this study was conducted in stages, information and identified patterns were synthesized to understand how to best measure ON. However, each stage of the process relied on self-report data at a single time point, which may have introduced common method bias (Podsakoff et al., 2012). Some preventative efforts, such as increasing psychological and proximal distance were taken in the study to reduce the bias, but future studies should collect data at multiple points in time from multiple sources. Lastly, participants were predominantly White college students, which limits the ability to generalize findings to different ages and racial/ethnic groups outside of this college student population.

Future Steps and Directions

My directions for future research. To extend this study, I plan to rerun the CFA with a new development sample once the concerns that arose in the present study are addressed. The goal is to change, modify and add items to the scale in order to establish a better model fit. For example, two of the factors had low alpha and omega values, which could be due to a low number of questions or heterogeneous constructs. Thus, I will revise and discard certain items after careful examination of each item (Tavakol & Dennick., 2011). Additionally, when collecting data with a new development sample, I will carefully examine the items that had large correlations based on the model modification indices. If two items appear to be asking the same question in a different way, I will remove one of those items. This will likely improve the model fit.

General directions for future research. Future research should replicate the findings of this study, confirming the factor structure with other samples (Floyd & Widaman, 1995). These studies should also compare ON to other well-established constructs to continue to contribute to construct validity evidence. Research regarding ON is in its infancy, and future research examining its relatedness to other constructs will help to establish whether or not ON is even its own distinct construct demonstrates in what way ON is related to (or not) other existing eating disorders, it will be important to understand what those similarities and differences are from an empirical standpoint. Although ON was compared to two other constructs, more convergent and discriminant validity evidence is needed. By comparing individuals with ON, AN, and OCD on related constructs, such as body image or obsessiveness, research could help clarify whether or not ON is a distinct construct (Gleaves, Graham, & Ambwani, 2013). Additionally, as Gleaves, Graham, & Ambwani (2013) suggested, future research would also help to clarify individuals

who may be at high risk for developing ON. The content and face validity evidence, in addition to the reliability evidence that has already been gathered in this study, is a necessary foundation for obtaining validity evidence in future research.

Although conclusions are made about the underlying factor structure of a construct based on CFA results, the structure could be a function of the sample that was used (DeVellis, 2012). Thus, it is recommended to test the factor structure with a cross-validation sample (Floyd & Widaman, 1995), confirming that the structure emerges with other representative groups of people. It will be important for the next study to replicate the ten-factor structure derived in this study, providing further support for the measure.

Lastly, the development of norms for the ONS is an important next step. The ONS is a trait-based measure of ON, and may be an important tool to help identify risk in individuals. This measure is not designed to diagnose individuals with ON, but developing norms is an important step such that raw score are essentially meaningless unless they are accompanied by relevant data that places the scores in a meaningful, interpretive context. Ideally, it will be important for ONS scores to be comparable to scores earned by members of a defined population, such as other college students or other individuals diagnosed with an eating disorder. Such norms should take the form of a cut-off score, such that those scoring above or below a certain score fall in an at-risk category for ON symptoms.

CONCLUSION

Orthorexia Nervosa is not currently recognized as a clinical diagnosis in the DSM-5, likely because there has been a paucity of research examining the construct of ON, as well as the is distinctiveness and similarities to other existing psychological disorders. Although there has been concern over the increasing prevalence of individuals who report symptomology related to ON, we have not had a measure that could comprehensively identify individuals who experience ON. This study created a new measure for ON with ten dimensions that is more inclusive than the existing measures. With the new measure of ON, there are some exciting directions to take and researchers and clinicians can learn more about individuals' ON symptomology, as well as help identify individuals who are experiencing symptoms of ON. This measure is also a first step in empirically identifying whether ON is a distinct disorder, or if it can be explained by an already existing psychological disorder, such as anorexia nervosa or obsessive compulsive disorder.

TABLES

Table 1

Orthorexia nervosa diagnostic criteria proposed by Moroze et al.

Criterion A: Obsessional preoccupation with eating "healthy foods," focusing on concerns regarding the quality and composition of meals. (Two or more of the following.)

- Consuming a nutritionally unbalanced diet owing to preoccupying beliefs about food "purity."
- Preoccupation and worries about eating impure or unhealthy foods and of the effect of food quality and composition on physical or emotional health or both.
- Rigid avoidance of foods believed by the patient to be "unhealthy," which may include foods containing any fat, preservatives, food additives, animal products, or other ingredients considered by the subject to be unhealthy.
- For individuals who are not food professionals, excessive amounts of time (eg., 3 or more hours per day) spent reading about, acquiring, and preparing specific types of foods based on their perceived quality and composition.
- Guilty feelings and worries after transgressions in which "unhealthy" or "impure" foods are consumed.
- Intolerance to other's food beliefs.
- Spending excessive amounts of money relative to one's income on foods because of their perceived quality and composition.

Criterion B: The obsessional preoccupation becomes impairing by either of the following:

- Impairment of physical health owing to nutritional imbalances (eg., developing malnutrition because of an unbalanced diet).
- Severe distress or impairment of social, academic, or vocational functioning owing to obsessional thoughts and behaviors focusing on patient's beliefs about "healthy" eating.

Criterion C: The disturbance is not merely an exacerbation of the symptoms of another disorder such as obsessive-compulsive disorder or of schizophrenia or another psychotic disorder.

Criterion D: The behavior is not better accounted for by the exclusive observation of organized orthodox religious food observance or when concerns with specialized food requirements are in relation to professionally diagnosed food allergies or medical conditions requiring a special diet.

Table 2

Demographic Data

Variable	EFA Sample	CFA Sample
Age (M, SD)	19.5 (2.04)	20.27 (3.30)
Race (%)		
American Indian/Native Alaskan	0.6	2.1
Asian, Asian-American	3.7	5.6
Black/African-American	2.7	5.3
Native Hawaiian/Pacific Islander	0.3	1.1
White	82.6	84.5
Bi-racial/Multiracial	7.6	4.3
Do Not Wish to Respond	2.7	2.9
Ethnicity (%)		
Hispanic	--	12.6
Not Hispanic	--	83.7
Do Not Wish to Respond	--	3.5
Sex (%)		
Male	29.6	27.0
Female	69.8	72.5
Transgender	0.4	0.3
Other	0.1	0.3
BMI (M, SD)	23.0 (3.8)	23.11 (4.57)

Table 3

Means and Standard Deviations

Item	EFA	CFA
	M (*SD)*	*M* (*SD*)
1.	3.2 (.94)	3.3 (.98)
2.	2.7 (.84)	2.7 (.87)
3.	3.2 (.96)	3.3 (1.0)
4.	2.9 (.91)	3.0 (1.0)
5.	2.5 (.83)	--
6.	3.7 (.79)	--
7.	2.6 (.92)	--
8.	2.2 (.73)	--
9.	2.5 (.84)	2.6 (.90)
10.	2.8 (.85)	2.9 (.96)
11.	2.5 (.92)	--
12.	2.1 (.87)	1.9 (.93)
13.	1.9 (.87)	1.7 (.76)
14.	1.9 (.78)	1.7 (.75)
15.	1.9 (.81)	1.8 (.89)
16.	2.1 (.94)	1.9 (1.1)
17.	2.0 (.85)	1.8 (.88)
18.	2.7 (1.0)	--
19.	2.8 (.98)	--
20.	2.2 (.80)	--
21.	2.1(.84)	1.9 (.89)
22.	2.6 (1.0)	--

23.	2.6 (.99)	--
24.	2.2 (.78)	--
25.	2.4 (.93)	--
26.	2.2 (1.1)	--
27.	2.2 (.99)	--
28.	2.2 (1.2)	1.9 (1.1)
29.	3.0 (1.1)	3.0 (1.1)
30.	2.7 (1.0)	2.6 (1.1)
31.	2.2 (.96)	2.0 (1.0)
32.	2.0 (.90)	1.9 (.98)
33.	2.1 (.97)	1.9 (.97)
34.	3.2 (1.2)	--
35.	1.9 (.82)	1.7 (.82)
36.	1.9 (.75)	1.7 (.83)
37.	1.9 (.75)	--
38.	2.0 (.85)	--
39.	3.4 (1.1)	--
40.	2.0 (.80)	1.6 (.79)
41.	2.8 (1.1)	2.5 (1.2)
42.	2.2 (1.0)	2.1 (1.1)
43.	1.8 (.75)	--
44.	3.4 (.95)	3.7 (.92)
45.	3.4 (94)	3.6 (.93)
46.	3.2 (.93)	--
47.	2.6 (1.2)	--
48.	2.5 (.99)	--
49.	3.8 (.77)	--

50.	3.3 (98)	--
51.	2.9 (.99)	--
52.	3.7 (.84)	--
53.	2.9 (.90)	2.7 (.96)
54.	2.4 (.83)	2.4 (.92)
55.	2.4 (.83)	2.4 (.95)
56.	2.8 (.99)	2.8 (1.1)
57.	2.3 (.81)	--
58.	2.0 (.81)	1.7 (.84)
59.	2.6 (1.2)	2.7 (1.3)
60.	3.3 (1.1)	--
61.	2.8 (1.0)	--
62.	2.3 (.98)	2.2 (1.1)
63.	2.2 (.96)	2.1 (1.1)
64.	1.9 (.80)	--
65.	1.9 (75)	--
66.	3.2 (1.0)	3.3 (1.1)
67.	3.2 (1.1)	3.4 (1.0)
68.	2.6 (1.0)	--
69.	2.3 (.94)	--
70.	2.6 (.98)	2.9 (1.0)
71.	2.2 (.91)	--
72.	2.6 (.93)	--
73.	2.5 (1.0)	--
74.	2.4 (.92)	2.5 (1.1)
75.	2.3 (.90)	2.4 (1.0)
76.	1.9 (.77)	--

77.	1.9 (.78)	--
78. *	4.4 (.77)	--
79.	1.9 (.82)	--
80.	2.0 (.83)	--
81.	3.0 (1.0)	3.3 (1.1)
82.	3.9 (.86)	--
83.	3.7 (.76)	--
84.	2.6 (.99)	--
85.	2.2 (.86)	--
86.	3.0 (.92)	--
87.	2.3 (.88)	--
88.	3.6 (1.0)	--
89.	2.3 (.99)	--
90.	2.7 (1.0)	2.7 (1.1)
91.	2.9 (1.0)	2.9 (1.1)
92.	2.2 (.92)	--
93.	2.3 (.92)	--
94.	3.4 (98)	3.4 (1.1)
95.	2.7 (.97)	2.6 (1.1)
96.	2.1 (.82)	--
97. *	1.9 (71)	--
98.	3.2 (1.1)	3.3 (1.2)
99.	2.6 (1.0)	2.7 (1.2)
100.	3.0 (1.1)	--
101.	2.0 (.82)	--
102.	2.3 (.96)	--
103.	1.9 (.78)	1.8 (.85)

104.	2.9 (1.0)	--

*=items were removed based on descriptive statistics

Table 4

Development sample 1 correlation table and reliabilities for orthorexia nervosa factors

Factor	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	.92									
2	.32	.90								
3	.31	.36	.84							
4	.50	.47	.46	.81						
5	.58	.56	.23	.43	.75					
6	.31	.21	-.14**	.16	.42	.89				
7	.71	.45	.34	.53	.58	.39	.70			
8	.40	.41	-.03**	.24	.38	.51	.44	.83		
9	.12**	..64	.37	.31	.39	-.03**	.18	.11**	.83	
10	.60	.31	.03**	.31	.76	.46	.57	.43	.13	.69

Note. ** indicates p>.05 for correlations. For reliabilities, the alpha is reported.

Table 5

Development sample 2 correlation table and reliabilities for orthorexia nervosa factors

Factor	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	(.89, .86)									
2	.58	(.91, .91)								
3	.36	.43	(.87, .87)							
4	.59	.67	.56	(.83, .84)						
5	.70	.55	.19	.49	(.76, .75)					
6	.38	.27	-.23	.17	.38	(.92, .92)				
7	.83	.63	.21	.54	.61	.50	(.63, .62)			
8	.54	.58	.08**	.30	.50	.37	.53	(.90, .91)		
9	.37	.75	.48	.62	.36	.05**	.28	.30	(.82, .82)	
10	.62	.46	-.00**	.39	.87	.48	.63	.48	.16	(.70, .69)

Note. ** indicates p>.05 for correlations. For reliabilities, alpha is reported first, followed by omega

Table 6

Results of the Parallel Analysis

Factor	Eigenvalue	95% Eigenvalue
1	25.634	1.023
2	8.300	0.968
3	4.720	0.927
4	2.212	0.893
5	1.871	0.862
6	1.562	0.833
7	1.495	0.806
8	1.238	0.781
9	1.159	0.757
10	1.010	0.734
11	0.930	0.711
12	0.832	0.690
13	0.790	0.669
14	0.764	0.649
15	0.724	0.630
16	0.683	0.611
17	0.653	0.592
18	0.591	0.574
19	0.562	0.557
20	0.529	0.540

Eigenvalue > 95% Eigenvalue = Significant Factor

Table 8

Means and standard deviations for the ONS, EHQ, and YFAS total scale scores

Scale	Total Scale Score	Standardized Scale Score
ONS (M, SD)	117.89 (23.27)	2.51 (0.50)
EHS (M, SD)	103.49 (15.22)	2.96 (0.44)
YFAS (M, SD)	9.37 (2.26)	0.43 (0.10)

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing. (1999). *Standards for educational and psychological testing*. American Educational Research Association. United States.

Aksoydan, E. & Camci, C. (2009). Orthorexia nervosa among Turkish performance artists. *Eating and Weight Disorders – Studies on Anorexia Bulimia and Obesity, 14(1*), 33-37.

Affenito, S. G., Dohm, F. A., Crawford, P. B., Daniels, S. R.& Striegel-Moore, R. H. (2002). Macronutrient intake in anorexia nervosa: The national heart, lung, and blood institute growth and health study. *Journal of Pediatrics, 141* (5), 701-705.

Bagci Bosi, A. T., Camur, D., & Guler, C. (2007). Prevalence of orthorexia nervosa in resident medical doctors in the faculty of medicine (Ankara, Turkey). *Appetite, 49*, 661-666.

Bandalos, D. & Boehm-Kaufman, M. (2008). *Four common misconceptions in exploratory factor analysis.* In Lance, C. & Vandenberg, R. (Eds.). *Statistical and Methodological Myths and Urban Legends.* New York: Routledge.

Bartrina, J. A. (2007). Orthorexia or when a healthy diet becomes an obsession. *Archivos LatinoAmericanos de Nutricion, 54*, 313-315.

Bratman, S. & Knight, D. (2000). *Health food junkies: Orthorexia nervosa - overcoming the obsession with healthful eating.* New York: Broadway Books.

Brotsky, S. R., & Giles, D. (2007). Inside the "pro-ana" community: a covert online participant observation. Eating Disorders: The Journal of Treatment & Prevention, 15(2), p. 93-109. doi: 10.1080/10640260701190600

Brytek-Matera, A. (2012). Orthorexia nervosa – an eating disorder, obsessive-compulsive disorder or disturbed eating habit? *Archives of Psychiatry and Psychotherapy, 1*, 55-60.

Brytek-Matera, A., Donini, L. M., Krupa, M., Poggiogalle, E., & Hay, P. (2015). Orthorexia nervosa and self-attitudinal aspects of body image in female and male university students. *Journal of Eating Disorders, 3*, 2.

Bulik, C. M., Sullivan, P. F., Tozzi, F., Furberg, H., Lichtenstein, P., Pedersen, N. L. (2006). Prevalence, heritability, and prospective risk factors for anorexia nervosa. *Archives of General Psychiatry, 63*, 305–312.

Catalina Zamora, M. C., Bote Bonaechea, B., Garcia Sanchez, F., & Rios Rial, B. (2005). Orthorexia nervosa: A new eating behavior disorder? *Actas Esp Psiquiatr., 33(1),* 66–68.

Chaki, B., Pal, S., & Bandyopadhyay, A. (2013). Exploring scientific legitimacy of orthorexia nervosa: A newly emerging eating disorder. *Journal of Human Sport & Exercise, 8* (4), 1045-1053.

Coleman, A. M. (2015). *Dictionary of Psychology*. Oxford University Press, Oxford, UK.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334. doi:10.1007/BF02310555

Curran P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychologial Methods, 1(1),* 16-29.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (Vol. 26). Thousand Oaks, CA: Sage Publications, Inc.

Donini, L. M., Marsili, D., Gaziani, M. P., Imbriale, M., & Cannella, C. (2004). Orthorexia nervosa: A preliminary study with a proposal for diagnosis and an attempt to measure the dimension of the phenomenon. *Eating Weight Disorders, 9*, 151–157.

Donini, L. M., Marsili, D., Gaziani, M. P., Imbriale, M., & Cannella, C. (2005). Orthorexia nervosa: Validation of a diagnosis questionnaire. *Eating Weight Disorders, 10,* 28–32.

Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (2011). Recall error: Sources and Bias reduction techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman, (Eds.). *Measurement Errors in surveys* (pp. 127-144). John Wiley & Sons, Inc.

Eisinga, R., Grotenhuis, M. t., & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International Journal of Public Health, p. 1-11*. doi: 0.1007/s00038-012-0416-3

Eriksson, L., Biagi, A., Marklund, B., & Lindgren, E. C. (2008). Social physique anxiety and sociocultural attitudes toward appearance impact on orthorexia test in fitness participants. *Scandinavian Journal of Medicine & Science in Sports, 18,* 389-394.

Fidan, T., Ertekin, V., Işikay, S. & Kirpinar, I. (2010). Prevalence of orthorexia among medical students in Erzurum, Turkey. *Comprehensive Psychiatry, 51,* 49-54.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299.

Ford, J. K., MacCallum, R. C. & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39,* 291-314.

Gearhardt, A. N., Corbin, W. R., & Brownell, K. D. (2009). Preliminary validation of the Yale Food Addiction Scale. *Appetite, 52,* 430-436.

Getz, L. (2009). Orthorexia: When Eating Healthy Becomes an Unhealthy Obsession. *Today's Dietitian, 11*(6), 40.

Gleaves, D. H., Graham, E. C., & Ambwani, S. (2013). Measuring "orthorexia": Development of the Eating Habits Questionnaire. *The International Journal of Educational and Psychological Assessment, 12*, 1-18.

Guttman, L. (1954). A new approach to factor analysis: the Radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258-348). New York, NY: Free Press.

Haberman, T. (1996). In defense of weight phobia as the central organizing motive in anorexia nervosa: Historical and cultural arguments for a culture-sensitive psychological conception. *International Journal of Eating Disorders, 19* (4), 317-334.

Hoelter, J. W. (1983). The analysis of covariance structures goodness-of-fit indices. *Sociological Methods and Research, 11*(3), 325-344.

Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon‟s Mechanical Turk. *Computers in Human Behavior*, *29*(4), 1749-1754.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179-185.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99-114. doi: 10.1007/s10869-011-9231-8

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. Thousand Oaks, CA: Sage Publications.

Pilzer, P. Z. (2002) *The Wellness Revolution: How to Make a Fortune in the Next Trillion Dollar Industry*. New York: John Wiley and Sons.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*141-151. doi:10.1177/001316446002000116

Kickbusch, I. & Payne, L. (2003). *Twenty-first century health promotion: The public health revolution meets the wellness revolution*. Oxford University Press. doi: 10.1093/heapro/dag418

Kinzl, J.F., Hauer, K., Traweger, C. & Kiefer, I. (2005). Orthorexia nervosa: einehaufigeEsssto¨rungbeiDia¨tassistentinnen? (Orthorexia nervosa: a frequent eating disorder in dieticians?) *Erna hrungs-Umschau, 52,* 436-439.

Kummer, A., Dias, F.M.& Teixeira, A.L. (2008). On the concept of orthorexia nervosa. *Scandinavian Journal of Medicine & Science in Sports, 18* (3), 395-396.

Lewinsohn, P. M., Seeley, J. R., Moerk, K. C., & Striegel-Moore, R. H. (2002). Gender differences in eating disorder symptoms in young adults. *International Journal of Eating Disorders, 32,* 426-440.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86,* 114-121.

Lindeman, M., Keskivaara, P., & Roschier, M. (2000). Assessment of magical beliefs about food and health. *Journal of Health Psychology, 5*, 195-209.

Little, L. M., Kluemper, D., Nelson, D. L., & Gooty, J. (2012). Development and validation of the Interpersonal Emotion Management Scale. *Journal of Occupational & Organizational Psychology, 85*, 407-420. doi:10.1111/j.2044-8325.2011.02042.x

Mathieu, J. (2005). What is orthorexia? *Journal of the American Dietetic Association, 10,* 1510-1512.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press.

Misra, M., Tsai, P., Anderson, E. J., Hubbard, J. L., Gallagher, K., Soyka, L. A.& Klibanski, A. (2006). Nutrient intake in community-dwelling adolescent girls with anorexia nervosa and in healthy adolescents. *American Journal of Clinical Nutrition, 84* (4), 698-706.

Moroze, R. M., Dunn, T. M., Holland, J. et al. (2014) Microthinking about micronutrients: a case of transition from obsessions about healthy eating to near-fatal "orthorexia nervosa" and proposed diagnostic criteria. *Psychosomatics.*

MPLUS Version 7.4. (2012). [Computer Software]. Los Angeles, CA: Muthén & Muthén.

Murphy, K. R., & Davidshofer, C. O. (1991). *Psychological testing*. Prentice-Hall.

Nicolosi, G. (2006). Biotechnologies, alimentary fears and the orthorexic society. *Tailoring Biotechnologies, 2(3),* 37-56.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory.* United States of America: McGraw-Hill.

Nymah, H. (2002). A direct question: is orthorexia a correct word for a wrong concept? *Lakartdningen, 99,* 433-434.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32,* 396-402.

Park, S. W., Kim, J. Y., Go G. J., et al. (2011). Orthorexia Nervosa with hypoatremia, subcutaneous emphysema, pneumomediastimum, pneumothorax, and pancytopenia. *Electrolyte Blood Press, 9,* 32-37.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539-569. doi:10.1146/annurev-psych-120710-100452

Podsakoff, P. M., MacKenzie, S. B., Lee, J-Y, & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903. doi: 10.1037/0021-9010.88.5.879

Rangel, C., Dukeshire, S., & MacDonald, L. (2012). Diet and anxiety: An exploration into the orthorexic society. *Appetite, 58,* 124-132.

Raykov, T. & Marcoulides, G. A. (2011). *Introduction to psychometric theory.* New York, NY: Taylor and Francis Group, LLC.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*(5), 529-540.

Schneider, N., Martus, P., Ehrlich, S., Pfeiffer, E., Lehmkuhl, U., & Salbach-Andrae, H. (2009). The assessment of body image distortion in female adolescents with anorexia nervosa: The development of a test for body image distortion in children and adolescents (BID-CA). *Eating and Weight Disorders – Studies on Anorexia, Bulimia and Obesity, 14 (2*), 128-136.

Segura-Garcia, C., Papaianna, M. C., Caglioti, F., Procopio, L., Nistico, C. G., Bombardiere, L., Ammendolia, A., Rizza, P., De Fazio, P., & Capranica, L. (2012). Orthorexia nervosa: A frequent eating disordered behavior in athletes. *Eating and Weight Disorders-Studies on Anorexia Bulimia and Obesity, 17* (4), 226-233.

Shapiro, J. R., Woolson, S. L., Hamer, R. M., Kalarchian, M. A., Marcus, M. D., & Bulik, C. M. (2007). Evaluating binge eating disorder in children: Development of the children's binge eating disorder scale (C-BEDS). *International Journal of Eating Disorders, 40,* 82-89.

Stice, E., Telce, C. F., & Rizvi, S. L. (2000). Development and validation of the eating disorder diagnostic scale: a brief self-report measure of anorexia, bulimia and binge-eating disorder. *Psychological Assessment, 12,* 123-131.

Striegel-Moore, R. H. & Bulik, C. M. (2007). Risk factors for eating disorders. *American Psychologist, 62* (3), 181-198.

Tavakol, M. & Dennick, R. (2011). Making sense of cronbach's alpha. *International Journal of Medical Education, 2,* 53-55. doi: 10.5116/ijme.4dfb.8dfd

Ullman, J. B., & Bentler, P. M. (2009). Structural equation modeling. In M. Hardy & A. Bryman (Eds.) *The Handbook of Data Analysis, Paperback Edition* (pp. 431-458). Thousand Oaks, CA: Sage Publications.

Vandereycken, W. (2011). Media hype, diagnostic fad or genuine disorder? Professionals' opinions about night eating syndrome, orthorexia, muscle dysmorphia, and emetophobia. *Eating Disorders, 19,* 145-155.

Varga, M., Dukay-Szabo, S., Tury, F., & van Furth, E. (2014) When eating healthy is not healthy: orthorexia nervosa and its measurement with the ORTO-15 in Hungary. *BioMed Central Psychiatry*, 14, 1-11.

Wilson, J. L., Peebles, R., Hardy, K. K., & Litt, I. F. (2006). Surfing for thinness: a pilot study of pro-eating disorder web site usage in adolescents with eating disorders. Pediatrics, 118(6). doi: 10.1542/peds.2006-1133.

Wood, J. M., Tataryn, D. J., Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354-365.