Appetite 214 (2025) 108227

Contents lists available at ScienceDirect

Appetite

journal homepage: www.elsevier.com/locate/appet

Tracking the shift from health to harm: Development and validation of a short screening tool for orthorexia nervosa (STONE)

Elina Mitrofanova [a,b,*], Hilda Mulrooney [a,c], Elizabeth Pummell [a], Andrea Petróczi [a,d]

[a] School of Life Sciences, Pharmacy and Chemistry, Faculty of Health, Science, Social Care and Education, Kingston University, UK
[b] School of Human Sciences, Faculty of Education, Health, and Human Sciences, University of Greenwich, Old Royal Naval College, Park Row, Greenwich, London, SE10 9LS, UK
[c] School of Human Sciences, London Metropolitan University, London, N78DB, UK
[d] Faculty of Health and Sport Sciences, Széchenyi István University, Győr, Hungary

ARTICLE INFO

Keywords:
Orthorexia Nervosa
Screening
Assessment
Validity
Reliability
Psychometric scale

ABSTRACT

Orthorexia Nervosa (ON), a problematic fixation on healthy eating, has captured researchers' attention for over a decade. We aimed to develop a brief screening tool for ON that captures physical appearance as a motivating factor, behavioural aspects (rigid control over food selection, consumption and preparation), and nutritional aspects (avoidance of foods considered "impure"). Using a sequential, iterative design, 687 participants completed a self-reported survey across four studies: item identification and selection through exploratory factor analysis (n = 248), testing factorial construct validity with confirmatory factor analysis (n = 127), discriminant validity via known group differences (n = 241), and test-retest reliability of two subsequent administrations of the selected items (n = 71). The final unidimensional version of the short Screening Tool for Orthorexia Nervosa (STONE) comprises eight items. It demonstrated excellent known-group validity and ability to differentiate ON from other types of strict dietary control (e.g., health-based or religious restrictions). Consistent with the view of ON as behaviours aimed at rigid dietary control, avoidance of "impure" foods, and motivation to enhance physical appearance, STONE scores positively related to measures of eating pathology and appearance orientation, while only weakly correlating with obsessive-compulsive tendencies. Based on its psychometric properties, STONE is recommended as a first-level screening tool for ON in research contexts and epidemiology studies among adults. Due to its brevity, it can be easily combined with other scales to explore ON or related phenomena. Future studies should examine convergent validity and test it among adolescents and in different cultural contexts.

1. Introduction

Orthorexia Nervosa (ON) refers to an unhealthy fixation with healthy eating, where diets are based on perceived quality and not quantity of foods (Bratman, 1997), and has captured the attention of many. Parallel to the worldwide attempt to assess the prevalence of ON in diverse populations, several sets of diagnostic criteria were proposed (Cena et al., 2019). Recently, a group of researchers published an agreement document on the definition and diagnostic criteria (Donini et al., 2022). According to the Donini et al.'s document (2022), ON is characterised by eating behaviour with self-imposed rigid rules where food exclusion is not due to following religious practices, delusional ideas and not due to economic conditions but personal preference, and which requires excessive time for planning and consuming one's meals. These rules reflect an individual's desire for self-defined "pure" or "healthy" nutrition with foods being excluded often referred to as processed, containing genetically modified components, or treated with pesticides. Engaging in behaviours reflective of ON leads to emotional, cognitive, and social consequences that have a negative impact on individuals' day-to-day functioning. Exclusion of foods considered "unhealthy" may lead to nutritional deficiencies and low body weight. The document specifically states that there are also factors that have been linked to ON in research that are yet to be fully understood. These include age, gender, socio-cultural status, body image, appearance anxiety, alcohol and drug addiction, physical exercise, veganism and vegetarianism, and engagement with social media. To date, no formal diagnosis of ON exists in the

* Corresponding author. School of Human Sciences, Faculty of Education, Health, and Human Sciences, University of Greenwich, Old Royal Naval College, Park Row, Greenwich, London, SE10 9LS, UK.
E-mail address: e.mitrofanova@greenwich.ac.uk (E. Mitrofanova).

https://doi.org/10.1016/j.appet.2025.108227
Received 13 April 2025; Received in revised form 9 July 2025; Accepted 10 July 2025
Available online 12 July 2025

Diagnostic and Statistical Manual for Mental Disorders (DSM-5) (American Psychiatric Association, 2013). ON should not be confused with a healthy diet. According to the World Health Organisation (WHO, 2020, para. 1), a healthy diet is a "… diet that helps to protect against malnutrition in all its forms, as well as noncommunicable diseases (NCDs), including diabetes, heart disease, stroke and cancer". This definition recommends balancing the energy intake with energy expenditure, limiting consumption of free sugars, salt, saturated and trans-fats, and including a variety of foods, such as fruits, vegetables, legumes, nuts, and whole grains. While both, ON and healthy eating, may appear similar on the surface, ON involves rigid rules whereas healthy eating emphasises flexibility and diversity of dietary intake. Healthy eating allows for occasional indulgences without guilt, while ON is marked by experiences of distress when deviating from the self-imposed standards (Cena et al., 2019).

Despite being an increasingly fast-growing research field and several existing reviews on ON (e.g., Atchison & Zickgraf, 2022; Horovitz & Argyrides, 2023; López-Gil et al., 2023), some contentions regarding the factors contributing to its onset and maintenance exist (Ng et al., 2024). For example, one such factor is the role of body image and physical appearance. As mentioned earlier, the agreement document by Donini and colleagues (2022) states that the panel has not reached a consensus about the association between ON and "physical shape or body image disturbances" (p. 3703). However, insights from qualitative research exploring ON suggest that motivation to maintain or enhance physical appearance was identified as an important factor for following a self-defined healthy diet (Mitrofanova, Pummell, et al., 2021) as well as striving for control over food preparation and consumption (Cheshire et al., 2020). Interestingly, Barthels et al. (2015) proposed that intended weight loss may be present while Dunn and Bratman (2016) suggested that the absence of a desire to lose weight is essential for ON diagnosis. More recent empirical work supports the link between ON, preoccupation with overweight, muscularity distortion (Pauzé et al., 2021) and overvaluation of weight and shape (Messer et al., 2022). Divergence between proposed diagnostic criteria and empirical findings may be attributable to several factors. It could be that while some individuals may initiate behaviours reflective of ON due to the desire for "clean" or "healthy" nutrition, others may be driven by a desire to improve physical appearance, indicating a spectrum of motivations behind ON. It could also be that societal ideals of thinness and muscularity may influence individuals to pursue "healthy" diet as a socially acceptable way to achieve desired physical appearance (White et al., 2020). Discrepancies in the conceptualisations of ON lead to different assessment instruments and subsequent prevalence rates reported across the studies that are not comparable. One factor that has been overlooked in all existing measures are reasons for development and maintenance of dietary behaviour characteristic of ON.

Considering that recent studies suggest that cognitions leading to food restriction based on "purity" may be more complex than proposed diagnostic criteria (Cinquegrani & Brown, 2018; Fixsen et al., 2020; Greville-Harris et al., 2019; McGovern et al., 2020; Oberle et al., 2021; Valente et al., 2020) an assessment instrument should be able to capture not just behaviours but also reasons for those behaviours. For example, when considering various reasons for food choices and their connection to possible ON, Depa and colleagues (2019) found desire to control weight and affect regulation to display the highest associations with ON. Furthermore, adopting a vegan diet for the 'wrong reasons' (i.e., as an evidence of self-control) as opposed to environmental and ethical concerns has been described as one of the signs of the onset of ON (Opitz, Newman, & Sharpe, 2022). Considering reasons and motives for eating when screening for disordered eating patterns is not a new idea. In fact, several questionnaires that assess reasons for eating have been developed to screen for disordered eating patterns characterised by restrictive eating, bingeing, purging and attempt at emotional regulation (e.g. Jackson et al., 2003; Van Strien et al., 1986). Furthermore, the questionnaire used to assess the range and severity of eating disorder behaviours EDE-Q (Fairburn & Beglin, 1994) includes items measuring dietary restraint based on food rules and fear-driven restriction. These measures helped to extend our understanding of eating disorders recognised in the DSM-5 beyond the list of symptoms and behaviours focusing on why individuals engage in these behaviours (Gomez & Perez, 2022), identify at-risk individuals (Bryant et al., 2021) and have helped clinicians to design tailored treatment plans (Baer et al., 2005).

Since ON's first conceptualisation, eleven questionnaires were developed using different definitions (Cena et al., 2019; Meule et al., 2020; Opitz et al., 2020; Valente et al., 2019). A recent systematic review of ON reports that all existing questionnaires show strengths (some more than others), but they also have limitations (Ng et al., 2024). The first questionnaire (Bratman Orthorexia Scale) was developed in 2000 and included six statements with yes/no responses. No validating procedures were followed in the development of this scale and no cut-off point exists for it. This fact, however, did not discourage its use in academic research. Bundros et al. (2006), for example, found the scores to be associated with Eating Attitude Test (EAT-26), Body Dysmorphic Disorder Questionnaire (BDDQ) and the Obsessive Compulsive Inventory (OCI-R). The scale has been criticised for a lack of appropriate psychometric procedures in its development and for a lack of reporting on its validity in the studies using this scale (Missbach et la., 2016). However, research comparing four self-report measures of ON found BOT to demonstrate convergent validity with the Eating Habits Questionnaire and Dusseldorf Orthorexia Scale and recommended its use over the more popular scale ORTO-15.

The ORTO-15 was developed by an Italian research team and was based on the items from the BOT (Donini et al., 2005). The ORTO-15 includes six BOT and nine additional items intended to reflect obsessive-compulsive traits. With possible responses on a 4-point Likert-type scale. The cut-off score was set at 40 with scores below indicating the presence of ON. However, subsequent research highlighted this cut-off score as problematic and suggested a cut-off score to be set at 35 (Ramacciotti et al., 2011). Validation that followed the construction of the tool suggested three dimensions of ON: cognitive-rational, clinical, and emotional. However, the study that attempted to confirm the factor structure of the 15, 11 and 9 items versions of ORTO-15 concluded that none of the three versions represent an acceptable model (Moller et al., 2019). Researchers suggested that the only version of ORTO-15 with a stable factor structure was a 7-item single factor. Furthermore, ORTO-15 has been criticised for overestimating the prevalence of ON (Reynolds, 2018) due to classifying as ON dieting behaviour which may not be reflective of the ON-specific pathology (Dunn et al., 2017). Validity, reliability and internal consistency of the tool have also been questioned (e.g. Barrada & Meule, 2024; Meule et al., 2020).

The Eating Habits Questionnaire (EHQ) was developed in the USA by Gleaves and his research team (2013) and was not based on diagnostic criteria but on the analysis of Bratman and Knight's case studies. The questionnaire contains 21 items with responses on a 4-point Likert-type scale. The authors suggested a three-factor structure: problems associated with healthy eating, knowledge of healthy eating, and feeling positively about healthy eating. However, later investigations into the structure of this scale present a different factor structure. For example, a four-factor structure has been proposed: healthy eating cognitions, dietary restriction, diet superiority, and social impairment (Mohamed Halim et al., 2020). Furthermore, another study has identified three items that loaded on the EHQ-Behaviours subscale instead of the originally intended EHQ-Problems subscale suggesting a need for further analysis of the internal structure of this questionnaire (Oberle et al., 2017).

The Dusseldorf Orthorexia Scale (DOS) was developed by Barthels and colleagues in 2015 in Germany and was also based on the case studies of Bratman and Knight. The questionnaire consists of ten items with possible responses on a four-point Likert-type scale. The suggested cut-off score was 30 with scores between 25 and 29 indicating being at

risk for ON. The scale was developed and validated in German. Studies investigating the internal structure of DOS present inconsistent results. While confirmatory factor analysis of the English version revealed a poorly fitted one-factor model (Chard et al., 21019), Meule et al. (2020) confirmed the unidimensional structure. The Chinese version of the scale demonstrated that a one-factor model did not fit the data very well and a three-factor structure was proposed instead (He et al., 2019). The scale has also been criticised for its inability to differentiate between patients with anorexia nervosa and those displaying ON (Barthels et al., 2017).

Teruel Orthorexia Scale (TOS) was developed and validated in Spain and is a 17-item questionnaire that assumes a bi-dimensional structure (i.e., healthy orthorexia [HeOn] and orthorexia nervosa [OrNe]) (Barrada & Roncero, 2018). There is no cut-off score. There is some support for the scale's factor structure. For example, Barthels et al. (2019) confirmed that the OrNe dimension was positively associated with negative affect whereas HeOn was positively associated with positive affect. However, the factor structure may vary across cultures. In a Greek adaptation study, researchers found that a three-factor model (comprising Healthy Orthorexia, Emotional Orthorexia, and Cognitive-Social Orthorexia) provided a better fit than the original two-factor model (Argyrides et al., 2024). There were also item-level concerns with various studies suggesting removal of some items. For example, items 9 and 13 were identified as problematic due to cross loadings in an English validation study (Chace and Kluck, 2022) while the adaptation study to French dropped items 9 and 15 (Lasson et al., 2023).

Bauer et al. (2019) developed the Barcelona Orthorexia Scale (BOS) in Spain, which includes 64 items with no cut-off score. The strength of this measure is that it was developed based on proposed diagnostic criteria available at the time of development. Development of this scale was based on the Delphi study methodology where Spanish and English-speaking experts in the field of eating disorders were asked for their opinions on the representativeness of each item. To date, BOS lacks psychometric validation, and the authors identified several limitations in the methodology of its development. Among these is the fact that not all experts that participated in the process ever published on ON and level of expertise differed; with some individuals being involved with eating disorders but not having any experience with ON.

Orthorexia Nervosa Inventory (ONI) was developed in the USA and includes 24 items, which need to be responded to on a 4-point Likert-type scale (Oberle et al., 2021). This questionnaire is based on the items from EHQ and DOS. The authors suggested a three-dimensional structure: ON behaviours, emotional stress, and physical and social impairment. This structure was confirmed in later studies (e.g. Fodor et al., 2025). It is the only available scale that includes items reflecting physical impairment from ON. This is a promising measure. However, it does not have a cut-off score.

Some of the questionnaires discussed demonstrate better psychometric properties than others, but none without limitations, which necessitates development of alternative tools. To summarise, DOS has been criticised for its inability to differentiate between ON and other eating disorders such as anorexia nervosa and for its limited incorporation of functional impairment. EHQ and TOS present uncertainty in terms of factor structure and cultural relevance of some items. Although ONI represents a more recent and promising measure (Oberle et al., 2021), it was not available at the time when STONE was conceived. In comparison with STONE, ONI lacks a validated cut-off score and remains relatively long, limiting its practicality as a rapid screening instrument in broad, multi-measure assessments. Moreover, existing tools, ONI included, tend to underemphasise motivational factors such as appearance-driven eating behaviours, which emerging evidence suggests play a key role in ON (Pauzé et al., 2021). Furthermore, to overcome the inability of some tools to differentiate between individuals with ON and other possible restrictive practices, the new measure should be able to demonstrate discriminative capacity across comparison groups. Thus, a new assessment tool to further our understanding of ON is needed (Ng et al., 2024).

1.1. Aims

The aim was to develop a short screening tool for ON. Short measures are associated with higher completion rates, reduced participant fatigue, and are often perceived as less intrusive (Rolstad et al., 2011). Furthermore, in large epidemiological research, inclusion of short measures allows for assessment of multiple domains without overloading participants. Building on previous studies (Mitrofanova, Pummell, et al., 2021; Mitrofanova, Mulrooney, & Petroczi, 2021), this study aimed to develop a new assessment tool that includes physical appearance as a motivating factor, behavioural aspects (rigid control over food selection, consumption and preparation), and nutritional aspects (avoidance of particular foods considered "impure").

1.2. Overview of the scale development process

An iterative sequential design was employed, with four independent studies (Fig. 1), conducted between 2018 and 2020. Study 1 aimed to generate the item pool and initial testing for psychometric properties. Study 2 examined the factor structure of the new scale. Study 3 was conducted to establish evidence for validity and re-examine the factor structure of the scale with the independent sample. Study 4 sought to establish test-retest reliability.

Ethical approval was granted for all studies by the Faculty Research Ethics Committee at Kingston University London. Participants in all studies were fully informed about the aims, procedures involved, their rights to withdraw and data treatment via the information sheet. Contact details of organisations offering psychological support were provided on the debriefing sheet.

2. Study One

The objective of this study was to generate and test the items assessing ON-related cognitions, behaviours, and beliefs.

2.1. Method

2.1.1. Measures

The list of items was developed based on qualitative interviews and nutritional assessments of individuals with suspected ON (Mitrofanova, Pummell, et al., 2021; Mitrofanova, Mulrooney, & Petroczi, 2021). Previous studies carried out by this research team explored individuals' 24-hour macro- and micro-nutrient intakes to identify a potential pattern of dietary restrictions reflective of ON and provided a qualitative understanding of the reasons, beliefs, and motives for following a diet indicative of ON. Themes generated from the qualitative interviews were systematically translated into item content. For example, items reflecting the impact of individuals' dietary restrictions on their social lives, interpersonal communication and their relationships with friends and family were derived from theme "Social". Items reflecting perceived dietary purity, avoidance of foods considered unhealthy, behavioural control over food preparation and consumption and striving to enhance physical appearance emerged from theme "Rules/Control" and from the 24-hour recall assessments; items reflecting the "Journey" theme captured the evolving role of food in one's life, relating to identity and self-worth. Putting findings from both qualitative and nutritional studies, common features were very particular beliefs about foods' harmful and health-enhancing properties, motivation to maintain or achieve physical appearance, and a desire for control over food preparation and consumption which impacted other aspects of individuals' lives. In all, 80 statements that reflected cognitions, behaviours, and beliefs of individuals with suspected ON were grouped under eight domains. These domains were 'preoccupation with healthy eating',

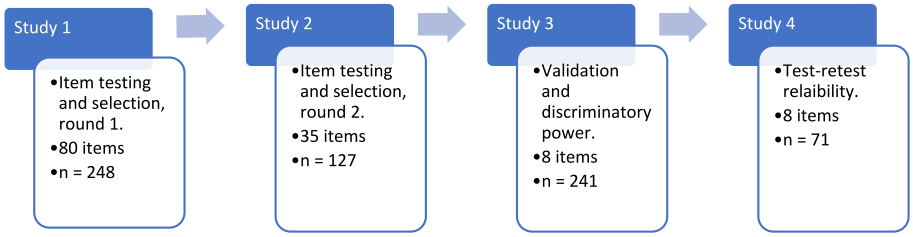

Fig. 1. Development process of STONE.

'compensatory behaviours', 'rigidity/control', 'physical activity', 'identity/self-worth', 'purity of the diet', 'social life', and 'physical appearance'. Participants were also asked to provide demographic and anthropometric information, indicate self-perceived weight status, report existing dietary restrictions, past diagnoses of an eating disorder, obsessive compulsive disorder, any medical condition that may impact their eating behaviours, and to complete the Bratman Orthorexia Test (BOT). Responses were not forced, allowing participants to skip questions they felt unsure about or uncomfortable with. The questionnaire with the information sheet can be found in the supplementary materials.

2.1.2. Participants and procedure

The sample in Study 1 consisted of 248 respondents. Participants' age ranged from 18 to 68 years ($M = 26$, $SD = 9.66$); 129 individuals identified as men and 119 identified as women. Of those participants that responded to the question about ethnicity, the most represented ethnic group was White English (n = 60, 24.1 %) and any other White background (n = 60, 24.1 %) followed by White British (n = 13, 5.2 %), Black African (n = 13, 5.2 %), Indian (n = 12, 4.8 %), Pakistani (n = 11, 4.4 %), Bangladeshi (n = 11, 4.4 %), any other Asian background (n = 10, 4 %), Chinese (n = 8, 3.2 %), Arab (n = 8, 3.2 %), Black Caribbean (n = 7, 2.8 %), and Mixed White and Black and Mixed White and Asian (both n = 6, 2.4 %). Most participants (n = 168, 67.5 %) indicated the UK as their country of residence. The majority identified the English language as at least one of the languages spoken at home (n = 146, 59.6 %). BMI of the sample ranged from 16.1 to 52.03 kg/m^2 (M = 23.20, SD = 4.23). Most participants reported not having any restrictions in their daily diets (n = 145, 58.5 %). Of those that reported restrictions, vegetarian and vegan diets were cited most often. Other restrictions included low carbohydrate, fat, sugar, dairy products, and "fast food". Eighteen individuals (7.3 %) reported an eating disorder diagnosis, while six participants (2.4 %) reported having an OCD diagnosis. Most individuals did not report any medical or psychological conditions impacting their diet (n = 210, 84.7 %). Adherence to religious diet was reported by 36 (14.5 %) participants. Interestingly, most participants in this sample answered affirmatively to at least five of the BOT items (n = 153, 61.7 %) indicating a possible presence of ON tendencies among the sample. No exclusion criteria were applied based on socio-demographic characteristics, however, all participants had to be at least 18 years old. Recruitment took place in-person and online among undergraduate university students in England. Students were approached via in-classroom announcements and asked to complete a hard copy of the survey. Online participants were contacted via email and social media platforms (Facebook and Instagram) and asked to complete an identical survey hosted on a closed platform (SurveyMonkey). Participants were informed that it would take approximately 20 min to complete the survey. Items were presented in a fixed order starting with demographic and anthropometric questions, followed by the scale under development (80 items) and the BOT. The final part of the survey asked if they were ever diagnosed with an eating disorder, obsessive-compulsive disorder, or if there was a known medical or psychological condition that impacted upon their food choices and if they were following any religious diets. Participants were encouraged to share the survey link with their acquaintances.

The sample size for the exploratory factor analysis (EFA) was determined based on published recommendations for scale development. A total of 248 participants completed the 80-item version of the screening tool, yielding a subject-to-variable ratio of approximately 3:1. While this is below the commonly cited 5–10 participants per variable rule of thumb (Gorsuch, 1983), several simulation studies suggest that lower ratios can still yield reliable factor solutions under certain conditions (e.g., when communalities are greater than .6) (MacCallum et al., 1999). In this study, due to a lack of meaningful factors resulting from the EFA, sample size for a hierarchical agglomerative cluster analysis was considered sufficient. While there are no strict sample size requirements for hierarchical cluster analysis, Formann (1984) recommends the minimum sample size to include no less than 2^k cases (k = number of variables). In this case $2^k = 2 \times 80 = 160$ cases. Finally, the sample size of 248 was considered sufficient for the EFA with 41 items remaining from the cluster analysis as it presented a 6:1 subject-to-variable ratio.

2.2. Results

Kaiser's measure of sampling adequacy for the scale indicated the data were appropriate for EFA (Kaiser, 1974). The value was .88 which Kaiser described as "meritorious". This iteration of the EFA produced 20 factors with eigenvalues >1.00. The analysis failed to reveal any meaningful factors consistent with the operationalisation of the ON. A hierarchical agglomerative cluster analysis was, therefore, performed for data analysis and revealed seven clusters. One cluster was omitted as it contained only one item. Then, each cluster was assessed by conducting an item-total correlation and identifying the items with the highest coefficients. Items with corrected item-total correlation values > .5 were retained. The resulting scale consisted of 41 items. Table A.1 in Appendix A presents the items with their respective cluster membership and item-total correlation values.

The next step involved subjecting the retained scale to the EFA. EFA of the 41-item with varimax rotation produced nine factors. However, two factors were removed as only two items loaded on factor 8 and one item on factor 9. Decision regarding the number of factors to retain was based on examining Cattell's (1966) scree-test and eigenvalues (>1). Furthermore, items that had large cross loadings, defined as secondary loadings ≥.30 on more than one factor, and items with loadings <.3 were removed (Tabachnick & Fidell, 2019). Although higher cut-offs (e.g., .40) are often recommended for more conservative analyses, a .3 threshold allows for the retention of items that may contribute meaningfully to emerging constructs, especially in the context of new instrument development. We followed these criteria as they allowed for methodological rigour without prematurely discarding potentially meaningful items at the early stages of scale development. The total number of items retained after the EFA was 33. At this point, the combinations of the items necessitated a review of descriptive labels of the factors. Table 1 presents the items with their respective factor loadings and Cronbach's alpha coefficients for each component. According to interpretations of alpha coefficients suggested by George and Mallory (2003) the values fall from acceptable (>.7) to good (>.8). Cronbach's alpha for the scale was .938, which is considered excellent.

Table 1
Factor structure and Cronbach's alpha coefficients of the seven factors.

	Items	F1 – Physical (rational) outcome - Health	F2 – Emotional outcome – identification with diet	F3 – Barriers to overcome to stick to the diet	F4 – Weight management	F5 – Food purity	F6 – Subject interest (nutrition)	F7 - Control
1	My diet has more health benefits than other diets.	.694						
2	Healthy eating is a large part of who I am.	.685						
3	I eat only healthy food.	.647						
4	My body is pure because of my healthy diet.	.561			.360			
5	I have a strict exercise routine to complement my diet.	.521					.334	
6	I feel a sense of achievement when I stick to my diet.		.722					
7	Being able to stick to my diet has a positive impact on my mood.		.638				.342	
8	I eat healthy because I want to improve the way I look.		.625		.401			
9	I feel better about myself when I manage to avoid slipping off my healthy diet.		.621		.301			
10	My chosen diet has a direct impact on my appearance.		.595		.380			
11	It's difficult to find a restaurant that serves the foods that I eat.			.690				
12	My family has to make me a separate meal/dish when eating together (e.g. Christmas, Easter).			.687				
13	Other people have mentioned that my diet is too restrictive.			.672				
14	My diet has many rules.			.594			.322	
15	I avoid food that I haven't prepared myself.			.484				
16	I'm very specific about my food choices.	.439		.480				
17	I go out less frequently since I began eating healthy.			.387				
18	I follow my diet in order not to gain weight.				.796			
19	My diet is designed to keep me at a specific weight.				.700			
20	The main motivation behind my food choices is weight management.		.351		.644			
21	If I eat something outside of my diet, I will try to make up for it and eat less or exercise more the next day.	.341			.614			
22	I avoid foods that were treated with pesticides.					.798		
23	I avoid genetically modified foods.					.773		
24	I avoid processed foods.	.321				.659		
25	It's important for me to know where the food I buy at the supermarket/ market comes from.					.604		
26	Most of my social interactions involve a discussion about my eating habits.						.682	
27	Nutrition is a hobby of mine.	.338	.424				.589	
28	I make sure that my diet is better than most people's diet.	.418					.540	
29	I actively seek the latest trends/ information/news in nutrition.						.486	.310
30	I spend a lot of time researching nutritional composition of foods.	.353		.317			.448	
31	I plan when to allow myself a treat outside of my diet.							.551
32	I measure every portion.							.492
33	I carefully monitor the nutritional composition of what I eat.	.376						.481
	Reliability (Cronbach's alpha)	.821	.844	.811	.805	.783	.791	.732

Despite the satisfactory values of internal consistency demonstrated by the factors and the scale, some factors still contained items reflecting several theoretical dimensions, which warranted further investigation.

3. Study Two

This study aimed to re-examine the structure of the scale using a reduced number of items with an independent sample. In addition, we

added new items to 'control' and 'appearance' subscales to better reflect the respective constructs.

3.1. Method

3.1.1. Measures

Two additional items were added to the scale. One item was added to "control" subscale "All my meals are planned" and one to "appearance" subscale "My diet is good for my skin". Both components were highlighted as reasons for adherence to the diet of choice in the qualitative study (Mitrofanova, Pummell, et al., 2021) and were important contributors to the conceptualisation of ON when developing the items for Study One. All other questions were identical to the questionnaire used in Study One.

3.1.2. Participants and procedure

The sample consisted of 127 participants (91 females, 36 males) aged from 18 to 68 years ($M = 31.05$, $SD = 11.06$). "Any other ethnic group" was the most represented category in this sample (n = 40, 31.5 %) followed by White English (n = 29, 22.8 %), Black British (n = 14, 11 %), Asian Indian (n = 8, 6.3 %), and Asian Pakistani (n = 6, 4.7 %). The UK was the country of residence for most participants (n = 99, 78 %). BMI of the sample ranged from 15.79 to 42.72 kg/m^2 (M = 24.21, SD = 4.97). Most individuals did not have any restrictions in their daily diets (n = 82, 64.6 %). Those who reported restrictions in their diet avoided sugar, dairy, carbohydrates, "junk food", and restricted their calorie intakes. Two individuals were following a vegetarian diet, two individuals were vegan, and two participants excluded gluten. Seven individuals reported an eating disorder diagnosis, two participants were diagnosed with anorexia nervosa, five participants had a diagnosis of bulimia nervosa in the past. Five participants reported an OCD diagnosis. Psychological and medical conditions influencing dietary choices were reported by nineteen participants. The most common conditions included depression and anxiety. Twelve people were following a religious diet (halal or kosher). Six individuals (4.7 %) answered affirmatively to at least five BOT statements. The sample was drawn from university students in England. Data collection followed identical steps to Study 1.

Using a new sample of 127 participants, resulted in a subject-to-variable ratio of approximately 3.6:1. Although this ratio is below the commonly recommended threshold of 5–10 participants per item (Gorsuch, 1983), MacCallum et al. (1999) suggest smaller samples may be adequate when certain conditions are met specifically, when communalities are moderate to high (≥.50) and factors are well-defined by at least three items with strong loadings.

3.2. Results

The value of Kaiser's measure of sampling adequacy for the scale (KMO = .86) indicated the data were appropriate for EFA (Kaiser, 1974). EFA with oblimin rotation was conducted to explore the internal structure of the scale. To determine the number of dimensions to retain, parallel analysis (PA) was used (Horn, 1965) and was conducted using R software (R Core Team, 2017). In addition, visual inspection of the scree plot, factor loadings, eigenvalues, and face validity of the items were taken into consideration. Only items with factor loadings above .30 were retained and considered for inclusion. The results from the PA, which can be seen in Appendix B Figure B.1, suggested the retention of three factors in the EFA since three eigenvalues from the sample were greater than those from the randomly generated datasets. These factors represented dimensions of "appearance" (7 items), "purity" (8 items), and "control" (6 items). After assessing items for face validity, five items were removed as they failed to demonstrate conceptual fit with the factors. For example, items "I feel better about myself when I manage to avoid slipping off my diet" and "I feel a sense of achievement when I stick to my diet" describe the affective element of adhering to a diet rather than enhancement of one's physical appearance and were, therefore, removed from the "appearance" dimension. Next, item "My diet is good for my skin" was included in the "appearance" component and item "It's important for me to know where the food I buy at the supermarket/market comes from" was assigned to the "control". Internal consistency reliability, assessed by Cronbach's alpha coefficients, were .84 for the "appearance", .82 for the "purity", .61 for the "control" subscales, and .84 for the whole scale. The resulting scale included 16 items. Table 2 presents retained items with their factor loadings.

To verify the factor structure, the three-factor scale was subjected to the Confirmatory Factor Analysis (CFA). Absolute and relative fit indices were generated and examined. Hooper and colleagues (2008) advocate reporting the following indices: Chi-squared ($\chi 2$) its degrees of freedom and p value, root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR), the comparative fit index (CFI) and one parsimony fit index such as parsimonious normed fit index (PNFI). Obtained values were compared with the acceptable thresholds of fit by indices recommended by Hooper et al. (2008), which were: $\chi 2$ with a non-significant p value ($p > .05$), RMSEA value smaller than .07, SRMR value smaller than .08, CFI value greater or equal to .95, PNFI within the .50 region while other fit indices achieve values over .90. However, the goodness of fit statistics revealed that the model demonstrated a poor fit to the data $\chi^2/df = 223.02/101 = 2.208$, $p < .00$; RMSEA .098; SRMR .113; CFI .817; PNFI .603, but normed chi-square statistics ($\chi^2/df = 2.208$) was within the acceptable range of <2.5. This necessitated further adjustments and assessment of the internal structure of the scale with the independent sample.

4. Study Three

At this stage, the resulting scale included three factors corresponding with conceptualisation of ON as a condition that presents as preoccupation with "clean" eating (purity), desire to enhance one's appearance via adherence to the diet of choice (appearance), preoccupation with the topic of nutrition, and rigid behaviours of food preparation and consumption (control). However, the three-factor model demonstrated a poor fit to the data in Study Two, which necessitated further

Table 2
Factor structure of the 16-item scale.

	Items	Appearance	Purity	Control
1	I follow my diet in order not to gain weight (A1).	.815		
2	I eat healthy because I want to improve the way I look (A2).	.812		
3	The main motivation behind my food choices is weight management (A3).	.778		
4	My chosen diet has a positive impact on my appearance (A4).	.675		
5	My diet is designed to keep me at a specific weight (A5).	.646		
6	I avoid processed foods (P1).		.777	
7	It's important for me to know where the food I buy at the supermarket/market comes from (C5).		.740	
8	I avoid genetically modified foods (P2).		.718	
9	I eat only healthy food (P3).		.698	
10	My diet has more health benefits than other diets (P4).		.685	
11	I avoid foods that were treated with pesticides (P5).		.657	
12	My diet is good for my skin (A6).		.555	
13	Other people have mentioned that my diet is too restrictive (C1).			.709
14	I avoid food that I haven't prepared myself (C2).			.364
15	I measure every portion (C3).			.343
16	I spend a lot of time researching nutritional composition of foods (C4).			.331

Note: "A" = Appearance, "P" = Purity, "C" = Control.

investigation into the scale's structure.

Past literature on ON consistently reflects an association with obsessive-compulsive tendencies and disordered eating habits (McComb & Mills, 2019). In addition, qualitative interviews conducted by this research team revealed the desire to improve one's appearance as one of the primary motivations for seeking to eat "healthily" (Mitrofanova, Pummell, et al., 2021). Therefore, to establish convergent validity of STONE, we hypothesised that the scale would correlate with the measures of obsessive-compulsive tendencies, eating pathology, and appearance orientation. Further, to establish discriminatory power of STONE, we hypothesised that individuals exhibiting orthorexic tendencies would score significantly higher than participants restricting their dietary intakes for other reasons and those without dietary restrictions. For example, individuals with medical conditions such as irritable bowel disease, food allergies, coeliac disease, diabetes often have dietary restrictions to manage symptom severity (Evert et al., 2019). Many avoid certain foods and engage in fasting as part of practicing religious beliefs (Düzçeker et al., 2021). Furthermore, some may limit their food intakes for professional reasons. For example, athletes may manipulate food intake to enhance sport performance (Stoel et al., 2021). While these examples involve abstaining from food, none of these practices are inherently pathological. STONE should, therefore, be able to differentiate between those with a diagnosed eating disorder, those without any dietary restrictions, and those with high scores on BOT if it is measuring ON behaviours, cognitions and motivations.

For practical application, the scale was validated against the scores on the BOT due to the absence of official diagnostic criteria. Assessing accuracy of a screening tool involves examining its ability to distinguish a presence of condition from its absence compared to the 'gold standard'. Receiver Operating Characteristic (ROC) analysis reports performance of a screening tool in terms of the sensitivity (probability of a positive result in people with a condition) and specificity (probability of a negative result in people without a condition) using a cut-off score to define "positive" and "negative" test results (McNeil et al., 1975). The ROC curve is a graph that provides combinations of sensitivity and specificity and runs from point 0 to point 1. A screening tool with no discriminating ability at all has a ROC graph that follows the diagonal line from point 0 to point 1 covering 50 % of the area under the curve (AUC) with accuracy approximating random guessing.

Therefore, the purpose of Study Three was threefold: (1) to confirm the factor structure for the scale in an independent sample, (2) to assess convergent and discriminant validity, (3) and to identify cut-off points for practical application.

4.1. Method

4.1.1. Measures

Participants provided demographic information, self-reported height and weight, the 16-item scale developed in the previous study, and various self-report measures to assess the validity of the scale (described below). Study materials were hosted on SurveyMonkey. Items were presented in random order to each participant.

4.1.1.1. The Eating Attitude Test (EAT-26; Garner et al., 1982). This questionnaire was used to identify pathological eating behaviours and attitudes. The test is not a diagnostic measure for eating disorders. The authors of the scale suggest that scoring above the cutoff point indicates a presence of a possible eating pathology but do not claim the scale's ability to establish an exact diagnosis. Responses are scored on a 6-point Likert-type scale ranging from "Never" to "Always". A total score was used for analysis with a score of 20 and higher indicating a tendency towards disordered eating. Internal consistency reliability (Cronbach's alpha) of this scale in this sample was .89. This measure was included in this study to demonstrate that STONE measures a distinct eating behaviour reflective of ON, not pathological eating behaviours and attitudes that represent possible presence of eating disorders like anorexia nervosa and bulimia nervosa. For example, if the scores of STONE and EAT-26 had a moderate correlation, it would suggest that STONE captures restrictive eating while still measuring a distinct condition.

4.1.1.2. The Obsessive Compulsive Inventory Revised (OCI-R; Foa et al., 2002). The questionnaire is used to assess symptoms of obsessive-compulsive disorder. This scale is an 18-item measure scored on a 5-point Likert scale with a score of 21 and higher suggesting the presence of obsessive-compulsive tendencies. In this sample internal consistency reliability (the Cronbach's alpha) was .92. Previous studies linked ON to obsessive compulsive tendencies (e.g., Koven & Abry, 2015). OCI-R was included to assess this link. We expected positive but low associations with STONE suggesting that ON behaviours are more reflective of eating pathology, rather than a condition of obsessive-compulsive nature.

4.1.1.3. Multidimensional Body-Self Relations Questionnaire Appearance Orientation subscale (MBSRQ-AO; Cash, 2015). In this study only the AO subscale of the MBSRQ was used. The items of this subscale measure the extent of preoccupation and investment into one's physical appearance and grooming behaviours. There are 12 items with responses ranging from "Definitely Disagree = 1" to "Definitely Agree = 5". Scores are calculated by estimating an average. There is no cut-off score but author-provided population average for males is 3.60 and 3.91 for females. Higher scores indicate greater investment in one's appearance. In this sample internal consistency reliability (Cronbach's alpha) was .90. This measure was included to assess the extent to which physical appearance is important in ON. Moderate correlation would suggest that ON may be driven by motivations to maintain and enhance one's appearance.

4.1.2. Participants and procedure

The sample included 241 individuals. All participants were at least 18 years old. The study used purposive sampling technique. Recruitment took place using Prolific.co. Prolific is an online platform for recruitment of participants with the aim to explicitly cater to researchers. Participants were contacted via Prolific using their IDs on the web site. Prolific allows researchers to not only post a "call" for participation in a study but also to limit the visibility of this "call" to particular individuals. Once the "call" is visible, it is entirely up to Prolific participants to take part (or not) in the study. Participants were selected based on an existing large dataset collected for previous (unrelated) research projects. This two-stage approach to recruitment was used to mitigate against the risk of falsely claiming a behaviour or condition for eligibility to participate in the study. Participants received a monetary reward for their participation (average reward rate = 10.55£ per hour). To assess convergent validity five groups of participants took part in this study: "BOT" – individuals that scored at least 5 points on BOT scale; "Medical" – individuals that reported having a medical condition (e.g. diabetes, irritable bowel syndrome) that impacts their daily diet; "Religious diet" – those following a religious diet (e.g. Kosher, Halal, Eastern Orthodox); "ED" – individuals self-identifying as having been diagnosed with an eating disorder; "Professional reasons" – individuals following a specific diet to maintain their weight for professional reasons (athletes, models); "Control" – a group of healthy adults that do not self-identify with any of the above criteria. Table 3 presents participants' groups and demographic information.

Sample size for the CFA was considered sufficient as it was greater than the rule of thumb of 10 cases per variable (Nunnally, 1967). This sample size was also sufficient for validity testing as a priori power analysis conducted using G*Power version 3.1.9.7 (Faul et al., 2007) indicated the required sample size to achieve 80 % power for detecting a medium effect, at a significance criterion of $\alpha = .05$, was $n = 82$ for

Table 3
Groups and demographic characteristics of participants.

Characteristics	n	%
Group		
BOT	44	18.3
Medical	42	17.4
Religious diet	29	12
ED	41	17
Professional/job-related reasons	45	18.7
Control	40	16.6
Ethnicity		
White English/Welsh/Scottish/Northern Irish/British	190	79
Indian	13	5.4
Pakistani	11	4.6
Black African/British	9	3.8
White and Asian	4	2.1
White and Black African/Caribbean	3	1.2
Bangladeshi	3	1.2
White and Black African	1	.4
Arab	1	.4
Other	6	2.5
Marital status		
Never having been married or in civil partnership	140	58.1
Married/civil partnership	91	37.8
Divorced/civil partnership dissolved	8	3.3
Separated from spouse or partner	2	.8
Children under 16 years old in household.		
No children	147	61
One child	46	19.1
Two children	29	12
Three children	16	6.6
Four children	2	.8
Five children	1	.4

Note: BOT – individuals that scored at least 5 points on BOT scale; Medical – individuals that reported having a medical condition (e.g. diabetes, irritable bowel syndrome) that impacts on their daily diet; Religious diet – those following a religious diet (e.g. Kosher, Halal, Eastern Orthodox); ED – individuals self-identifying as having been diagnosed with an eating disorder; Prof. reasons – individuals that follow a specific diet to maintain their weight for professional reasons (athletes, models); Control – a group of healthy adults that do not self-identify with any of the above criteria.

Pearson's correlation. Using the same parameters in G Power sample size for a one-way ANOVA was calculated to be n = 216, which is below the sample size in this study.

4.2. Results

4.2.1. CFA

Kaiser's measure of sampling adequacy for the scale (.89) indicated the data were appropriate for analysis (Kaiser, 1974). Initial fit indices were poor, χ^2/df (555.84/101) = 5.503, $p < .00$; RMSEA .137 SRMR .125; CFI .744; PNFI .595. Modification indices were examined and error terms with values > 20 on the same factor were allowed to correlate. However, the model still did not demonstrate a satisfactory fit χ^2/df (405.425/96) = 4.223, $p < .00$; RMSEA .116 SRMR .110; CFI .826; PNFI .629.

After observing the model fit of the 16-item version, the decision was made to test if single-factor shorter version of the scale could be a viable solution. Several items (A3, A4, A6) showed high modification indices with more than one other item suggesting an additional covariances with items measuring purity of the diet (A4 and P4) and items measuring the control dimension (A6 and C5). Purity dimension also contained items with high modification indices suggesting an additional covariance with items on the control dimension (P1 and C2, P2 and C5) and with items on the same variable (P2 and P5). The CFA was conducted on the scale that consisted of items that did not demonstrate multiple modification indices with values > 20. This version represents an alternative to the 16-item version consisting of three subscales and includes only one hypothetical dimension underlying all items – ON; and consisted of items A1, A2, A5, P3, P5, C1, C3, and C4. This decision integrated both statistical evidence and theoretical considerations. Although previous research has conceptualized ON through multidimensional frameworks encompassing behavioral, cognitive, and emotional domains, this analysis indicates these theoretical dimensions may demonstrate substantial empirical overlap. Rather than functioning as distinct constructs, these components appear to coalesce around a single underlying factor, suggesting that pursuit of dietary purity manifests as an integrated psychological phenomenon. This unidimensional structure may reflect that cognitive preoccupation with food purity, behavioral rigidity in eating patterns, and motivation to enhance/maintain one's physical appearance operate as mutually reinforcing elements of a cohesive phenomenon. Initially, model fit indices still did not meet the recommended values, χ^2/df(92.50/20) = 4.63, $p < .00$; RMSEA .123 SRMR .064; CFI .872; PNFI .603. Modification indices were examined for values above 10 and associated error terms were allowed to correlate (e1↔e2, MI = 19.159; e1↔e8, MI = 15.585; e1↔e3, MI = 15.141; e2↔e7, MI = 13.808).

The model fit indices improved after allowing for the correlation of the error terms, χ^2(92.50/20) = 4.63, $p < .00$; RMSEA .123 SRMR .064; CFI .872; PNFI .603. Fig. 2 below presents the model and each item's contribution to measuring orthorexic tendencies.

The resulting Screening Tool for Orthorexia Nervosa (STONE) (8 items) demonstrated a very good internal reliability consistency expressed as Cronbach's alpha coefficient of .82.

4.2.2. Validity and discriminatory power of STONE

As predicted, STONE demonstrated significant correlations with the measures administered to assess the convergent validity. There was a moderate positive correlation between scores on the STONE and the EAT-26 (r = .454, p < .001), indicating a medium effect size. Approximately 21 % of the variance in EAT-26 scores was shared with STONE scores (r^2 = .21). There was a small but statistically significant positive correlation between STONE and OCI-R scores (r = .179, p = .01), with approximately 3 % of the variance shared between the two measures (r^2 = .03), suggesting a weak association with obsessive–compulsive symptoms. A moderate positive correlation was also found between STONE and MBSRQ-AO scores (r = .344, p = .01), indicating a medium effect size, with 12 % of the variance in appearance orientation scores explained by STONE scores (r^2 = .12). The obtained correlations with indices of disordered eating (EAT-26) and obsessive-compulsive tendencies (OCI-R) were consistent with the previous studies using other ON measurement tools (e.g. McComb & Mills, 2019). Correlation with appearance orientation (MBSRQ-AO) is consistent with the view of ON proposed by the qualitative findings (Mitrofanova, Pummell, et al.,

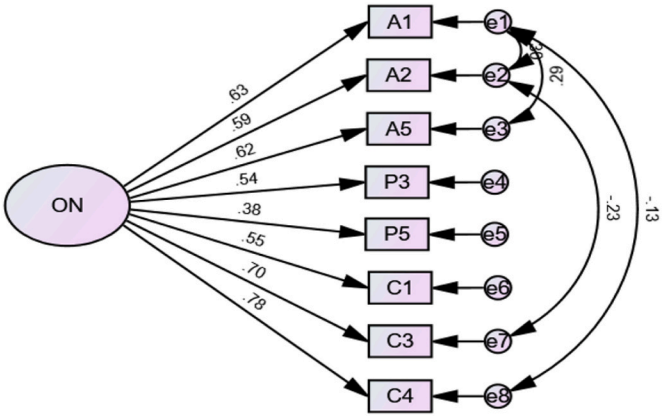

Fig. 2. CFA 1-factor structure goodness-of-fit model.

2021) that suggested exaggerated focus on appearance as a motive behind the drive for "healthy" nutrition.

One-way ANOVA indicated a significant difference between the groups on their performance with STONE, F (5, 235) = 9.34, p = .0001. A Gabriel's *post hoc* procedure revealed that the group with suspected ON scored significantly higher than all other groups (medical p = .0001; religious diet p = .010; eating disorder diagnosis p = .034; weight maintainers for professional reasons p = .047; and control group p = .0001). Table 4 presents means and standard deviations for all groups' post-hoc comparisons.

4.2.3. Accuracy of STONE

The area under the ROC curve (AUC) for the STONE, differed significantly from .50 (p < .0001). STONE was able to predict possible ON status among the control group and those with possible ON (BOT group). Fig. 3 presents the ROC curve.

Fig. 3 suggests that there would be an 83 % likelihood that a randomly selected individual from the BOT group would score above the cut-off point, compared with the control group (AUC = .830, SE = .045, 95 % CI .741, .918). To establish the cut-off scores, we examined the curve and identified the coordinates that reflect minimum distance from the left-upper corner of the unit square. To allow for flexibility in application of the scale, three cut-off points were selected, where point A prioritises specificity over sensitivity, point B reflects the balance between specificity and sensitivity, and point C favours sensitivity over specificity. True and false positive rates and positive and negative likelihood ratios for each cut-off point are presented in Table 5.

Considering cut-off point B, STONE identified 80 % of the individuals in the BOT group as having possible ON. The positive likelihood ratio at this point predicts that individuals who scored above the cut-off value are 3.55 times more likely to have ON. Negative likelihood ratio of .282 indicates a 28.2-fold decrease in the odds of having ON in a person with a negative screening test result.

5. Study Four

The last study focused on the test-retest reliability assessment of STONE.

5.1. Methods

5.1.1. Measures

The measure included only the newly developed 8-item STONE (Appendix C section).

5.1.2. Participants and procedure

The sample included participants from the control and BOT groups that took part in Study 3 and responded to the "call" for participation (n = 71). There were 34 individuals in the Control group and 37 in the BOT group. The recruitment process took place three weeks after completion of Study 3 and was identical. Participants recruited via Prolific received monetary compensation with average reward per hour = £18.75. Participants' Prolific ID was used to match their data to their scores from the previous study. The measure included only STONE. Pearson correlation coefficients between the scores of the scale from Study 3 and this study were calculated to assess test-retest reliability with coefficients of 1 considered a perfect reliability, >.75 as excellent reliability, from .60 to .74 as good, and from .40 to .59 as fair (Cicchetti, 1994).

Table 4
Post-hoc comparison of groups' performance on STONE.

Group	*M*	*SD*
BOT	30.20	7.41
Medical	23.19	5.84
Religious diet	24.38	6.42
ED	25.46	8.08
Prof. reasons	25.73	7.92
Control	20.08	6.22

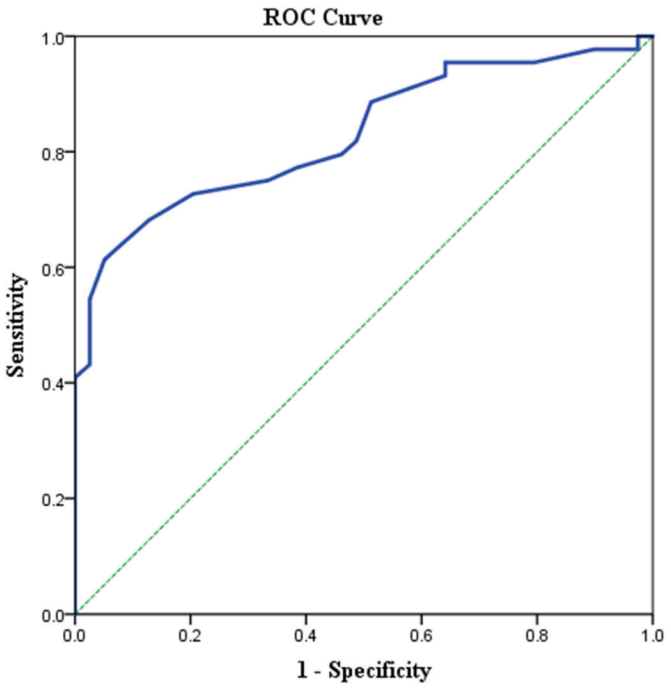

Fig. 3. ROC curve of STONE to predict possible ON between the control group and individuals that scored at least 5 points on BOT.

Table 5
Cut-off values and associated calculations.

Cut-off point	Cut-off value	Sensitivity	1-Specificity	TPR	FPR	PLR	NLR
A	28.5	.614	.051	.931	.037	12.04	.083
B	25.5	.727	.205	.800	.186	3.55	.282
C	20.5	.886	.513	.661	.833	1.73	.579

Note: TPR = true positive rate; FPR = false positive rate; PLR = positive likelihood ratio; NLR = negative likelihood ratio.

5.2. Results

A good test-retest reliability was observed between the scores of the control, r = .604, p = .01, and the BOT groups, r = .660, p = .01. However, there was a small discrepancy between the number of cases scoring above the cut-off in Study Three and this study (two cases). Several factors may account for this shift, including natural fluctuations in eating behaviours, situational changes (e.g., health concerns, exposure to dietary messaging), or response variability due to mood or context at the time of retest. The shift in classification of two participants suggests that scores close to the cut-off value may be influenced by these situational or behavioural fluctuations. However, life changes were not assessed in the retest phase which limits the ability to interpret the results. Future studies should include brief behavioural or lifestyle inventories to contextualize changes in STONE scores over time and strengthen the interpretive value of reliability estimates. Table 6 reports the true positive/negative and false positive/negative rates from Studies Three and Four. Cut-off point B was used to estimate classification. Because not all participants from Study Three responded to the call for participation, which resulted in unequal sample sizes in Studies Three and Four, values are presented in percentages.

Table 6
Percentage of participants scoring above the "B" cut-off point in Studies 3 and 4.

	Study 3		Study 4	
	True positive (BOT)	True negative (BOT)	True positive (BOT)	True negative (BOT)
Predicted positive (STONE)	72.73 %	20.51 %	81.08 %	32.35 %
Predicted negative (STONE)	27.27 %	79.49 %	18.92 %	67.65 %

6. Discussion

The aim of the studies described in this article was to develop an independent short screening tool for ON that would include physical appearance as a motivation behind individuals' restrictions on their diet, reflect avoidance of foods perceived as "unhealthy" or "impure" and also rigid behavioural practices surrounding food preparation and food shopping. These components align with the evidence from empirical studies of ON that conceptualize it as a multifaceted condition driven by internalised appearance ideals and concerns (Messer et al., 2022). For example, items such as "I follow my diet in order not to gain weight", "I eat healthy because I want to improve the way I look", and "My diet is designed to keep me at a specific weight" tap into appearance-driven motives. From a cognitive-behavioural perspective, STONE items reflect ON as a condition that is driven by beliefs about food's quality and reinforced by rigid practices around food consumption and preparation. Items like "I eat only healthy food" and "I avoid foods that were treated with pesticides" reflect purity- and health-focused motives. Finally, items such as "Other people have mentioned that my diet is too restrictive", "I measure every portion", and "I spend a lot of time researching nutritional composition of foods" capture the behavioural rigidity and obsessive tendencies characteristic of orthorexia and related to compulsive control seen in cognitive-behavioural and obsessive-compulsive frameworks. As summarised in Fig. 4, our newly developed screening tool for ON has good psychometric properties.

The resulting scale includes eight items and is a unidimensional measure of ON. The scale demonstrated a good model fit and results supported the internal consistency. STONE showed associations with the measures of eating pathology, obsessive-compulsive tendencies and investment in appearance offering evidence for the scale's convergent validity. STONE was able to differentiate between the groups with the group of individuals that self-identified with BOT statements scoring significantly higher than all other groups, which points to the scale's discriminant validity. We also offered three cut-off points for the practical application of the scale. In practice, scales that are highly sensitive run a risk of identifying false positive cases, while high specificity entails a risk of missing true positive cases (Akobeng, 2007). The first cut-off value (A) was set prioritising specificity over sensitivity and could be used in research scenarios when high importance is placed on ON status as an inclusion criterion. Using cut-off score "A" could potentially ensure exclusion of false positives from the ON sample. The second cut-off score (B) was set to represent a reasonable balance between sensitivity and specificity. The third cut-off score (C) was set prioritising sensitivity and could be used in cases when the scale is used to identify individuals at risk of developing ON. In practice, this cut-off score would be useful for screening individuals involved in disciplines or occupations (e.g. modelling, bodybuilding) where circumstances might accentuate their ON tendencies with potential adverse consequences for an individual's health. This could constitute a preventative tool against the development of pathological eating behaviour.

The test-retest study indicated that the scale identified more individuals scoring above the cut-off point in the group of those scoring high on the BOT than in Study Three. This finding was unexpected considering characteristics of ON behaviours discovered in nutritional and qualitative studies conducted by this research team (Mitrofanova, Pummell, et al., 2021; Mitrofanova, Mulrooney, & Petroczi, 2021). Individuals described their diets and their food-related behaviours as stable with little variation. If the behaviours and diets are the same, then one possible explanation for the variation in scores obtained three weeks apart could be that even though the behaviours may remain stable, their self-perceptions might have changed. This aspect must be researched further in the future, as to date, studies do not report on progression/development of ON over time. Considering the discrepancies observed in the test-retest study, a single application of the scale might not be enough to form a diagnosis. It is recommended that at-risk individuals are assessed at two timepoints. The scale should be used in combination with anthropometric and nutritional measures. Combining these measures will provide a clearer picture of the level of potential physical impairment caused by ON.

The scale developed in this article offers an alternative instrument to assess ON and before discussing the limitations of the presented studies several advantages are worth mentioning. The structure of the scale was explored using a sample of individuals that self-identified with the BOT statements, which were developed based on the diagnostic criteria and ON definition proposed by Bratman (2017) in contrast to most studies that attempted to develop diagnostic tools for ON using student samples (Barrada & Roncero, 2018; Donini et al., 2005; Gleaves et al., 2013) who cited their sample characteristics as one of the limitations of their

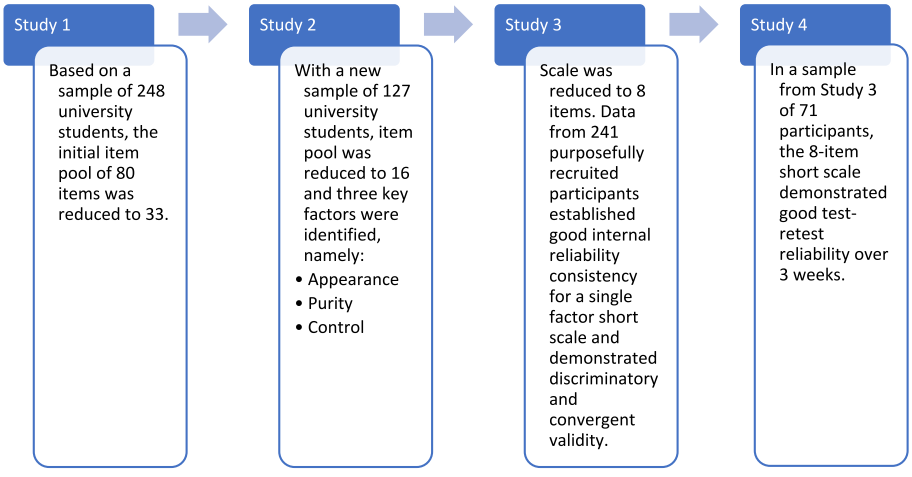

Fig. 4. Summary of the outcomes of the four studies included in the development process of STONE.

designs. Another advantage is that all items were developed based on qualitative interviews with an exploration of dietary intakes of individuals displaying orthorexic tendencies. This method has not been previously utilised in the development of measures. For example, for the development of the EHQ scale items were generated based on the Bratman's case studies and were agreed on via consultation with graduate researchers familiar with the symptoms of ON. Generating items informed by the individuals in particular behavioural circumstances, however, offers an opportunity to ground the concept of ON in real-life observations and enhance the quality of the measure (Rowan & Wulff, 2007).

6.1. Limitations and future directions

There are some limitations to the developed measure worth noting. First, the final 8-item scale does not capture negative affect and physical impairment dimensions of ON proposed by Dunn and Bratman (2016). Items reflecting negative affect were included in the initial list of items. However, these items were eliminated at the EFA stage. The items capture behavioural and motivational aspects of ON but will not assess whether hypothesised orthorexic behaviours cause any impairment in individual's social or occupational functioning. The exclusion of these items may result in underestimating ON severity among individuals experiencing emotional distress and psychosocial dysfunction. Also, application of the scale alone would not be sufficient to identify whether adherence to self-defined "healthy" diet causes any adverse physical consequences (e.g. malnutrition). Statements reflecting social impairment and compensatory behaviours were included in the process of development but were eliminated from the scale during the EFA. STONE is recommended for use in research settings to identify individuals who may exhibit ON tendencies. In other settings, assessment should include measures of negative affect, physical health consequences, and nutritional assessment. For example, its use could be complemented with other tools that capture impairment and affective distress, such as the ONI, which includes dedicated subscales assessing emotional and physical impairment. One direction for future research would be to generate and test clearly worded items reflecting these factors. Second, evaluating concurrent validity of the scale was not feasible within the series of studies we conducted to date and described in this article. Considering that independently developed questionnaires measure the same construct, in the future, the scale's performance needs to be evaluated with the existing measures of ON (e.g. EHQ, TOS). Future studies should aim to address this gap. It would be particularly informative to compare the scales' performance with the TOS (Barrada & Roncero, 2018). TOS scale claims to differentiate between the drive for healthy diet and the pathological dimension when this drive negatively impacts individual's functioning. Comparing the scale to TOS would offer additional information on its ability to place individuals on a spectrum from a healthy interest in nutrition to the point where this interest impairs their functioning.

Last, but not least, data collection was carried out in the UK using adult population. Therefore, validation of STONE in adolescents, and in different cultural contexts is warranted. Disordered eating often has its roots in childhood and adolescence (Volpe et al., 2016). Some studies indicate a similar trend for the onset of ON indicating a higher prevalence among younger individuals (Fidan et al., 2010; McComb & Mills, 2016). However very few studies use adolescent samples. Considering adolescence as an age of vulnerability to disordered eating (Favaro et al., 2009) and ON, early screening is essential for prevention.

Culture is an important influence on individuals' food choices, which can symbolise identity, values, lifestyle, availability of resources, social norms and environmental attitudes of a group (Enriquez & Archila-Godinez, 2021). Furthermore, our understanding of disordered eating practices that were initially considered to be a product of exposure to "Western beauty ideals" in other countries has now evolved to be viewed as a result of local industrialisation and urbanisation with their own culture-specific presentations (Pike & Dunne, 2015). To date, research exploring ON in various cultural contexts is limited with most studies focusing on comparing prevalence of ON among individuals from different cultural backgrounds and yields inconsistent results (e.g., Gramaglia et al., 2017; Parra-Fernández et al., 2019). These inconsistencies may be due to the use of different questionnaires in such studies, which does not allow for comparison. Validating STONE in different cultures could allow for exploration of culture-specific trajectories of ON.

6.2. Recommended use

STONE is a new 8-item tool we recommend as a rapid screening for ON in the research context. Due to its brevity, it can be easily combined with other scales in a battery of psychometric tests to explore ON or related phenomena.

Consistent with the view of ON as a condition that presents as a set of behaviours aimed at rigid dietary control over food-related practices, avoidance of foods considered "impure" and motivation to enhance one's physical appearance, STONE scores were positively related to measures of obsessive-compulsive tendencies, eating pathology and appearance orientation. While the role of body image and desire to achieve a certain physique have been implicated in eating disorders, the current understanding of ON lacks clarity about the role of physical appearance as a motive for rigid food rules (Messer et al., 2022). The latest set of proposed diagnostic criteria recognised rigid dietary control and exclusion of "impure" foods among Criteria A1 (Donini et al., 2022). Low body weight and thin or muscular ideals were suggested by the same document to be a consequence of ON rather than a motivating factor. STONE may be useful in research that aims to explore the consequences of ON dietary restrictions and clarify the role of appearance and related concepts. Given growing interest in the relationship between ON and sociocultural pressures (e.g., clean eating, body ideals), STONE is also well-positioned for use in studies exploring the appearance-related motivations behind restrictive eating and may be especially relevant for testing models grounded in objectification theory or body image disturbance, which remain debated in ON research (Messer et al., 2022; White et al., 2020).

In longitudinal research, STONE can be used to track the development and stability of behaviours reflective of ON over time, particularly when combined with nutritional assessments or qualitative interviews. This could help clarify how motivational drivers (e.g., purity, appearance) evolve and whether they contribute to the development of functional impairment.

In epidemiology research, STONE could be used to explore ON tendencies among individuals that may be considered vulnerable due to the nature of their professional occupation (e.g. fashion models, athletes, nutritionists, dietitians, ballet dancers) or due to a presence of a health condition (e.g. diabetes, irritable bowel syndrome, coeliac disease). Considering that such individuals may restrict their diet for reasons other than ON, using a tool that can differentiate these reasons from ON would allow for screening in these vulnerable groups.

6.3. Conclusion

STONE is a short tool to screen for orthorexic tendencies in a research context. It demonstrated good psychometric properties and an excellent ability to discriminate between individuals that self-identified with statements reflective of ON symptoms and those that may be restricting their dietary intake for other reasons. By offering a concise measure grounded in understanding of motivation, cognitions and behaviours indicative of ON, STONE has the potential to enhance conceptual clarity, advance empirical research, and contribute to a more nuanced understanding of this condition—ultimately informing public health strategies that distinguish between healthful eating and disordered patterns masked by wellness ideals. The scale does not capture

negative emotional and physical consequences of ON, which suggests that it should be used in combination with other measures to capture these aspects. Future research efforts should focus on validating the scale with different cultural contexts, particularly in non-Western populations where cultural beliefs about health and diet may influence the expression of orthorexic behaviours. Additionally, research in adolescent populations is warranted, given emerging evidence that orthorexic attitudes can begin early and may be shaped by social media, school-based health messaging, and body image concerns. Validating the tool in this age group could facilitate early identification of risk trajectories. STONE could be used in research targeting high-risk professional groups such as athletes, fitness influencers, nutritionists, and dancers, where certain dietary restrictions are normalised and do not necessarily reflect pathology. STONE can help differentiate between performance-oriented dietary control and pathological restriction, allowing for early identification of those at risk for orthorexia nervosa within these occupational groups.

CRediT authorship contribution statement

Elina Mitrofanova: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Hilda Mulrooney:** Writing – review & editing, Project administration, Investigation. **Elizabeth Pummell:** Writing – review & editing, Project administration, Investigation. **Andrea Petróczi:** Writing – review & editing, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Ethical approval

Ethical approval was granted for all studies by the Faculty Research Ethics Committee at Kingston University London.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.appet.2025.108227.

Appendix A

Table A1
Retained Items with their Corrected-Item-Total-Correlation coefficients.

Cluster	Items	Corrected Item-Total Correlation
A	Healthy eating is a large part of who I am.	.747
A	My food choices are based on a desire to maximise my health.	.737
A	Nutrition is a hobby of mine.	.690
A	I carefully monitor the nutritional composition of what I eat.	.669
A	I carefully check the ingredients before I buy a food item.	.662
A	I frequently seek information about nutrition (e.g. on the internet, reading books on nutrition).	.651
A	My food choices are based on my desire to maximise my fitness performance.	.650
A	I'm very specific about my food choices.	.649
A	I spend a lot of time researching nutritional composition of foods.	.648
A	I plan my meals in advance.	.617
A	My diet has more health benefits than other diets.	.617
A	I have a strict exercise routine to complement my diet.	.613
A	I actively seek the latest trends/information/news in nutrition.	.550
A	I think about healthy eating while doing something else.	.531
B	My body is pure because of my healthy diet.	.644
B	I try to keep my body as pure as possible.	.634
B	I eat only healthy food.	.587
B	I plan when to allow myself a treat outside of my diet.	.564
B	I make sure that my diet is better than most people's diet.	.504
C	My diet has many rules.	.658
C	It's difficult to find a restaurant that serves the foods that I eat.	.556
C	Other people have mentioned that my diet is too restrictive.	.539
C	I avoid food that I haven't prepared myself.	.539
C	My family has to make me a separate meal/dish when eating together (e.g. Christmas, Easter).	.536
C	I measure every portion.	.529
C	If I wasn't eating the way I do, people wouldn't be interested in me.	.523
C	I go out less frequently since I began eating healthy.	.520
C	Most of my social interactions involve a discussion about my eating habits.	.515
D	I feel a sense of achievement when I stick to my diet.	.720
D	Being able to stick to my diet has a positive impact on my mood.	.662
D	I feel better about myself when I manage to avoid slipping off my healthy diet.	.627
D	My chosen diet has a direct impact on my appearance.	.606
D	I eat healthy because I want to improve the way I look.	.565
E	It's important for me to know where the food I buy at the supermarket/market comes from.	.629
E	I avoid foods that were treated with pesticides.	.586

(*continued on next page*)

Table A1 (*continued*)

Cluster	Items	Corrected Item-Total Correlation
E	I avoid processed foods.	.509
E	I avoid genetically modified foods.	.506
F	I follow my diet in order not to gain weight.	.687
F	My diet is designed to keep me at a specific weight.	.630
F	The main motivation behind my food choices is weight management.	.574
F	If I eat something outside of my diet, I will try to make up for it and eat less or exercise more the next day.	.551

Appendix B

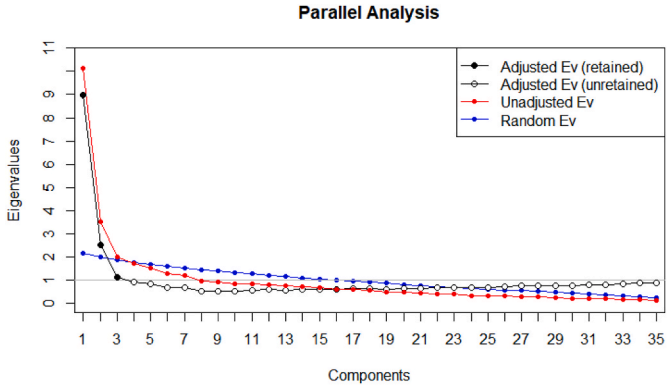

Fig. B.1. Parallel analysis of the Scale (35 items).

Appendix C

Table C1
8-item STONE.

Item	Dimension
I follow my diet in order not to gain weight (A1).	Appearance
I eat healthy because I want to improve the way I look (A2).	Appearance
My diet is designed to keep me at a specific weight (A5).	Appearance
I eat only healthy food (P3).	Purity
I avoid foods that were treated with pesticides (P5).	Purity
Other people have mentioned that my diet is too restrictive (C1).	Control
I measure every portion (C3).	Control
I spend a lot of time researching nutritional composition of foods (C4).	Control

Data availability

Data will be made available on request.

References

Akobeng, A. K. (2007). Understanding diagnostic tests 1: Sensitivity, specificity, and predictive values. *Acta Pædiatrica, 96*, 338–341. https://doi.org/10.1111/j.1651-2227.2006.00180.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596

Argyrides, M., Anastasiades, E., Maïano, C., & Swami, V. (2024). Greek adaptation of the teruel orthorexia scale (TOS) in adults from the Republic of Cyprus: A bidimensional model May not be universal. *Appetite, 194*, Article 107180. https://doi.org/10.1016/j.appet.2023.107180

Atchison, A. E., & Zickgraf, H. F. (2022). Orthorexia nervosa and eating disorder behaviors: A systematic review of the literature. *Appetite, 177*, Article 106134. https://doi.org/10.1016/j.appet.2022.106134

Baer, R. A., Fischer, S., & Huss, D. B. (2005). Mindfulness and acceptance in the treatment of disordered eating. *Journal of Rational-Emotive and Cognitive-Behavior Therapy, 23*, 281–300. https://doi.org/10.1007/s10942-005-0015-9

Barrada, J. R., & Meule, A. (2024). Orthorexia nervosa: Research based on invalid measures is invalid. *Journal of Global Health, 14*, Article 03007. https://doi.org/10.7189/jogh.14.03007

Barrada, J. R., & Roncero, M. (2018). Bidimensional structure of the orthorexia: Development and initial validation of a new instrument. *Anales de Psicología/Annals of Psychology, 34*(2), 283–291. https://doi.org/10.6018/analesps.34.2.299671

Barthels, F., Meyer, F., Huber, T., & Pietrowsky, R. (2017). Orthorexic eating behaviour as a coping strategy in patients with anorexia nervosa. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 22*, 269–276. https://doi.org/10.1007/s40519-016-0329-x

Barthels, F., Meyer, F., & Pietrowsky, R. (2015). Orthorexic eating behavior. A new type of disordered eating. *Ernahrungs Umschau, 62*(10), 156–161. https://doi.org/10.4455/eu.2015.029

Bauer, S. M., Fusté, A., Andrés, A., & Saldaña, C. (2019). The barcelona orthorexia scale (BOS): Development process using the Delphi method. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 24*, 247–255. https://doi.org/10.1007/s40519-018-0556-4

Bratman, S. (1997). Original essay on orthorexia. Retrieved September 2018 from https://www.orthorexia.com/original-orthorexia-essay/.

Bratman, S. (2017). Bratman orthorexia self-test. June 2020 at https://www.orthorexia.com/the-authorized-bratman-orthorexia-self-test/. (Accessed 13 July 2020).

Bryant, E., Miskovic-Wheatley, J., Touyz, S. W., Crosby, R. D., Koreshe, E., & Maguire, S. (2021). Identification of high risk and early stage eating disorders: First validation of

a digital screening tool. *Journal of Eating Disorders, 9*, 1–10. https://doi.org/10.1186/s40337-021-00464-y
Cash, T. F. (2015). *Multidimensional body–self relations questionnaire (MBSRQ). Encyclopedia of feeding and eating disorders*. Singapore: Springer. https://doi.org/10.1007/978-981-287-087-2_3-1
Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10
Cena, H., Barthels, F., Cuzzolaro, M., Bratman, S., Brytek-Matera, A., Dunn, T., Varga, M., Missbach, B., & Donini, L. M. (2019). Definition and diagnostic criteria for orthorexia nervosa: A narrative review of the literature. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 24*(2), 209–246. https://doi.org/10.1007/s40519-018-0606-y
Chace, S., & Kluck, A. S. (2022). Validation of the Teruel orthorexia scale and relationship to health anxiety in a US sample. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 27*(4), 1437–1447. https://doi.org/10.1007/s40519-021-01272-8
Cheshire, A., Berry, M., & Fixsen, A. (2020). What are the key features of orthorexia nervosa and influences on its development? A qualitative investigation. *Appetite, 155*, Article 104798. https://doi.org/10.1016/j.appet.2020.104798
Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284
Cinquegrani, C., & Brown, D. H. K. (2018). 'wellness' lifts Us above the food chaos': A narrative exploration of the experiences and conceptualisations of orthorexia nervosa through online social media forums. *Qualitative Research in Sport, Exercise and Health, 10*(5), 585–603. https://doi.org/10.1080/2159676X.2018.1464501
Donini, L. M., Barrada, J. R., Barthels, F., Dunn, T. M., Babeau, C., Brytek-Matera, A., … Lombardo, C. (2022). A consensus document on definition and diagnostic criteria for orthorexia nervosa. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 27*(8), 3695–3711.
Donini, L. M., Marsili, D., Graziani, M. P., Imbriale, M., & Cannella, C. (2005). Orthorexia nervosa: Validation of a diagnosis questionnaire. Eating and weight disorders – Studies on anorexia. *Bulimia and Obesity, 10*(2), e28–e32. https://doi.org/10.1007/BF03327537
Dunn, T. M., & Bratman, S. (2016). On orthorexia nervosa: A review of the literature and proposed diagnostic criteria. *Eating Behaviors, 21*, 11–17. https://doi.org/10.1016/j.eatbeh.2015.12.006
Dunn, T. M., Gibbs, J., Whitney, N., & Starosta, A. (2017). Prevalence of orthorexia nervosa is less than 1%: Data from a US sample. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 22*, 185–192. https://doi.org/10.1007/s40519-016-0258-8
Düzçeker, Y., Akgül, S., Durmaz, Y., Yaman, M., Örs, S., Tüzün, Z., Büyüktuncer, Z., & Kanbur, N. (2021). Is Ramadan fasting correlated with disordered eating behaviours in adolescents? *Eating Disorders, 29*(1), 74–87. https://doi.org/10.1080/10640266.2019.1642032
Enriquez, J. P., & Archila-Godinez, J. C. (2021). Social and cultural influences on food choices: A review. *Critical Reviews in Food Science and Nutrition, 62*(13), 3698–3704. https://doi.org/10.1080/10408398.2020.1870434
Evert, A. B., Dennison, M., Gardner, C. D., Garvey, W. T., Lau, K. H. K., MacLeod, J., Mitri, J., Pereira, R. F., Rawlings, K., Robinson, S., Saslow, L., Uelmen, S., Urbanski, P. B., & Yancy, W. S., Jr. (2019). Nutrition therapy for adults with diabetes or prediabetes: A consensus report. *Diabetes Care, 42*(5), 731–754. https://doi.org/10.2337/dci19-0014
Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*(4), 363–370. https://doi.org/10.1002/1098-108X(199412)16:4<363::AID-EAT2260160405>3.0.CO;2-%23
Favaro, A., Caregaro, L., Tenconi, E., Bosello, R., & Santonastaso, P. (2009). Time trends in age at onset of anorexia nervosa and bulimia nervosa. *The Journal of Clinical Psychiatry, 70*(12), Article 21711. https://doi.org/10.4088/JCP.09m05176blu
Fixsen, A., Cheshire, A., & Berry, M. (2020). The social construction of a concept—orthorexia nervosa: Morality narratives and psycho-politics. *Qualitative Health Research, 30*(7), 1101–1113. https://doi.org/10.1177/1049732320911364
Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment, 14*(4), 485–496. https://doi.org/10.1037/1040-3590.14.4.485
Fodor, A., Varga, B. A., & Rigó, A. (2025). Beyond healthy eating: Introducing ONI-Hu, the Hungarian version of the orthorexia nervosa inventory. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 30*(1), 1–13. https://doi.org/10.1007/s40519-025-01745-0
Formann, A. K. (1984). *Die Latent-Class-Analyse: Einführung in die Theorie und Anwendung*. Weinheim: Beltz.
Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psychological Medicine, 12*(4), 871–878. https://doi.org/10.1017/S0033291700049163
George, D., & Mallory, P. (2003). In *SPSS for windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
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. May 2020 at https://www.researchgate.net/publication/263808892_Measuring_Orthorexia_Development_of_the_Eating_Habits_Questionnaire.
Gomez, F., & Perez, M. (2022). Motivation for dietary restraint, and disordered eating among adults experiencing food insecurity in the United States. *International Journal of Eating Disorders, 55*(6), 838–844. https://doi.org/10.1002/eat.23730
Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale. NJ: Lawrence Erlbaum Associates.
Greville-Harris, M., Smithson, J., & Karl, A. (2019). What are people's experiences of orthorexia nervosa? A qualitative study of online blogs. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 1–10. https://doi.org/10.1007/s40519-019-00809-2
He, J., Ma, H., Barthels, F., & Fan, X. (2019). Psychometric properties of the Chinese version of the düsseldorf orthorexia scale: Prevalence and demographic correlates of orthorexia nervosa among Chinese university students. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 24*, 453–463. https://doi.org/10.1007/s40519-019-00656-1
Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 32*, 179–185.
Horovitz, O., & Argyrides, M. (2023). Orthorexia and orthorexia nervosa: A comprehensive examination of prevalence, risk factors, diagnosis, and treatment. *Nutrients, 15*(17), 3851. https://doi.org/10.3390/nu15173851
Jackson, B., Cooper, M. L., Mintz, L., & Albino, A. (2003). Motivations to eat: Scale development and validation. *Journal of Research in Personality, 37*(4), 297–318. https://doi.org/10.1016/S0092-6566(02)00574-3
Kaiser, H. (1974). An index of factor simplicity. *Psychometrika, 39*, 31–36.
Lasson, C., Rousseau, A., Vicente, S., Goutaudier, N., Romo, L., Roncero, M., & Barrada, J. R. (2023). Orthorexic eating behaviors are not all pathological: A French validation of the Teruel orthorexia scale (TOS). *Journal of Eating Disorders, 11*(1), 65. https://doi.org/10.1186/s40337-023-00764-5
López-Gil, J. F., Tárraga-López, P. J., Hershey, M. S., Lopez-Bueno, R., Gutierrez-Espinoza, H., Soler-Marin, A., … Victoria-Montesinos, D. (2023). Overall proportion of orthorexia nervosa symptoms: A systematic review and meta-analysis including 30 476 individuals from 18 countries. *Journal of Global Health, 13*. https://doi.org/10.7189/jogh.13.04087
MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84
McGovern, L., Gaffney, M., & Trimble, T. (2020). The experience of orthorexia from the perspective of recovered orthorexics. Eating and weight disorders – Studies on anorexia. *Bulimia and Obesity*. https://doi.org/10.1007/s40519-020-00928-1
McNeil, B. J., Keller, E., & Adelstein, S. J. (1975). Primer on certain elements of medical decision making. *New England Journal of Medicine, 293*(5), 211–215. https://doi.org/10.1056/NEJM197507312930501
Messer, M., Liu, C., McClure, Z., Mond, J., Tiffin, C., & Linardon, J. (2022). Negative body image components as risk factors for orthorexia nervosa: Prospective findings. *Appetite, 178*, Article 106280. https://doi.org/10.1016/j.appet.2022.106280
Meule, A., Holzapfel, C., Brandl, B., Greetfeld, M., Hessler-Kaufmann, J. B., Skurk, T., Quadflieg, N., Schlegl, S., Hauner, H., & Voderholzer, U. (2020). Measuring orthorexia nervosa: A comparison of four self-report questionnaires. *Appetite, 146*, Article 104512. https://doi.org/10.1016/j.appet.2019.104512
Mitrofanova, E., Mulrooney, H., & Petroczi, A. (2021). Assessing psychological and nutritional impact of suspected orthorexia nervosa: A cross-sectional pilot study. *Journal of Human Nutrition and Dietetics, 34*(1), 42–53. https://doi.org/10.1111/jhn.12797
Mitrofanova, E., Pummell, E. K., Mulrooney, H. M., & Petróczi, A. (2021). Using behavioural reasoning theory to explore reasons for dietary restriction: A qualitative study of orthorexic behavioural tendencies in the UK. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.685545
Mohamed Halim, Z., Dickinson, K. M., Kemps, E., & Prichard, I. (2020). Orthorexia nervosa: Examining the eating habits Questionnaire's reliability and validity, and its links to dietary adequacy among adult women. *Public Health Nutrition, 23*(10), 1684–1692. https://doi.org/10.1017/S1368980019004282
Moller, S., Apputhurai, P., & Knowles, S. R. (2019). Confirmatory factor analyses of the ORTO 15-, 11- and 9-item scales and recommendations for suggested cut-off scores. Eating and weight disorders – Studies on anorexia. *Bulimia and Obesity, 24*, 21–28. https://doi.org/10.1007/s40519-018-0515-0
Ng, Q. X., Lee, D. Y. X., Yau, C. E., Han, M. X., Liew, J. J. L., Teoh, S. E., … Chee, K. T. (2024). On orthorexia nervosa: A systematic review of reviews. *Psychopathology*, 1–14.
Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
Oberle, C. D., De Nadai, A. S., & Madrid, A. L. (2021). Orthorexia nervosa inventory (ONI): Development and validation of a new measure of orthorexic symptomatology. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 26*, 609–622.
Oberle, C. D., Marcell, H. S., & Noebel, N. A. (2021). Orthorexia nervosa and substance use for the purposes of weight control, conformity, and emotional coping. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 1–9.
Oberle, C. D., Samaghabadi, R. O., & Hughes, E. M. (2017). Orthorexia nervosa: Assessment and correlates with gender, BMI, and personality. *Appetite, 108*, 303–310. https://doi.org/10.1016/j.appet.2016.10.021
Opitz, M. C., Newman, E., Mellado, A. S. A. V., Robertson, M., & Sharpe, H. (2020). The psychometric properties of orthorexia nervosa assessment scales: A systematic review and reliability generalization. *Appetite*. , Article 104797.. https://doi.org/10.1016/j.appet.2020.104797
Pauzé, A., Plouffe-Demers, M. P., Fiset, D., Saint-Amour, D., Cyr, C., & Blais, C. (2021). The relationship between orthorexia nervosa symptomatology and body image attitudes and distortion. *Scientific Reports, 11*(1), Article 13311. https://doi.org/10.1038/s41598-021-92569-2

Pike, K. M., & Dunne, P. E. (2015). The rise of eating disorders in Asia: A review. *Journal of eating disorders, 3*, 1–14. https://doi.org/10.1186/s40337-015-0070-2

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Ramacciotti, C. E., Perrone, P., Coli, E., Burgalassi, A., Conversano, C., Massimetti, G., & Dell'Osso, L. (2011). Orthorexia nervosa in the general population: A preliminary screening using a self-administered questionnaire (ORTO-15). *Eating and weight disorders-Studies on Anorexia, Bulimia and Obesity, 16*, e127–e130. https://doi.org/10.1007/BF03325318

Reynolds, R. (2018). Is the prevalence of orthorexia nervosa in an Australian university population 6.5. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity, 23*, 453–458. https://doi.org/10.1007/s40519-018-0535-9

Rolstad, S., Adler, J., & Rydén, A. (2011). Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health, 14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Rowan, N., & Wulff, D. (2007). Using qualitative methods to inform scale development. *Qualitative Report, 12*(3), 450–466. Retrieved July 2020 from https://nsuworks.nova.edu/tqr/vol12/iss3/7.

Tabachnick, B. G., & Fidell, L. S. (2019). In *Using multivariate statistics* (7th ed.). Pearson.

Valente, M., Syurina, E. V., & Donini, L. M. (2019). Shedding light upon various tools to assess orthorexia nervosa: A critical literature review with a systematic search. Eating and weight disorders – Studies on anorexia. *Bulimia and Obesity, 24*, 671–682. https://doi.org/10.1007/s40519-019-00735-3

Valente, M., Syurina, E. V., Muftugil-Yalcin, S., & Cesuroglu, T. (2020). "keep yourself alive": From healthy eating to progression to orthorexia nervosa. A mixed methods study among young women in the Netherlands. *Ecology of Food and Nutrition*. https://doi.org/10.1080/03670244.2020.1755279

Van Strien, T., Frijters, J. E., Bergers, G. P., & Defares, P. B. (1986). The dutch eating behaviour questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behaviour. *International Journal of Eating Disorders, 5*(2), 295–315. https://doi.org/10.1002/1098-108X(198602)5:2<295::AID-EAT2260050209>3.0.CO;2-T

Volpe, U., Tortorella, A., Manchia, M., Monteleone, A. M., Albert, U., & Monteleone, P. (2016). Eating disorders: What age at onset? *Psychiatry Research, 238*, 225–227. https://doi.org/10.1016/j.psychres.2016.02.048

White, M., Berry, R., & Rodgers, R. F. (2020). Body image and body change behaviors associated with orthorexia symptoms in males. *Body Image, 34*, 46–50. https://doi.org/10.1016/j.bodyim.2020.05.003

World Health Organization. (2020). Healthy diet. https://www.who.int/news-room/fact-sheets/detail/healthy-diet.