U.PORTO

MESTRADO EM PSICOLOGIA
PSICOLOGIA CLÍNICA E DA SAÚDE

Three Decades of Eating Disorder Examination-Questionnaire: A Systematic Review

Maria Luís Rocha Conde

M

2025

Universidade do Porto

Faculdade de Psicologia e de Ciências da Educação

THREE DECADES OF EATING DISORDER EXAMINATION-QUESTIONNAIRE: A SYSTEMATIC REVIEW

Maria Luís Rocha Conde

Junho 2025

Dissertação apresentada no Mestrado em Psicologia, área de especialização de Psicologia Clínica e da Saúde, Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto, orientada pela Professora Doutora **Eva Martins da Conceição** (FPCEUP) e coorientada pelo Professor Doutor **Jorge Fernando Pereira Sinval**.

AVISOS LEGAIS

O conteúdo desta dissertação reflete as perspetivas, o trabalho e as interpretações do/da autor/a no momento da sua entrega. Esta dissertação pode conter incorreções, tanto concetuais como metodológicas, que podem ter sido identificadas em momento posterior ao da sua entrega. Por conseguinte, qualquer utilização dos seus conteúdos deve ser exercida com cautela.

Ao entregar esta dissertação, o/a autor/a declara que a mesma é resultante do seu próprio trabalho, contém contributos originais e são reconhecidas todas as fontes utilizadas, encontrando-se tais fontes devidamente citadas no corpo do texto e identificadas na secção de referências. O/A autor/a declara, ainda, que não divulga na presente dissertação quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor ou de propriedade industrial

Agradecimentos

Esta dissertação representa o culminar de um percurso académico desafiante, mas profundamente enriquecedor e transformador, que não teria sido possível sem o apoio, incentivo e presença de várias pessoas.

Em primeiro lugar, agradecer à Professora Doutora Eva Conceição, pela sua orientação rigorosa, disponibilidade constante e incentivo ao pensamento crítico. A sua exigência e generosidade foram fundamentais para o desenvolvimento deste trabalho. Agradecer, também, ao Professor Doutor Jorge Sinval por ter contribuído, na qualidade de coorientador, com a sua experiência e competência. À Catarina Gomes, por ter aceitado fazer parte deste desafio e me apoiar mais do que pode imaginar, sempre com uma palavra amiga.

Aos meus professores, agradeço por toda a aprendizagem que contribuiu para o meu crescimento académico e pessoal.

Aos meus familiares, em especial aos meus pais, por cada palavra de encorajamento, compreensão e silêncio cúmplice nas horas de estudo. Obrigada por acreditarem em mim e pelo amor incondicional. Este trabalho também vos pertence.

Aos amigos de sempre e aos que a vida trouxe pelo caminho, obrigada por me lembrarem que há vida para além das sebentas e dos exames, mas por também se juntarem a mim na biblioteca sempre que era preciso. Que nos continuemos a apoiar em todas as horas e que, juntos, consigamos celebrar cada pequena conquista!

E, por fim, a todos os que cruzaram este percurso de forma breve mas significativa - com um gesto, uma palavra ou um exemplo - o meu mais sincero agradecimento.

Resumo

As Perturbações da Alimentação e da Ingestão caracterizam-se por alterações significativas no comportamento alimentar/ingestão. Dadas as comorbilidades médicas e elevadas taxas de mortalidade associadas às perturbações do espetro do peso, é imperativo ter acesso a medidas de avaliação ajustadas e garantir a sua eficácia e adequação. Posto isto, é fundamental que instrumentos como o Eating Disorder Examination-Questionnaire (EDE-Q 6.0; uma das medidas de autorrelato mais utilizadas na investigação e prática clínica) se encontrem adaptados para diferentes populações. Este estudo procura sistematizar os dados atuais relativos às medidas psicométricas (consistência interna, fiabilidade teste-reteste, validade de critério, validade de construto e validade concurrente), estrutura fatorial e invariância do EDE-Q 6.0 em diferentes categorias diagnósticas, culturas, género e populações clínicas e não clínicas.

Esta revisão foi conduzida de acordo com as guidelines PRISMA, utilizando as plataformas PubMed, Scopus e APA PsycNet. Foram incluídos estudos com amostras de adultos e cujo objetivo foi examinar pelo menos uma das propriedades psicométricas do EDE-Q 6.0, estrutura fatorial e/ou invariância da medida. Foram excluídos estudos com populações não adultas, versões diferentes do EDE-Q e que não se dedicavam ao estudo de, pelo menos, uma propriedade psicométrica.

Foram encontrados 35 estudos que investigaram as propriedades psicométricas do EDE-Q 6.0. Embora o EDE-Q 6.0 apresente boa consistência interna, fiabilidade teste-reteste e validade, a estrutura fatorial revelou-se uma questão crítica — a estrutura original de Fairburn não foi suportada pelos resultados da CFA ou EFA. A invariância da medida continua pouco explorada, com resultados preliminares que sugerem falta de equivalência entre amostras clínicas e não clínicas, género e grupos étnicos/raciais.

Há necessidade de investigação adicional robusta sobre a estrutura fatorial e invariância, nomeadamente invariância longitudinal. É também importante que estudos futuros se concentrem em análises com amostras mais diversificadas.

Palavras-chave: Comportamento Alimentar; Eating Disorder Examination-Questionnaire; Medidas Psicométricas; Estrutura Fatorial; Invariância

Abstract

Eating Disorders (ED) are characterized by significant changes in eating behaviour/ingestion. Given the medical comorbidities and high mortality rates associated with weight spectrum disorders, it is imperative to have access to adjusted assessment measures and to ensure their effectiveness and appropriateness. It is therefore essential that instruments such as the Eating Disorder Examination-Questionnaire (EDE-Q 6.0; one of the most widely used self-report measures in research and clinical practice) are adapted for different populations. As a result, this study seeks to systematize the current data regarding the psychometric measures (internal consistency, test-retest, criterion validity, construct validity and concurrent validity), factor structure and invariance of the EDE-Q 6.0 in different diagnostic categories, cultures, gender and clinical and non-clinical populations.

This review was conducted in accordance with the PRISMA guidelines, using the PubMed, Scopus and APA PsycNet platforms. We included studies with adult samples whose aim was to examine at least one of the psychometric properties of the EDE-Q 6.0, factor structure and/or measurement invariance. Studies using non-adult populations, versions other than the EDE-Q 6.0 and which were not dedicated to the study of at least one psychometric property were excluded.

Thirty-five studies were found to investigate the psychometric properties of the EDE-Q 6.0. Although the EDE-Q 6.0 shows good internal consistency, test-retest reliability and validity, the structural validity has been proven to be a critical issue — Fairburn's original structure was not supported by CFA or EFA results. The measurement invariance remains little explored, with initial results suggesting a lack of equivalence between clinical and non-clinical samples, gender and ethnic/racial groups.

Based on the results, there is a need for additional robust research into its factor structure and measurement invariance, namely longitudinal invariance. It is also important for future studies to focus on analyses with more diverse samples.

Keywords: Eating Behaviour, Eating Disorder Examination-Questionnaire, Psychometric Measures, Factor Structure; Invariance

Table of Contents

I – Conceptual framework

1.1 Eating Disorders (ED)

Eating disorders (ED) can be broadly defined as serious mental disorders characterized by significant changes in eating behaviour or food intake (American Psychiatric Association; APA, 2013), with this behaviour being impacted by social, demographic and cultural aspects, perception, previous experiences and the individual's nutritional status (Köster, 2009).

Based on the DSM-V (APA, 2013), the following disorders are considered eating disorders: pica, mericism or rumination disorder, anorexia nervosa (AN), bulimia nervosa (BN), binge eating disorder (BED), avoidant/restrictive food intake disorder (ARFID), other specified feeding or eating disorders (OSFED) and eating disorders not otherwise specified (EDNOS).

Despite the clear differentiation in the diagnostic categorization of EDs, it is essential to take into account the similarities and significant number of central common features in their characterization (Fairburn & Harrison, 2003; Fairburn, 2008). Thus, AN, BN and BED are understood by Fairburn (2008), according to the Transdiagnostic Model, as having a core psychopathology of a cognitive nature, common to the different diagnoses. In other words, transdiagnostic characteristics capable of "maintaining" the ED, regardless of the diagnostic category. In addition, research carried out on the evolution of EDs has revealed a tendency towards temporal migration - over time, patients pass through more than one diagnostic category (Fairburn & Harrison, 2003).

In terms of the prevalence of eating disorders, it increases in the transition phase that occurs from adolescence to adulthood (15-25 years) (Schmidt et al., 2016), thus reaching a "range of young people for whom health problems with this level of impairment would not be anticipated" (Vaz, Conceição & Machado, 2009, p.189). Despite the possible variations resulting from the sample and methods used, there is a common understanding that there is a higher prevalence of eating disorders in young females (Fairburn & Harrison, 2003).

In addition to clinical suffering resulting from EDs, there is also impairment in domains such as physical health and psychosocial functioning (APA, 2013). Thus, other disorders such as depression and anxiety, the consumption of illicit substances and a higher risk of suicidal ideation are often associated with eating disorders (Swanson et al., 2011). Additionally, numerous psychiatric disorders such as autism spectrum disorders (Huke et al., 2013), personality disorders, obsessive-compulsive disorder and attention deficit hyperactivity disorder are prominent in these patients (Rikani et al., 2013).

Based on the literature, as a result of the prolonged state of malnutrition, the excessive practice of physical activity and the presence of binge eating and/or purging behaviours, medical comorbidities can be developed. These can vary in severity, affecting different systems such as cardiac, gastrointestinal and metabolic, as well as bone, oral and reproductive health (Hambleton et al., 2022). It is common, particularly in Anorexia Nervosa, for cardiovascular complications to arise, which result in a high proportion of deaths in these patients (Giovinazzo et al., 2019). This disorder is also commonly associated with a higher risk of heart failure due to severe malnutrition, dehydration and electrolyte imbalances (Gosseaume et al., 2019). At a gastrointestinal level, patients often reveal a wide variety of complaints such as early satiety, postprandial discomfort (Holt et al., 1981), constipation, abdominal fullness, acute intestinal occlusion and swelling of the salivary glands (Zipfel et al., 2006). Still in the field of medical comorbidities, the binge eating behaviours characteristic of BN and BED are associated with higher rates of metabolic syndrome (Bulik, Sullivan & Kendler, 2002; Roehrig et al., 2009) - a group of factors that increase the risk of heart disease, diabetes mellitus, stroke, among others (NHLBI, 2022). From the rapid literature review carried out by Hambleton et al. (2022), it was possible to identify evidence of bone loss and/or low bone mineral density. In particular, in patients with AN, high rates of bone resorption follow chronic malnutrition and consequent osteoporosis, increased risk of fractures and scoliosis (Hambleton et al., 2022; Zaina et al., 2018). As with the gastrointestinal disorders mentioned above, purging behaviours (namely self-induced vomiting) are associated with oral health disorders (Hambleton et al., 2022), such as dental erosion (Hermont et al., 2014). It is also worth highlighting the comorbidity between EDs and reproductive health dysfunctions and infertility, since this type of patient tends to have lower birth rates and higher risk rates for spontaneous and induced abortions, premature births, caesarean deliveries and intrauterine growth restrictions compared to the control group (Pasternak et al., 2012; Linna et al., 2013).

On the other end of the weight spectrum disorders, candidates for bariatric surgery have a high frequency of psychopathologies (such as anxiety and mood disorders, binge eating and bulimia nervosa (Kudel et al., 2019; Mitchell et al., 2015)), given their specific psychosocial characteristics and the social stigma surrounding obesity (Keeton et al., 2020). Considering the potential impact, there is a need to assess the client's psychopathology and preparation for the procedure in the preoperative period (Flores, 2014; Smaidi & Gonçalves, 2016; Keeton et al., 2020).

After medical comorbidities, the second most common cause of mortality among patients with EDs, namely AN, is suicide (Arcelus et al., 2011; Sullivan, 1995). Similarly, rates of

suicidal behaviour are higher in patients with BN and BED compared to the general population (Crow et al., 2009; Forrest et al., 2017). Therefore, suicide rates are typically high in the different disorders and higher in patients with an associated psychiatric disorder (Smith, Zuromski & Dodd, 2018; Mayes et al., 2014).

With that being said, EDs have some of the highest mortality rates among psychiatric disorders, which have a significant personal, interpersonal, social and economic impact (van Hoeken & Hoek, 2020; Weigel, Löwe & Kohlmann, 2019), becoming, in recent decades, a public health problem and, as such, the focus of attention of the scientific community (Smink, van Hoeken & Hoek, 2012).

1.2 Importance of self-report measures for assessing disorders

Given the medical comorbidities associated with these disorders in the weight spectrum (from underweight to obesity), it is imperative to have access to assessment measures adjusted to different populations and cultures and to ensure their effectiveness and appropriateness in detecting EDs (Dahlgren, Wisting & Rø, 2017).

Different authors defend the irreplaceable role of diagnostic interviews in the formal attribution or non-attribution of a diagnosis of a disorder defined by the DSM-V (Dahlgren, Wisting & Rø, 2017; Fairburn & Beglin, 1994; Passi, Bryson & Lock, 2003). The Eating Disorder Examination (EDE) (Fairburn & Cooper, 1993) is recognized as the best validated research-based interview on EDs and is strongly recommended in research and evaluation of treatment outcomes in these disorders (Williamson et al., 1995). However, this procedure, like other interviews, takes longer (an average of 30 to 60 minutes) and is more expensive, given the need for prior training on the part of the researchers. In addition, the participant may omit information because it is a more intrusive methodology (Fairburn & Beglin, 1994).

As an alternative, and in order to overcome the limitations mentioned above, self-report measures have emerged, which are advantageous in terms of cost and time effectiveness. As such, these instruments are easier to distribute and administer and may even be more valid when it comes to “secret”, “shameful” or socially undesirable behaviours, since in an interview, sharing can generate resistance in the individual (Fairburn & Beglin, 1994). In this field, the Eating Disorder Examination Questionnaire (EDE-Q) stands out as one of the most commonly used self-report measures in research and clinical practice (Fairburn & Beglin, 1994).

1.3. Eating Disorder Examination-Questionnaire

Having been developed from the interview (EDE), this instrument is shorter in duration and does not require the presence of or administration by a professional, which can be appealing to young people (Black & Wilson, 1996).

Thus, in 1994, a first version of the EDE-Q appeared with 36 items (Fairburn & Beglin, 1994) which assess the main symptoms of EDs and a series of psychopathologies related to eating. Subsequently, after several revisions, the EDE-Q 6.0 was published in 2008, being considered the current and widely used version. It comprises 28 items that seek to assess the frequency of certain behaviours and attitudes related to eating disorders over the last 28 days. These are scored using a 7-point ordinal scale:

- 0 (no days); 1 (1-5 days); 2 (6-12 days); 3 (13-15 days); 4 (16-22 days); 5 (23-25 days); 6 (every day)
- 0 (not at all); 1-2 (a little); 3-4 (moderately); 5-6 (markedly)

This instrument has a factor structure of four subscales based on responses to 22 items: the *Restraint subscale* comprises 5 items focused on food rules and attempts to avoid eating; the *Eating Concern subscale* comprises 5 items focused on fear of losing control, eating in secret and feeling guilty about eating; the *Shape Concern subscale* comprises 8 items focused on fear of getting fat and dissatisfaction with body shape, among others; the *Weight Concern subscale* comprises 5 items focused on the importance of weight and sensitivity to weight gain. To each subscale is assigned a score and, based on the weighted average of the subscale scores, a total score is obtained, with a higher score indicating the presence of more characteristics associated with EDs (Fairburn & Beglin, 1994; Machado et al., 2014). The remaining 6 items seek to assess the presence and frequency of binge eating, self-induced vomiting, misuse of laxatives, misuse of diuretics and physical exercise aimed at weight control (specific items) (Byrne et al., 2010).

Over the years, studies using Confirmatory Factor Analysis (CFA) and Exploratory Factor Analysis (EFA) have not always supported this four-factor structure. In result, alternative, shorter forms of the EDE-Q have been proposed (Grilo et al., 2013; Kliem et al., 2016; Gideon et al., 2016; Carey et al., 2019; Zohar, 2021).

1.4. Measure Validation

Given the transdiagnostic model mentioned above, there is a need to consider eating disorders as part of a single diagnosis and not as qualitatively distinct diagnostic categories. That said, it is essential that the instruments used to assess eating disorders are in line with the transdiagnostic model and are adapted for different diagnostic categories and samples.

As it is an instrument used to obtain descriptive information on symptoms, validate other assessments and diagnose eating disorders (AN, BN and EDNOS), the EDE-Q must be cross-validated for the different diagnostic categories and for both clinical and non-clinical samples (Grilo, Masheb & Wilson, 2001; Reas, Grilo & Masheb, 2006).

Similarly, the EDE-Q has increasingly been used to assess the effectiveness of ED treatment (Carter & Fairburn, 1998; Walsh et al., 2004), but also of surgical treatment for obesity (bariatric surgery) (de Lucena, de Souza & Alchieri, 2012), and there is evidence that this population undergoes considerable changes in terms of eating-related dimensions (Nunes et al., 2006; de Oliveira & Yoshida, 2009). This, once again, highlights the need for research into validating the measure.

In recent decades, there has been an increase in the population prevalence of eating disorders in males (Hudson et al., 2007; Mitchison et al., 2014; Mitchison & Mond, 2015) and gender minorities (Diemer et al., 2015; Guss et al., 2017; Simone et al., 2020), highlighting the need to clarify gender differences in the factor structure of the EDE-Q, which is widely used. It should also be noted that the evidence (Mitchison & Mond, 2015; Stanford & Lemberg, 2012) points to the existence of key differences between genders in terms of the presentation of the clinical picture. That being said, and despite the fact that several researchers have recently sought to include more diverse populations, there is still a lack of studies assessing eating disorders in these groups (Murray et al., 2017).

In terms of the prevalence of EDs, epidemiological studies (Hoek, 2002, 2006) point to a non-random distribution of these disorders, which are more common in young women, in Western and industrialized societies. That said, EDs are not evenly distributed across cultures and over time. Despite the American Psychological Association Multicultural Guidelines (APA, 2017) being against the application of concepts and theories as universal, the factor structure of the EDE-Q is used in the assessment of EDs in individuals from different cultures, even though it is rarely examined in ethnically diverse populations (Serier, Smith & Yeater, 2018). When the invariance of the measure between ethnicities is tested, the results obtained are often mixed, with some studies verifying it (Belon et al., 2011) and others not (Belon et al.,

2015; Kelly et al., 2012). Once again, there is a need to study the validity of this instrument, in this case, for populations from different countries/cultures.

Considering the above, in order for a survey to be valuable and of use, it is essential to study the psychometric properties, factor structure and measurement invariance in different populations, ensuring its reliability and validity.

Recognizing this, in 2012, Berg and collaborators prepared the first systematic review of the psychometric properties of both the EDE-Q and the EDE, evaluating the psychometric support of the two instruments and providing recommendations for future research. In conclusion, both instruments demonstrated reliability of scores. There was evidence that scores on the EDE and EDE-Q correlate with scores on measures of similar constructs and support for using the instruments to distinguish between cases and non-cases. The authors highlighted the need to broaden the generalizability of the findings.

Although several authors have studied the EDE-Q and its properties in recent years, this information is, to the best of our knowledge, currently not systematized, which justifies the relevance of this systematic review.

1.5. Validity Evidence

When it comes to the psychometric properties of an instrument, the Internal Consistency, for example, indicates whether items in a test/scale, that are intended to measure the same construct, produce consistent scores — in other words, it evaluates the consistency of results across factors within a test. In general, all the items on such measures are supposed to reflect the same underlying construct, so the participants' scores on those items should be related to each other (Tang et al., 2014; El Hajjar, 2018). In turn, Test-Retest Reliability corresponds to the degree to which values are consistent through any repeated test. The most direct way of estimating reliability is to manage or administer the test two times to the identical set of themes and then correlate the two measurements at each time. Construct Validity is the degree in which a test measures a theoretical construct that is intended to be measured (Wuensch, 2012). According to construct validity theory, a construct is implicitly defined by its position in a network of other constructs, that is deduced from theory and based on scientific laws — the "nomological network" (Cronbach & Meehl, 1995). In this context, both convergent and discriminant validity are included: the first concerns the degree in which the scale is statistically associated with preexisting scales that are theoretically related, whilst the second concerns the extent that the scale is uncorrelated with preexisting scales that are theoretically unrelated (Lac,

2016). Criterion Validity is a psychometric property characterized by the extent to which people's scores on a measure are correlated with other variables (known as criteria) that one would expect them to be correlated with. In turn, Concurrent Validity is evidenced if the scale is statistically related to an outcome, with the caveat that both factors are administered cross-sectionally (at the same time) (Lac, 2016). Cross-cultural validity refers to whether measures (in most cases psychological constructs) that were originally generated in a single culture are applicable, meaningful, and thus equivalent in another culture (Matsumoto, 2003). Structural Validity assesses how closely the organizational structure of a set of definitions of psychiatric disorders matches how the disorders present themselves in clinical samples (Jacobs & Krueger, 2015). Finally, the requirement of Measurement Invariance presupposes that the function that relates psychological abilities to test scores should be invariant over groups (Mellenbergh, 1989; Meredith, 1993).

II – Purpose of the Study

This paper aims to systematize the validity evidence of the EDE-Q 6.0 including internal consistency, test-retest reliability, criterion validity, concurrent validity, construct validity, factor structure and measurement invariance across different diagnoses, genders, cultures, clinical and non-clinical samples.

To better understand the aim of this study, the following research question was developed: "What is the currently available validity evidence of the EDE-Q 6.0 for the adult population?"

III – Method

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (Page et al., 2021). The review protocol was preregistered in Prospero (CRD420250655798). The search for articles was carried out in March 2025 in three online platforms, namely PubMed, Scopus and APA PsycNet. The search was limited to journal articles involving human participants, with no restrictions placed on the publication date and language.

The following keywords were applied in the search query: "EDE-Q", "Eating Disorder Examination-Questionnaire", "Reliability", "Validity", "Validation", "Psychometric

Properties", "Psychometric Validation", "Factor Structure", "Structural Validity", "Factor Analysis", "CFA", "EFA", and "Measurement Invariance" (see Supplementary Material - File 1 for the specific search query applied to each database).

The outputs of the search were managed using Rayyan (Ouzzani et al., 2016). Duplicate articles were removed first with an automatic tool and then verified manually. Two researchers (MC and CG) independently conducted the initial screening of articles based on the information in the titles and abstracts, excluding studies irrelevant to the review. To determine eligibility for inclusion, the full texts of the potential studies were independently examined using the Rayyan tool. Any disagreements between the researchers' assessments were resolved through group discussion involving both researchers and one other researcher with doctoral degree in psychology (EC), who reached a consensus. The Mendeley Reference Manager software was used to store, organize, and manage all selected bibliographic references for the final sample.

Based on the PI_O Model, the population, intervention and outcome to be analysed were defined:

- Population: Adults (aged over 16 years old, since the prevalence of eating disorders increases in the transition phase that occurs from adolescence to adulthood) (Schmidt et al., 2016);
- Intervention: Studies that carry out the psychometric validation of the EDE-Q 6.0
- Outcome: At least one of psychometric property, factor structure and/or measurement invariance of the EDE-Q 6.0

3.1. Inclusion and Exclusion Criteria

Thus, the inclusion criteria for this systematic review were as follows: (1) studies involving adults (aged over 16 years old); (2) studies examining at least one of the psychometric properties of the EDE-Q 6.0 (internal consistency, reliability, criterion validity, construct validity and concurrent validity), its factor structure and/or measurement invariance; (3) studies published in journals (journal articles).

On the other hand, the exclusion criteria for this systematic review were as follows: (1) studies with samples under 16 years of age; (2) studies using a version of the EDE-Q other than EDE-Q 6.0; (3) studies that use the EDE-Q 6.0 but do not evaluate at least one of the psychometric properties, factor structure or measurement invariance; (4) Studies that use the EDE-Q 6.0 to validate another instrument; (5) qualitative studies, systematic or scoping reviews, meta-analyses, commentaries, theses, book chapters, case studies, or case series.

3.2. Quality Assessment (risk of bias)

To assess the risk of bias for all the articles that met the inclusion criteria, the COSMIN Risk of Bias Checklist (Mokkink et al., 2024) was applied. The COSMIN Risk of Bias checklist is a tool designed to assess the methodological quality of single studies included in systematic reviews of patient-reported outcome measures (PROMs), and is part of the COSMIN Guideline for conducting systematic reviews of PROMs. The primary goal of assessing the methodological quality in a systematic review is to screen for risk of bias in the included studies. The term "risk of bias" aligns with the Cochrane methodology for systematic reviews of interventions and diagnostic test accuracy studies, as outlined in the *Cochrane Handbook for Systematic Reviews of Interventions* (2019). It reflects the extent to which the results of a study can be considered trustworthy, based on its methodological quality. The checklist includes standards referring to design requirements and preferred statistical methods of studies on measurement properties. For each psychometric measure, a COSMIN Risk of Bias "box" was developed containing all standards needed to assess the quality of a study addressing that specific property. These standards encompass both preferred statistical methods based on Classical Test Theory (CTT) and Item Response Theory (IRT). In total, the checklist consists of ten boxes, covering standards for PROM development and for the nine measurement properties.

Two researchers (MC and CG) independently assessed the risk of bias rating each item from "Inadequate" to "Very good" or "NA" (not applicable). Discrepancies between the researchers' assessments were resolved through discussion and consensus.

The level of agreement among raters was assessed using Cohen's kappa coefficient (κ), ranging from -1 to +1: ≤ 0 (no agreement); 0.01-0.20 (slight/poor agreement); 0.21-0.40 (fair agreement); 0.41-0.60 (moderate agreement); 0.61-0.80 (substantial agreement); 0.81-0.99 (almost perfect agreement), and 1 (perfect agreement) (McHugh, 2012).

IV – Results

A total of 632 articles were identified. After removing duplicates, 399 articles remained and underwent initial screening based on title and abstract. Subsequently, the full texts of 67 articles were retrieved for detailed assessment, resulting in 35 articles that met the outlined inclusion criteria for this systematic review. Fig. 1 presents the PRISMA flow diagram displaying the number of studies included in each phase of the selection process.

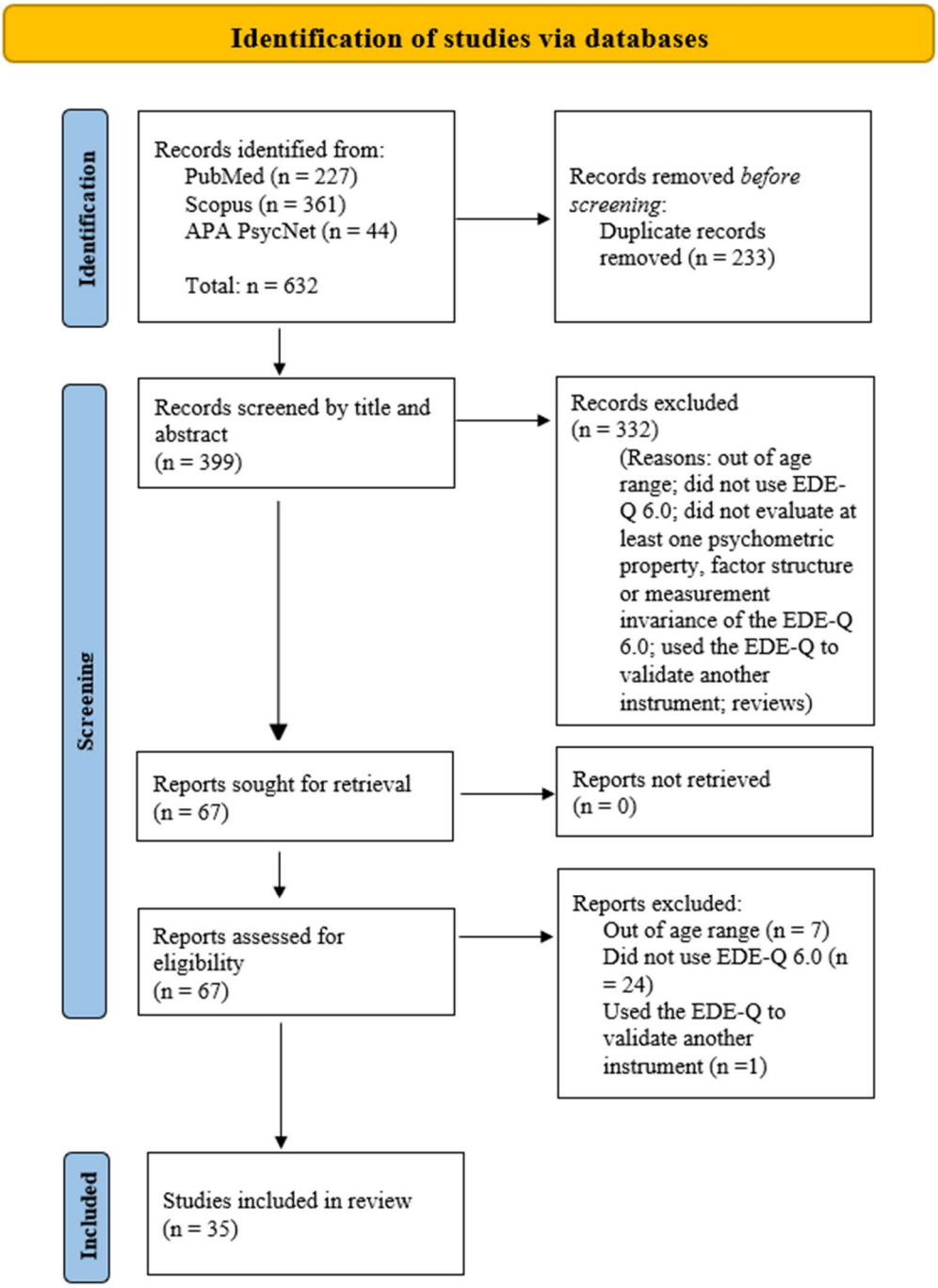

Fig. 1. *Flow diagram of the study selection process.*

4.1 Quality assessment

According to the COSMIN Risk of Bias Checklist assessment, the quality of the included studies varied, but the criteria in each box were mostly assessed with ratings ranging from Adequate to Very Good (see Supplementary Material - File 2). Globally, the agreement index showed an almost perfect concordance between the reviewers, reaching 99.1% ($\kappa = 0.986$, $p < 0.001$).

A detailed analysis based on each COSMIN Risk of Bias Checklist Box revealed the following results:

- Structural Validity: The agreement index showed an almost perfect concordance between the reviewers, reaching 99.1% ($\kappa = 0.985$, $p < 0.001$);
- Internal Consistency: The agreement index showed a perfect concordance between the reviewers, reaching 100% ($\kappa = 1$, $p < 0.001$);
- Cross cultural validity/Measurement Invariance: The agreement index showed an almost perfect concordance between the reviewers, reaching 98.8% ($\kappa = 0.979$, $p < 0.001$);
- Reliability: The agreement index showed a perfect concordance between the reviewers, reaching 100% ($\kappa = 1$, $p < 0.001$);
- Criterion validity: The agreement index showed a perfect concordance between the reviewers, reaching 100% ($\kappa = 1$, $p < 0.001$);
- Construct validity: The agreement index showed an almost perfect concordance between the reviewers, reaching 96.3% ($\kappa = 0.836$, $p < 0.001$).

4.2. Characterization of the studies

The studies included in the present systematic review are summarized in Table 1. In general, the studies included adults from the community ($n = 25$), clinical settings ($n = 6$), or a combination of both ($n = 4$). Most of these studies employed cross-sectional designs ($n = 29$), while only 6 out of the 35 studies had longitudinal designs.

Table 1. *Characterization of the studies included in the systematic review*

Authors (year)	Country	Population	N	Age M (SD)	Sex (% females)	Design
Aldubayan et al. (2023)	Saudi Arabia	Community sample	549	25.12 (4.89)	83.6%	Cross-sectional study
Habashy et al. (2022)	USA	Community sample	1981	19.18 (1.45)	70%	Cross-sectional study
Jenkins et al. (2024)	USA	Clinical sample (AN)	804	25.65 (9.51)	97.0%	Cross-sectional study
Dufour et al. (2025)	Canada	Clinical sample (AN, BN, OSFED, ARFID, BED)	1197	27.9 (10.08)	95%	Cross-sectional study
De Oliveira Júnior et al. (2022)	Brazil	Community sample	1409	27.23 (5.41) 26.68 (5.17)	0%	Longitudinal study
Prnjak & Jukic (2020)	Croatia	Community sample	279	24.61 (5.68)	77.1%	Cross-sectional study
McEntee et al. (2020)	USA	Community sample	1173	19.3 (1.4) 19.6 (1.8) 19.4 (1.5) 19.5 (1.7)	73.5%	Cross-sectional study
Klimek et al. (2020)	USA	Community sample	962	23.68 (3.73)	50.2%	Cross-sectional study
Taib et al. (2021)	Malaysia	Community sample	595	21.9 (1.2)	57%	Longitudinal study

Table 1. *(continued)*

Authors (year)	Country	Population	N	Age M (SD)	Sex (% females)	Design
Otani et al. (2020)	Japan	Clinical (AN, BN) and Non-clinical sample	148	29.6 (13.7) 30.9 (11.3) 30.7 (10.0)	100%	Cross-sectional study
Isomaa et al. (2016)	Finland	Community sample	133	46.1 (9.5)	51.1%	Cross-sectional study
Zohar et al. (2017)	Israel	Community sample	292	33.39 (14.52)	82%	Cross-sectional study
Chan & Leung (2015)	China	Community sample	310	20.75 (1.81)	54.2%	Cross-sectional study
Carey et al. (2019)	UK	Community sample	1075	19.77 (1.73)	79.2%	Cross-sectional study
Rand-Giovannetti et al. (2020)	USA	Community sample	940	20.34 (3.74)	69.9%	Cross-sectional study
Barnes et al. (2012)	UK	Clinical and Non-clinical sample	569		95.8% 91.8%	Cross-sectional study
Friborg et al. (2023)	Norway	Community sample	1076	36.2 (9.5)	100%	Cross-sectional study
Rø et al. (2010)	Norway	Community sample	670	24.8 (6.9)	100%	Longitudinal study

Table 1. *(continued)*

Authors (year)	Country	Population	N	Age M (SD)	Sex (% females)	Design
Allen et al. (2011)	Australia	Clinical (AN, BN, EDNOS) and Non-clinical sample	455	26.02 (9.09) 21.03 (5.85)	100%	Cross-sectional study
Darcy et al. (2013)	USA	Community sample	1637	20.87 (1.66)	59.43%	Cross-sectional study
Contreras-Valdez et al. (2022)	Mexico	Community sample	2092	29 (13) 30 (12) 33 (14) 33 (14)	61.2% 60.7%	Cross-sectional study
Phillips et al. (2018)	USA	Clinical sample (AN)	169	32.0 (13.2) 35.5 (14.0)	100%	Cross-sectional study
Compte et al. (2023)	USA	Community sample	1624	42.1 (15.1) 38.3 (14.4)	34.7%	Cross-sectional study
Knight & Preston (2023)	UK	Community sample	1638	27 (8.44)	36%	Cross-sectional study
Laskowski et al. (2023)	Germany	Clinical sample (AN, BN, BED, EDNOS)	188	32.5 (13.3)	0%	Cross-sectional study
Rose et al. (2013)	USA	Community sample	91	19 (1.16)	48.4%	Longitudinal study
Unikel Santoncini et al. (2017)	Mexico	Community sample	330	19.3 (2.5)	100%	Cross-sectional study

Table 1. *(continued)*

Authors (year)	Country	Population	N	Age M (SD)	Sex (% females)	Design
Rø et al. (2015)	Norway	Clinical (AN, BN, EDNOS) and Non-clinical sample	2465	28.15 (8.46) 31.57 (10.69)	100%	Cross-sectional study
Rica et al. (2022)	Spain	Community sample	796	19.8 (2.8)	0%	Cross-sectional study
Parker et al. (2016)	Australia	Clinical sample (bariatric surgery candidates)	405	43.8 (11.6)	79.3%	Cross-sectional study
Baceviciene et al. (2020)	Lithuania	Community sample	382	24 (6.4)	75.1%	Longitudinal study
McLean et al. (2022)	Australia	Community sample	1271	29.98 (9.97)	82%	Longitudinal study
Sahlan et al. (2022)	USA / Iran	Community sample	1040	19.04 (1.38) 20.11 (1.32)	100%	Cross-sectional study
Parker el al. (2015)	Australia	Clinical sample (bariatric surgery patients)	108	46.0 (12.2)	80.6%	Cross-sectional study
Mahmoodi et al. (2016)	Iran	Community sample	516	23.71 (3.14)	100%	Cross-sectional study

Note. N = number of participants in each study; M = Mean; SD = Standard deviation; AN = Anorexia Nervosa; BN = Bulimia Nervosa; OSFED = Other Specified Feeding or Eating Disorders; ARFID = Avoidant/Restrictive Food Intake Disorder; BED = Binge Eating Disorder; EDNOS = Eating Disorders Not Otherwise Specified

4.3. Internal Consistency

The studies included in the present systematic review that examined the internal consistency of the EDE-Q 6.0 are summarized in Table 2.

Twelve out of the thirty-five studies examined the internal consistency. These studies included the following samples: 310 university students (Chan & Leung, 2015), 228 adults with ED, a community sample of 211 adults (Allen et al., 2011), a community sample of 279 adults (Prnjak & Jukic, 2021), 169 women with AN (Phillips et al., 2018), 330 mexican female students (Unikel Santoncini et al., 2018), 133 Finnish adults (Isomaa et al., 2016), 670 adult women (Rø et al., 2010), 91 adults (Rose et al., 2013), 382 undergraduate and graduate students (Baceviciene et al., 2020), a control sample of 1791 women, 620 adult women with an ED (Rø et al., 2015) , 278 vegeterians, 580 vegans, 413 omnivores (McLean et al., 2022) and 516 female college students from Iran (Mahmoodi et al., 2016).

The global scale demonstrated excellent internal consistency (Streiner, 2003) in every study, with alphas ranging between 0.83–0.97. In turn, the four subscales demonstrated acceptable to excellent internal consistency (Streiner, 2003) in ten of the twelve studies with alphas ranging from 0.70 to 0.97. The two exceptions were an alpha of 0.63 in a sample of university participants (Chan & Leung, 2015) and an alpha of 0.58 in the Weight Concern subscale in a sample of female college students from Iran (Mahmoodi et al., 2016).

4.4. Test-Retest Reliability

The studies included in the present systematic review that examined the Test-retest Reliability of the EDE-Q 6.0 are summarized in Table 2.

The test–retest reliability scores over a 7-day interval have been examined by two studies (Rø et al., 2010; Rose et al., 2013) using Spearman's correlation coefficients. These studies included the following samples: 670 adult women; 91 adults (47 men and 44 women). For the global scale, the Spearman's correlation coefficients ranged from 0.89 to 0.93, which indicates a strong to very strong test-retest reliability (Ridder et al., 2021). For the four subscales, scores ranged from 0.68 to 0.93 (moderate to very strong) (Ridder et al., 2021). The Restraint subscale revealed, respectively, a very strong and strong test-retest reliability in the samples of 670 and 44 adult women (0.90 and 0.81) (Rø et al., 2010; Rose et al., 2013), and a moderate value in the samples of 91 adults (0.79) and 47 men (0.76) (Rose et al., 2013). For the Eating Concern scale, the Spearman's correlation coefficients were strong in the samples of 670 adult women

(0.82), 91 adults (0.80) and 44 women (0.83) (Rø et al., 2010; Rose et al., 2013), but moderate for the 47 men (0.68) (Rose et al., 2013). With scores ranging from 0.86 to 0.93, the Shape Concern subscale revealed a strong to very strong test-retest reliability in all samples. For the Weight Concern subscale, test-retest reliability was very strong in the sample of 44 women (0.91) (Rø et al., 2010; Rose et al., 2013), and moderate in the samples of 670 adult women (0.86), 47 men (0.85) and 91 adults (0.75) (Rø et al., 2010; Rose et al., 2013).

Two other groups of researchers examined the test–retest reliability scores over a 14-day interval using Intraclass Correlation Coefficients (ICC) in samples of 382 undergraduate and graduate students (Baceviciene et al., 2020), 278 vegeterians, 580 vegans and 413 omnivores (McLean et al., 2022). In the first study (Baceviciene et al., 2020), with scores ranging from 0.66 to 0.91, test-retest reliability was moderate to excellent for the global scale and its four subscales. In the second study (McLean et al., 2022), the scores ranged from 0.36 to 0.61 revealing a poor to moderate test-retest reliability (Koo & Li, 2016).

4.5. Concurrent Validity

The study included in the present systematic review that examined the Concurrent Validity of the EDE-Q 6.0 is summarized in Table 3.

One out of the thirty-five studies examined the Concurrent Validity of the EDE-Q in a sample of 382 Lithuanian undergraduate and graduate students (Baceviciene et al., 2020), by testing the associations with tools of similar constructs. The analysis demonstrated these associations in the expected direction between the LT-EDE-Q 6.0 scores and the Lithuanian version of the Multidimensional Body-Self Relations Questionnaire Appearance Scales (LT-MBSRQ-AS), the Lithuanian version of the World Health Organization Quality of Life-BREF Questionnaire (LT-WHOQOL-BREF), the Lithuanian version of M. Rosenberg's Self-Esteem Scale (RSES), and BMI measures which indicate adequate concurrent validity.

4.6. Construct Validity

The studies included in the present systematic review that examined the Construct Validity of the EDE-Q 6.0 (n = 4) are summarized in Table 3.

Three out of the thirty-five studies examined the Convergent Validity of the EDE-Q in the following samples: community sample of 279 adults (Prnjak & Jukic, 2021), 1271 adults from three different dietary groups (McLean et al., 2022) and 516 female college students from

Iran (Mahmoodi et al., 2016). In the first study (Prnjak & Jukic, 2021), the results showed a high correlation (r = 0.63) between the EDE-Q and EAT-26 which supports the convergent validity of the EDE-Q. For the second study (McLean et al., 2022), convergent correlations between EDE-Q subscale scores and the EDE-QS summary score ranged from strong to very strong across dietary groups. In the third study (Mahmoodi et al., 2016), the EDE-Q and all its subscale showed moderate to strong positive correlation with the BES scores.

Three groups of researchers examined the Discriminative Validity of the EDE-Q in the following samples: 382 Lithuanian undergraduate and graduate students (Baceviciene et al., 2020), 1271 adults from three different dietary groups (McLean et al., 2022) and 516 female college students from Iran (Mahmoodi et al., 2016), using Pearson Correlations Coefficients. In the first study (Baceviciene et al., 2020), results ranged from 0.31 (restraint subscale) to 0.43 (shape concern subscale), indicating a weak to moderate correlation. In the second study (McLean et al., 2022), the correlations between the EDE-Q global and subscales scores and DASS subscales were weak to moderate across dietary groups. In the third study (Mahmoodi et al., 2016), the EDE-Q and its subscale successfully distinguished underweight, students with healthy weight, and overweight participants, demonstrating acceptable discriminative validity.

4.7. Criterion Validity

The study included in the present systematic review that examined the Criterion Validity of the EDE-Q 6.0 is summarized in Table 3.

One out of the thirty-five studies examined the Criterion Validity of the EDE-Q in samples of 1791 female controls and 620 adult women with ED (Rø et al., 2015). ROC analysis demonstrated an AUC of 0.93 (95% CI=0.91–0.94), which strongly supported the criterion validity of the EDE-Q. This indicated there was a 93% probability that a randomly selected ED case would obtain a higher EDE-Q score than a non-clinical control.

4.8. Structural Validity

The studies included in the present systematic review that examined the Structural Validity of the EDE-Q 6.0 are summarized in Table 4 and 5. Twenty-nine studies examined the Structural Validity of the EDE-Q. In ten studies, only CFA was used. In five studies, only EFA was used. In fourteen studies, both CFA and EFA were used.

CFA results indicated an inappropriate / worse fit of the four-factors or that it failed to converge to the original model of EDE-Q in twenty-three out of the twenty-four studies, meaning the imposed structure did not represent or describe the data well or, at least, as well as the other models. However, in a community sample of 219 adults (Prnjak & Jukic, 2021), Restraint and Weight Concern factors were shown to have a satisfactory level of fit indices. Moreover, in this study, results showed lower suitability of the one-, two-, three-factor, and brief 8-item models when compared to original subscale factors.

EFA results showed that the factor solutions did not follow Fairburn's original factor structure in any of the samples.

The analysis revealed the retention of one factor in the following samples: 1197 ED patients (Dufour et al., 2025) and 1409 Brazilian cisgender gay and bisexual adult men (de Oliveira Junior et al., 2023).

A group of researchers (Darcy et al., 2013) found that, in a sample of 229 men, the data was best described by a two-factor model with 19 items.

The analysis revealed the retention of three factors in the following samples:

- 549 Saudi adults (shape and weight Cconcern, restraint, and eating concern with 14 items) (Aldubayan et al., 2023);
- 1981 US undergraduate students (dietary restraint, preoccupation and eating concern, and shape/weight overvaluation with 10 items) (Habashy et al., 2023);
- 148 ED patients (restraint and eating concern, dissatisfaction with shape and weight, and self-esteem based on shape and weight) (Otani et al., 2021);
- 292 community volunteers (20 items) (Zohar et al., 2017);
- 851 female students (with 18 items) (Carey et al., 2019);
- 224 male students (with 16 items) (Carey et al., 2019);
- 569 adults (shape/weight concern, eating concern, and restraint) (Barnes et al., 2012);
- 1076 Norwegian women (Friborg et al., 2013);
- 432 male competitive athletes (21 items) (Darcy et al., 2013);
- 544 female competitive athletes (19 items) (Darcy et al., 2013);
- 429 women (18 items) (Darcy et al., 2013);
- 1638 adults (14 items) (Knight et al., 2023);
- 278 vegeterians (eating concern, restraint, and weight and shape concern with 16 items) (McLean et al., 2022);
- 580 vegans (eating concern, restraint, and weight and shape concern with 15 items) (McLean et al., 2022);

- 413 omnivores (eating concern, restraint, and weight and shape concern with 16 items) (McLean et al., 2022).

EFA results revealed a four-factor solution in the following samples:

- 595 Malaysian university students (with 22 items) (Taib et al., 2021);
- 405 adults seeking LAGB (dietary restraint, eating concern, shape/ weight overvaluation, and appearance concern with 13 items) (Parker et al., 2016);
- 279 adults from the community (Prnjak & Jukic, 2021);
- 169 women with AN (Phillips et al., 2018);
- 108 adults who had undergone LAGB (Parker et al., 2015).

One of the studies even suggested a five-factor solution, retaining 17 of the original 22 items, in a sample of 188 German adult men with ED (Laskowski et al., 2023).

4.9. Measurement Invariance

The studies included in the present systematic review that examined the Measurement Invariance of the EDE-Q 6.0 are summarized in Table 6. Only three out of the thirty-five studies examined the Measurement Invariance of the EDE-Q. In a sample of adults with ED and adults from the community (Allen et al., 2011) the unconstrained and constrained models were significantly different, suggesting that measurement parameters were not equivalent across the eating disorder and community groups. In the second study (Baceviciene et al., 2020), invariance analyses across gender groups in a sample of undergraduate and graduate students revealed a statistical difference between unconstrained and fully constrained models, suggesting that measurement parameters were also not equivalent. In the third study (Sahlan et al., 2022), tests revealed the EDE-Q as non-invariant across ethnic/racial groups in samples of Iranian and US college women.

Table 2. *Internal Consistency and Test-retest Reliability of the EDE-Q 6.0 (n = 12)*

		Internal Consistency (Cronbach's Alpha)					**Reliability (Test-Retest)**											
							Spearman's correlation coefficients							Intraclass Correlation Coefficient					
Authors (year)	**Samples**	Global	R	EC	SC	WC	Sample	Global	R	EC	SC	WC	*p*	Sample	Global	*R*	*EC*	*SC*	*WC*
Chan & Leung (2015)	310 university participants	.93	ranging from .63 to .88															
Allen et al. (2011)	228 adults with ED		.79	.75	.88	.80												
	211 adults from a community sample		.81	.86	.93	.89												
Prnjak & Jukic (2021)	279 participants from a community sample	.93	.79	.84	.89	.84												
	215 women from a community sample	.93																
Philips et al. (2018)	169 women with AN	.96	.89	.81	.90	.84												
Unikel Santoncini et al. (2018)	330 mexican female students	.97	.84	.81	.92	.86												
Isomaa et al. (2016)	133 Finnish adults		.78	.76	.89	.81												
Rø et al. (2010)	670 adult women	.94	.75	.78	.90	.81	n = 159	.93	.90	.82	.91	.86	< .01						
Rose et al. (2013)	91 adults	Time 1: .89; Time 2: .90	Time 1: .73; Time 2: .83	Time 1: .79; Time 2: .86	Time 1: .87; Time 2: .92	Time 1: .82; Time 2: .87	n = 91	.92	.79	.80	.91	.75							
	47 men	Time 1: .83; Time 2: .87	Time 1: .74; Time 2: .86	Time 1: .73; Time 2: .77	Time 1: .86; Time 2: .89	Time 1: .77; Time 2: .82	n = 47	.89	.76	.68	.93	.85							
	44 women	Time 1: .91; Time 2: .92	Time 1: .75; Time 2: .81	Time 1: .79; Time 2: .89	Time 1: .87; Time 2: .93	Time 1: .83; Time 2: .89	n = 44	.90	.81	.83	.86	.91							
Baceviciene et al. (2020)	382 undergraduate and graduate students	.94	.83	.75	.88	.83								n = 382	.90	.66	.84	.91	.90
Rø et al. (2015)	1791 female controls	.94																	
	620 adult women with an ED	.93																	

Table 2. *(continued)*

		Internal Consistency (Cronbach's Alpha)					Reliability (Test-Retest)											
							Spearman's correlation coefficients							Intraclass Correlation Coefficient				
Authors (year)	**Samples**	Global	R	EC	SC	WC	Sample	Global	R	EC	SC	WC	*p*	Sample	Global	*R*	*EC*	*SC*	*WC*
McLean et al. (2022)	278 Vegeterians	.96	.84	.87	.93	.86								n = 278	.41	.46	.55	.55	.47
	580 Vegans	.96	.86	.86	.93	.86								n = 580	.43	.49	.52	.61	.50
	413 Omnivores	.94	.85	.85	.90	.84								n = 413	.36	.48	.51	.49	.43
Mahmoodi et al. (2016)	516 female college students from Iran	.91	.78	.73	.81	.58													

Note. N = number of participants in each study; R = Restraint; EC = Eating Concern; SC = Shape Concern; WC = Weight Concern; *p* = p-value; ED = Eating Disorder; AN = Anorexia Nervosa

Table 3. *Validities of the EDE-Q 6.0 (n = 5)*

		Criterion Validity				Construct Validity						
							Discriminative or known-groups validity (Pearson correlations coefficients)					
Authors (year)	**Samples**	sample	AUC	sensitivity	specificity	Convergent Validity	Global	R	EC	SC	WC	**Concurrent Validity**
Prnjak & Jukic (2021)	279 participants from a community sample					Good convergent validity						
Baceviciene et al. (2020)	382 undergraduate and graduate students							.31	.33	.43	.40	Adequate concurrent validity
Rø et al. (2015)	1791 female controls 620 adult women with ED	n = 2465	.93	.86	.86							
McLean et al. (2022)	278 Vegeterians 580 vegans 413 Omnivores					Convergent correlations ranged from strong to very strong across dietary groups	Correlations were weak to moderate in strength across dietary groups.					
Mahmoodi et al. (2016)	516 female college students from Iran					The EDE-Q and all its subscale showed moderate to strong positive convergent correlation	The EDE-Q and its subscale successfully distinguished underweight, students with healthy weight, and overweight participants.					

Note. N = number of participants in each study; R = Restraint; EC = Eating Concern; SC = Shape Concern; WC = Weight Concern; AUC = Area Under the ROC Curve; ED = Eating Disorder

Table 4. *Structural Validity (CFA) of the EDE-Q 6.0 (n = 24)*

		Structural Validity											
		Confirmatory Factor Analysis (CFA)											
Authors (year)	**Samples**	Sample	χ2 (df)	χ2 / df	*p*	RMSEA	CFI	TLI	SRMR	NFI	IFI	GFI	NNFI
Aldubayan et al. (2023)	549 Saudi adults (total sample)	n = 549				.133	.732	.694					
	120 adults from Riyadh	n = 120				.140	.727	.689					
	137 adults from Makkah	n = 137				.151	.635	.584					
	124 adults from Madinah	n = 124				.139	.762	.730					
	132 adults from AlSharqiya	n = 132				.128	.734	.697					
Jenkins et al. (2024)	804 women with an ED	n = 804	Did not converge										
Dufour et al. (2025)	1197 people with an ED	n = 1197	3282.26 (203)			.113	.754	.720					
McEntee et al. (2021)	177 Latino men	n = 177	867.77 (202)			.14	.74	.70	.096				
	133 non-Latino White men	n = 133	707.75 (202)			.14	.74	.69	.123				
	554 Latina women	n = 554	1923.38 (202)			.12	.81	.78	.074				
	309 non-Latina White women	n = 309	Did not converge										
Otani et al. (2021)	148 ED patients	n = 148				.166	.691						
Klimek et al. (2021)	962 cisgender sexual minority men and women	problematic model as indicated by a review of factor correlations, factor loadings, and variances											
Zohar et al. (2017)	292 community volunteers	n = 292	1306.8 (170)		< .001	.15	.77	.66	.32				
Chan & Leung (2015)	310 university participants	n = 310	1163.98 (202)	5.76	< .001	.12	.78	.75		.75	.78		
Carey et al. (2019)	851 female students	n = 851	Invalid due to Heywood cases										
	224 male students	n = 224	Invalid due to Heywood cases										
Rand-Giovannetti et al. (2020)	940 undergraduate psychology students	n = 940	1653.581 (203)			.087	.937	.928					
Barnes et al. (2012)	569 adults	n = 569	1866.48 (404)	4.62		.08	.86						

Table 4. *(continued)*

		Structural Validity										
		Confirmatory Factor Analysis (CFA)										
Authors (year)	**Samples**	Sample	χ2 (df)	χ2 / df	*p*	RMSEA	CFI	TLI	SRMR	NFI	IFI	GFI	NNFI
Friborg et al. (2013)	1076 Norwegian women	n = 1076	4382 (202)			.090							.981
Darcy et al. (2013)	432 male competitive athletes	n = 432	750.77 (202)		< .001	.0793	.9627	.9574	.1081				
	544 female competitive athletes	n = 544	1157.74 (202)		< .001	.0933	.9794	.9764	.0670				
	229 men	n = 229	529.02 (202)		< .001	.0842	.9694	.9650	.0861				
	429 women	n = 429	1185.82 (202)		< .001	.1067	.9561	.9497	.0711				
Contreras-Valdez et al. (2022)	684 mexican women	n = 684	1507.23 (202)	7.46	< .001	.10	.76		.09				.72
	433 mexican men	n = 433	758.13 (202)	3.75	< .001	.08	.72		.09				.68
	591 mexican women	n = 591	1434.45 (202)	7.11	< .001	.10	.74		.11				.70
	382 mexican men	n = 382	728.77 (202)	3.61	< .001	.08	.72		.10				.68
Compte et al. (2023)	1060 cisgender gay men	n = 1060	results revealed a nonpositive definite matrix solution, suggesting that this model was unacceptable										
	528 cisgender lesbian women	n = 528	results revealed a nonpositive definite matrix solution, suggesting that this model was unacceptable										
Knight & Preston (2023)	1638 adults	n = 420	2407.554 (183)		< .001	.12	.791	.762	.097				
Rica et al. (2022)	796 spanish male university students	n = 796	results revealed a nonpositive definite matrix solution, suggesting that this model was unacceptable										
Parker et al. (2016)	405 adults seeking laparoscopic adjustable gastric banding	n = 405	results revealed a nonpositive definite matrix solution, suggesting that this model was unacceptable										
Allen et al. (2011)	228 adults with ED	n = 228	1659.83 (203)		< .001	.18	.84			.83		.61	
	211 adults from a community sample	n = 211	1195.27 (203)		< .001	.15	.93			.92		.66	
Prnjak & Jukic (2021)	279 participants from a community sample	n = 274	898.51 (182)		< .001	.120	.808	.779	.779				
	215 female participants from a community sample	n = 210	789.45 (182)		< .001	.126	.797	.766	.766				

Table 4. *(continued)*

Authors (year)	Samples	Structural Validity											
		Confirmatory Factor Analysis (CFA)											
		Sample	χ2 (df)	χ2 / df	*p*	RMSEA	CFI	TLI	SRMR	NFI	IFI	GFI	NNFI
Unikel Santoncini et al. (2018)	330 mexican female students	n = 330	1223 (203)		< .001	.13	.82	.80	.07				
Baceviciene et al. (2020)	382 undergraduate and graduate students	n = 382		9.730	<.0001	.151	.710	.670				.660	
McLean et al. (2022)	278 Vegeterians	n = 278	1138.34 (202)		<.001	.13	.82	.80		.80		.72	
	580 vegans	n = 580	1857.73 (202)		<.001	.12	.85	.83		.84		.76	
	413 Omnivores	n = 413	1400.46 (202)		<.001	.12	.83	.80		.81		.76	
Parker et al. (2015)	108 adults who had undergone Laparoscopic Adjustable Gastric Banding (LAGB)	n = 108	The original four-factor solution could not be estimated due to a covariance matrix that was not positive definite										

Note. N = number of participants in each study; R = Restraint; EC = Eating Concern; SC = Shape Concern; WC = Weight Concern; ED = Eating Disorder; χ2 = Chi-square Statistic; df = Degrees of freedom; *p* = p-value; RMSEA = Root-Mean-Square Error of Approximation; CFI= Comparative Fit Index; TLI =Tucker Lewis Index; SRMR = Standardized Root-Mean-Square Residual; NFI = Normed Fit Index; NNFI = Non Normed Fit Index; IFI = Incremental Fit Index; GFI = Goodness of Fit Index;

Table 5. *Structural Validity (EFA) of the EDE-Q 6.0 (n = 19)*

Authors (year)	Samples	Structural Validity					
		EFA					
		sample	KMO	Bartlett's test		Number of items and factor structure	% of the total variance explained
				χ2 (df)	*p*		
Aldubayan et al. (2023)	549 Saudi adults	n = 275	.907	4083.5 (120)	< .001	14 items ; 3 factors: Shape and Weight Concern, Restraint and Eating Concern	
Habashy et al. (2023)	1981 US undergraduate students	n = 990	.94	19171.5 (120)	< .001	10 items ; 3 factors: Dietary Restraint, Preoccupation and Eating Concern and Shape/Weight Overvaluation	63.1 % of the total variance explained
Dufour et al. (2025)	1197 people with an ED	n = 1197				1 factor	
de Oliveira Junior et al. (2023)	1409 Brazilian cisgender gay and bisexual adult men	n = 704	.91	9727.33 (231)	< .001	22 items ; 1 factor	40.2% of the total variance explained
Otani et al. (2021)	148 ED patients	n = 148				3 factors: Restraint and Eating Concern, Dissatisfaction with Shape and Weight, and Self-Esteem Based on Shape and Weight	65.63% of the total variance explained
Taib et al. (2021)	595 Malaysian university students	n = 595				22 items ; 4 factors: Restraint, Shape/Weight Concerns, Eating Concerns, and Shape/Weight Overvaluation	63.8% of the total variance explained
Zohar et al. (2017)	292 community volunteers	n = 292				20 items ; 3 factors: Restraint, Weight/Shape Concern, and Eating Concern	65.77% of the total variance explained
Chan & Leung (2015)	310 university participants	n = 310				1 factor	

Table 5. *(continued)*

Authors (year)	Samples	EFA					
		sample	KMO	Bartlett's test		Number of items and factor structure	% of the total variance explained
				$\chi2$ (df)	*p*		
Carey et al. (2019)	851 female students	n = 851	.92		< .001	18 items ; 3 factors: Shape and Weight Concerns, Preoccupation and Eating Concern, and Restriction	66.26% of the total variance explained
	224 male students	n = 224	.91		< .001	16 items; 3 factors: Shape and Weight Concerns, Preoccupation and Eating Concern, and Restriction	67.38% of the total variance explained
Barnes et al. (2012)	569 adults	n = 569				3 factors: Shape/Weight Concerns, Eating Concern, and Restraint	50.3% of the total variance explained
Friborg et al. (2013)	1076 Norwegian women	n = 1076				4 factors	68.1% of the total variance explained
Darcy et al. (2013)	432 male competitive athletes	n = 432				21 items ; 3 factors	59.53% of the total variance explained
	544 female competitive athletes	n = 544				19 items ; 3 factors	69.96% of the total variance explained
	229 men	n = 229				19 items ; 2 factors	60.06% of the total variance explained
	429 women	n = 429				18 items ; 3 factors	69.36% of the total variance explained
Knight & Preston (2023)	1638 adults	n = 420	.92	5564.052 (120)	< .001	14 items ; 3 factors: Shape and Weight Concern, Preoccupation and Eating Concern, and Restriction	69% of the total variance explained
Laskowski et al. (2023)	188 German adult men with ED	n = 188	.80			17 items ; 5 factors: Restraint, Body Dissatisfaction,Weight Concern, Preoccupation, and Importance	68% of the total variance explained
Parker et al. (2016)	405 adults seeking laparoscopic adjustable gastric banding	n = 405	.80	3325.71 (231)	< .001	13 items ; 4 factors: Dietary Restraint, Eating Concern, Shape/Weight Overvaluation and Appearance Concern	56.7% of the total variance explained
Prnjak & Jukic (2021)	279 participants from a community sample	n = 279	.92	4379.38 (231)	< .001	4 factors	
Philips et al. (2018)	169 women with AN	n = 169	.923			4 factors	64.2% of the total variance explained
McLean et al. (2022)	278 Vegeterians	n = 278	.91		<.001	16 items ; 3 factors: Eating Concern, Restraint, and Weight and Shape Concern	67.76% of the total variance explained
	580 vegans	n = 580	.93		<.001	15 items ; 3 factors: Eating Concern, Restraint, and Weight and Shape Concern	69.05% of the total variance explained
	413 Omnivores	n = 413	.90		<.001	16 items ; 3 factors: Eating Concern, Restraint, and Weight and Shape Concern	64.38% of the total variance explained
Parker et al. (2015)	108 adults who had undergone Laparoscopic Adjustable Gastric Banding (LAGB)	n = 108	.83	1371.27 (231)	<.001	14 items; 4 factors: Dietary Restraint, Eating Concern, Shape/Weight Overvaluation, and Appearance Concern	64.8% of the total variance explained

Note. N = number of participants in each study; $\chi2$ = Chi-square Statistic; df = Degrees of freedom; *p* = p-value; KMO = Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy; ED = Eating Disorder; AN = Anorexia Nervosa

Table 6. *Measurement Invariance of the EDE-Q 6.0 (n = 3)*

		Measurement Invariance								
		Unconstrained model								
Authors (year)	**Samples**	sample	χ2 (df)	*p*	*χ2 /df*	GFI	AGFI	TLI	CFI	RMSEA
Allen et al. (2011)	228 adults with ED	n = 228	2298.33 (406)	< .001						
	211 adults from a community sample	n = 211								
Baceviciene et al. (2020)	382 undergraduate and graduate students	n = 382		<.0001	5.782	.636	.546	.645	.688	.112
		n = 95 (men)		<.0001	3.539	.611	.516	.652	.694	.164
		n = 287 (women)		<.0001	8.021	.645	.557	.643	.686	.157
Sahlan et al. (2022)	709 college women from the US	n = 709	Tests showed the EDE-Q as non-invariant between Iranian and US women							
	331 college women from Iran	n = 331								

Table 6. (*continued*)

		Measurement Invariance										
		Constrained model									Difference between Unconstrained and Constrained models	
Authors (year)	**Samples**	sample	χ2 (df)	*p*	*χ2 /df*	GFI	AGFI	TLI	CFI	RMSEA	χ2 (df)	*p*
Allen et al. (2011)	228 adults with ED	n = 228	3109.57 (406)	< .001							811.24 (50)	< .001
	211 adults from a community sample	n = 211										
Baceviciene et al. (2020)	382 undergraduate and graduate students	n = 382		<.0001	5.576	.622	.560	.661	.681	.110		
Sahlan et al. (2022)	709 college women from the US	n = 709	Tests showed the EDE-Q as non-invariant between Iranian and US women									
	331 college women from Iran	n = 331										

Note. N = number of participants in each study; χ2 = Chi-square Statistic; df = Degrees of freedom; *p* = p-value; RMSEA = Root-Mean-Square Error of Approximation; CFI= Comparative Fit Index; TLI =Tucker Lewis Index; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; ED = Eating Disorder;

V – Discussion

This systematic review provides a thorough analysis of the psychometric properties, structural validity and measurement invariance of the EDE-Q 6.0 in adults.

In terms of reliability and validity, our findings generally supported the results presented in the work carried out by Berg et al. (2012), revealing good to excellent internal consistency, strong test-retest reliability scores over a 7 and 14-day interval and good concurrent, construct and criterion validity.

The structural validity of the EDE-Q 6.0, however, has appeared to be a critical issue - most CFA results did not support the imposed structure of the original model and EFA results showed that the factor solutions did not follow Fairburn's original factor structure in any of the samples. Using the original structure in samples where it reveals an inappropriate / worse fit can lead to misinterpretations of clinical profiles, incorrect classifications and poorly targeted interventions. This raises the importance of testing the structural validity and the possible need to revise or reformulate the factor structure, as has been seen over the years with proposals for alternative and/or short form structures.

In addition, the measurement invariance remains little explored, with initial results suggesting a lack of equivalence between clinical and non-clinical samples, gender and ethnic/racial groups. It should also be noted that numerous studies dedicated to the validation of the EDE-Q in samples from different countries, genders and clinical groups did not study the measurement invariance of this instrument. There was also no evidence of the study of the measurement longitudinal invariance, which is a huge limitation, since without longitudinal invariance, scores should not be compared in repeated measurements (e.g. pre, post, follow up).

Lastly, it is necessary to point out the relatively small number of articles (n = 35) dedicated to studying the psychometric properties, factor structure and invariance of the EDE-Q 6.0 in adults, considering the breadth of its use.

5.1. Strengths, Limitations and Future Directions

The current study exhibits significant strengths. First and foremost, it aimed to fill a gap in the systematization of the available data concerning the validity evidence of the EDE-Q 6.0, which justifies the relevance of this systematic review. By encompassing the work carried out by numerous research groups (that has now been collected, organized and critically analyzed), the systematic review offers an integrated view of the evolution in this field of knowledge,

making it possible to identify consensus, gaps, tendencies and contrasts in the current literature. By following the PRISMA protocol, in which the process of selecting studies (first, based on the title/abstract and then based on the full text) and assessing their quality is carried out by more than one researcher (in this case 2 researchers, with a third in cases of disagreement), the risk of bias is minimized. Another key point was the lack of restrictions on the date and language of publication of the articles, allowing the inclusion of a greater number of studies.

As expected, this review also presents some limitations. Despite efforts to do so, the risk of bias cannot be completely eliminated. Furthermore, although the aim of this paper was to gather as much knowledge as possible to answer the research question, relevant studies may have been excluded because they were not available in the databases used. In retrospect, in terms of methodology, and given the specific nature of this systematic review, it could have been conducted based on the PRISMA-COSMIN — Guideline for reporting systematic reviews of outcome measurement instruments (OMIs) (Elsman et al., 2024) —, which is an extension of the PRISMA (2020) guidelines.

Based on the results, future studies should broaden the sources of validity evidence to be explored (e.g. response processes and consequences of testing). There is a need for additional robust research into the factor structure and measurement invariance, namely longitudinal invariance, of the EDE-Q 6.0. Given that the processes of adapting the instrument have not always been carried out rigorously, it is suggested that future studies make use of rigorous guidelines such as the ITC Guidelines for Translating and Adapting Tests (2017). It is also important to focus on analyses with more diverse populations (not only women and/or western samples), considering the increasing prevalence of these disorders in men and gender minorities, as well as their non-uniform distribution in different cultures. That being said, future studies should also consider exploring alternative models and possibly proposing adaptations of the instrument that could more accurately reflect the experience of symptoms in different populations.

VI – Conclusions

Unfortunately, Eating Disorders have been and continue to represent a serious health condition, considering the clinical suffering and impairment of physical health and psychosocial functioning. Thus, it is imperative to have access to assessment measures adjusted to different populations and cultures and to ensure their effectiveness and appropriateness in detecting EDs.

This highlights the need for self-report instruments, such as the EDE-Q, to be adapted for different samples and in line with the Transdiagnostic Model.

Overall, the EDE-Q 6.0 revealed good psychometric properties, but a poor fit of the four-factor structure and preliminary results suggesting a lack of equivalence between clinical and non-clinical samples, gender and ethnic/racial groups. In addition to systematizing current evidence, these findings also set the context for future investigation with the purpose of increasing and diversifying this field of knowledge.

References

Aldubayan, K., Ghafouri, K., Mutwalli, H., Kutbi, H. A., & Mumena, W. A. (2023). Validity and Consistency of the Arabic Version of the Eating Disorder Examination Questionnaire (EDE-Q) among Saudi Adults. *Healthcare (Basel, Switzerland)*, *11*(7), 1052. https://doi.org/10.3390/healthcare11071052

Allen, K. L., Byrne, S. M., Lampard, A., Watson, H., & Fursland, A. (2011). Confirmatory factor analysis of the Eating Disorder Examination-Questionnaire (EDE-Q). *Eating behaviors*, *12*(2), 143–151. https://doi.org/10.1016/j.eatbeh.2011.01.005

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

American Psychological Association. (2017). Multicultural Guidelines: An Ecological Approach to Context, Identity, and Intersectionality. Retrieved from: http://www.apa.org/about/policy/multicultural-guidelines.pdf

Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Archives of general psychiatry*, *68*(7), 724–731. https://doi.org/10.1001/archgenpsychiatry.2011.74

Baceviciene, M., Balciuniene, V., & Jankauskiene, R. (2020). Validation of the Lithuanian version of the Eating Disorder Examination Questionnaire 6.0 in a student sample. *Brain and behavior*, *10*(3), e01555. https://doi.org/10.1002/brb3.1555

Barnes, J., Prescott, T., & Muncer, S. (2012). Confirmatory factor analysis for the Eating Disorder Examination Questionnaire: Evidence supporting a three-factor model. *Eating behaviors*, *13*(4), 379–381. https://doi.org/10.1016/j.eatbeh.2012.05.001

Becker, A. E., Thomas, J. J., Bainivualiku, A., Richards, L., Navara, K., Roberts, A. L., Gilman, S. E., Striegel-Moore, R. H., & HEALTHY Fiji Study Group (2010). Validity and reliability of a Fijian translation and adaptation of the Eating Disorder Examination

Questionnaire. *The International journal of eating disorders*, *43*(2), 171–178. https://doi.org/10.1002/eat.20675

Belon, K. E., McLaughlin, E. A., Smith, J. E., Bryan, A. D., Witkiewitz, K., Lash, D. N., & Winn, J. L. (2015). Testing the measurement invariance of the eating disorder inventory in nonclinical samples of Hispanic and Caucasian Women. *International Journal of Eating Disorders, 48*, 262–270. https://doi.org/10.1002/eat.22286

Belon, K. E., Smith, J. E., Bryan, A. D., Lash, D. N., Winn, J. L., & Gianini, L. M. (2011). Measurement invariance of the Eating Attitudes Test-26 in Caucasian and Hispanic women. *Eating behaviors*, *12*(4), 317–320.https://doi.org/10.1016/j.eatbeh.2011.07.007

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: a systematic review of the literature. *The International journal of eating disorders*, *45*(3), 428–438. https://doi.org/10.1002/eat.20931

Black, C. M., & Wilson, G. T. (1996). Assessment of eating disorders: interview versus questionnaire. *The International journal of eating disorders*, *20*(1), 43–50. DOI: 10.1002/(SICI)1098-108X(199607)20:1<43::AID-EAT5>3.0.CO;2-4

Borsboom, D., Romeijn, J.-W., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods*, *13*(2), 75-98. https://doi.org/10.1037/1082989X.13.2.75

Buchwald, H., Avidor, Y., Braunwald, E., Jensen, M. D., Pories, W., Fahrbach, K., & Schoelles, K. (2004). Bariatric surgery: a systematic review and meta-analysis. *JAMA*, *292*(14), 1724–1737. https://doi.org/10.1001/jama.292.14.1724

Bulik, C. M., Sullivan, P. F., & Kendler, K. S. (2002). Medical and psychiatric morbidity in obese women with and without binge eating. *The International journal of eating disorders*, *32*(1), 72–78. https://doi.org/10.1002/eat.10072

Byrne, S. M., Allen, K. L., Lampard, A. M., Dove, E. R., & Fursland, A. (2010). The factor structure of the eating disorder examination in clinical and community samples. *The International journal of eating disorders*, *43*(3), 260–265. https://doi.org/10.1002/eat.20681

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456-466.10.1037/0033-2909.105.3.456

Carey, M., Kupeli, N., Knight, R., Troop, N. A., Jenkinson, P. M., & Preston, C. (2019). Eating Disorder Examination Questionnaire (EDE-Q): Norms and psychometric properties in U.K. females and males. *Psychological assessment*, *31*(7), 839–850. https://doi.org/10.1037/pas0000703

Carter, J. C., & Fairburn, C. G. (1998). Cognitive-behavioral self-help for binge eating disorder: a controlled effectiveness study. *Journal of consulting and clinical psychology*, *66*(4), 616–623. https://doi.org/10.1037//0022-006x.66.4.616

Chan, C. W., & Leung, S. F. (2015). Validation of the Eating Disorder Examination Questionnaire: an online version. *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association*, *28*(6), 659–665. https://doi.org/10.1111/jhn.12275

Compte, E. J., McGuire, F. H., Brown, T. A., Lavender, J. M., Murray, S. B., Capriotti, M. R., Flentje, A., Lubensky, M. E., Lunn, M. R., Obedin-Maliver, J., & Nagata, J. M. (2023). Investigating the factor structure and measurement invariance of the eating disorder examination questionnaire (EDE-Q) among cisgender gay men and lesbian women from the United States. *Journal of eating disorders*, *11*(1), 164. https://doi.org/10.1186/s40337-023-00880-2

Contreras-Valdez, J. A., Freyre, M. Á., & Mendoza-Flores, E. (2022). The eating disorder examination questionnaire for adults from the Mexican general population: Reliability and validity. *PloS one*, *17*(4), e0266507. https://doi.org/10.1371/journal.pone.0266507

Cronbach, L. J., & Meehl, P. E. (1995). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302. DOI: 10.1037/h0040957.

Crow, S. J., Peterson, C. B., Swanson, S. A., Raymond, N. C., Specker, S., Eckert, E. D., & Mitchell, J. E. (2009). Increased mortality in bulimia nervosa and other eating disorders. *The American journal of psychiatry*, *166*(12), 1342–1346. https://doi.org/10.1176/appi.ajp.2009.09020247

Darcy, A. M., Hardy, K. K., Crosby, R. D., Lock, J., & Peebles, R. (2013). Factor structure of the Eating Disorder Examination Questionnaire (EDE-Q) in male and female college athletes. *Body image*, *10*(3), 399–405. https://doi.org/10.1016/j.bodyim.2013.01.008

De Ridder, W. A., van Kooij, Y. E., Vermeulen, G. M., Slijper, H. P., Selles, R. W., Wouters, R. M., & the Hand-Wrist Study Group (2021). Test-retest Reliability and Construct Validity of the Satisfaction with Treatment Result Questionnaire in Patients with Hand and Wrist Conditions: A Prospective Study. *Clinical orthopaedics and related research*, *479*(9), 2022–2032. https://doi.org/10.1097/CORR.0000000000001794

Diemer, E. W., Grant, J. D., Munn-Chernoff, M. A., Patterson, D. A., & Duncan, A. E. (2015). Gender identity, sexual orientation, and eating-related pathology in a National sample of college students. *Journal of Adolescent Health, 57*(2), 144–149. https://doi.org/10.1016/j.jadohealth.2015.03.003

Dufour, R., Steiger, H., & Booij, L. (2025). Examining Dimensionality and Item-Quality of the Eating Disorder Examination Questionnaire in Individuals With Eating Disorders Using Item Response Theory Analysis. *The International journal of eating disorders*, *58*(2), 349–361. https://doi.org/10.1002/eat.24330

El Hajjar, S. T. (2018). Statistical analysis: Internal-consistency reliability and construct validity. *International Journal of Quantitative and Qualitative Research Methods, 6*, 27–38.

Elsman, E. B. M., Mokkink, L. B., Terwee, C. B., Beaton, D., Gagnier, J. J., Tricco, A. C., Baba, A., Butcher, N. J., Smith, M., Hofstetter, C., Lee Aiyegbusi, O., Berardi, A., Farmer, J., Haywood, K. L., Krause, K. R., Markham, S., Mayo-Wilson, E., Mehdipour, A., Ricketts, J., Szatmari, P., … Offringa, M. (2024). Guideline for reporting systematic reviews of outcome measurement instruments (OMIs): PRISMA-COSMIN for OMIs 2024. *Journal of clinical epidemiology*, *173*, 111422. https://doi.org/10.1016/j.jclinepi.2024.111422

Fairburn, C. (2008). *Cognitive Behavior Therapy and Eating Disorders*. The Guilford Press.

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/10.1002/1098108X(199412)16:4<363::AIDEAT2260160405>3.0.CO;2-%23

Fairburn, C. G., & Cooper, Z. (1993). The Eating Disorder Examination. In C. G. Fairburn & G. T. Wilson (Eds.), *Binge eating: Nature, assessment, and treatment* (12th ed., pp. 317–360). Guilford Press.

Fairburn, C. G., & Harrison, P. J. (2003). Eating disorders. *Lancet*, *361*(9355), 407–416. https://doi.org/10.1016/S0140-6736(03)12378-1

Flores C. A. (2014). Psychological assessment for bariatric surgery: current practices. *Arquivos brasileiros de cirurgia digestiva*, *27(*1), 59–62. https://doi.org/10.1590/s0102-6720201400s100015

Forrest, L. N., Zuromski, K. L., Dodd, D. R., & Smith, A. R. (2017). Suicidality in adolescents and adults with binge-eating disorder: Results from the national comorbidity survey

replication and adolescent supplement. *The International journal of eating disorders*, *50*(1), 40–49. https://doi.org/10.1002/eat.22582

Friborg, O., Reas, D. L., Rosenvinge, J. H., & Rø, Ø. (2013). Core pathology of eating disorders as measured by the Eating Disorder Examination Questionnaire (EDE-Q): the predictive role of a nested general (g) and primary factors. *International journal of methods in psychiatric research*, *22*(3), 195–203. https://doi.org/10.1002/mpr.1389

Gideon, N., Hawkes, N., Mond, J., Saunders, R., Tchanturia, K., & Serpell, L. (2016). Development and Psychometric Validation of the EDE-QS, a 12 Item Short Form of the Eating Disorder Examination Questionnaire (EDE-Q). *PloS one*, *11*(5), e0152744. https://doi.org/10.1371/journal.pone.0152744

Giovinazzo, S., Sukkar, S. G., Rosa, G. M., Zappi, A., Bezante, G. P., Balbi, M., & Brunelli, C. (2019). Anorexia nervosa and heart disease: a systematic review. *Eating and weight disorders: EWD*, *24*(2), 199–207. https://doi.org/10.1007/s40519-018-0567-1

Gosseaume, C., Dicembre, M., Bemer, P., Melchior, J-C., & Hanachi, M. (2019). Somatic complications and nutritional management of anorexia nervosa. *Clinical Nutrition Experimental*, *28*, 2–10 https://doi.org/10.1016/j.yclnex.2019.09.001

Grilo, C. M., Henderson, K. E., Bell, R. L., & Crosby, R. D. (2013). Eating disorder examination-questionnaire factor structure and construct validity in bariatric surgery candidates. *Obesity Surgery*, *23*(5), 657–662. https://doi.org/10.1007/s11695-012-0840-8

Grilo, C. M., Masheb, R. M., & Wilson, G. T. (2001). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *Journal of consulting and clinical psychology*, *69*(2), 317–322. https://doi.org/10.1037//0022-006x.69.2.317

Grilo, C. M., Reas, D. L., Hopwood, C. J., & Crosby, R. D. (2015). Factor structure and construct validity of the Eating Disorder Examination-Questionnaire in college students: further support for a modified brief version. *The International journal of eating disorders*, *48*(3), 284–289. https://doi.org/10.1002/eat.22358

Guss, C. E., Williams, D. N., Reisner, S. L., Austin, S. B., & Katz-Wise, S. L. (2017). Disordered weight management behaviors, nonprescription steroid use, and weight perception in transgender youth. *Journal of Adolescent Health*, *60*(1), 17–22. https://doi.org/10.1016/j.jadohealth.2016.08.027

Habashy, J., Benning, S. D., Renn, B. N., Borgogna, N. C., Lawrence, E. M., & Kraus, S. W. (2023). Psychometric properties of the eating disorder examination questionnaire: Factor analysis and measurement invariance by race/ethnicity and gender. *Eating behaviors*, *48*, 101696. https://doi.org/10.1016/j.eatbeh.2022.101696

Hambleton, A., Pepin, G., Le, A., Maloney, D., National Eating Disorder Research Consortium, Touyz, S., & Maguire, S. (2022). Psychiatric and medical comorbidities of eating disorders: findings from a rapid review of the literature. *Journal of eating disorders*, *10*(1), 132. https://doi.org/10.1186/s40337-022-00654-2

Harrison, R., Jones, B., Gardner, P., & Lawton, R. (2021). Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. *BMC health services research*, *21*(1), 144. https://doi.org/10.1186/s12913-021-06122-y

Hermont, A. P., Oliveira, P. A., Martins, C. C., Paiva, S. M., Pordeus, I. A., & Auad, S. M. (2014). Tooth erosion and eating disorders: a systematic review and meta-analysis. *PloS one*, *9*(11), e111123. https://doi.org/10.1371/journal.pone.0111123

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). John Wiley & Sons.

Hoek, H. W. (2002). Distribution of eating disorders. In C. G. Fairburn & K. D. Brownell (Eds.). *Eating Disorders and Obesity: A Comprehensive Handbook* (2nd ed., pp. 233-237). The Guilford Press.

Hoek H. W. (2006). Incidence, prevalence and mortality of anorexia nervosa and other eating disorders. *Current opinion in psychiatry*, *19*(4), 389–394. https://doi.org/10.1097/01.yco.0000228759.95237.78

Holt, S., Ford, M. J., Grant, S., & Heading, R. C. (1981). Abnormal Gastric Emptying in Primary Anorexia Nervosa. *British Journal of Psychiatry*, *139*(6), 550–552. https://doi.org/10.1192/bjp.139.6.550

Hudson, J. I., Hiripi, E., Pope, H. G., Jr, & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological psychiatry*, *61*(3), 348–358. https://doi.org/10.1016/j.biopsych.2006.03.040

Huke, V., Turk, J., Saeidi, S., Kent, A., & Morgan, J. F. (2013). Autism spectrum disorders in eating disorder populations: a systematic review. *European eating disorders review : the journal of the Eating Disorders Association*, *21*(5), 345–351. https://doi.org/10.1002/erv.2244

International Test Commission (2017). The ITC Guidelines for Translating and Adapting Tests (Second edition). www.InTestCom.org

Isomaa, R., Lukkarila, I. L., Ollila, T., Nenonen, H., Charpentier, P., Sinikallio, S., & Karhunen, L. (2016). Development and preliminary validation of a Finnish version of the Eating Disorder Examination Questionnaire (EDE-Q). *Nordic journal of psychiatry*, *70*(7), 542–546. https://doi.org/10.1080/08039488.2016.1179340

Jacobs, K. L., & Krueger, R. F. (2015). The importance of structural validity. In P. Zachar, D. Stoyanov, M. Aragona, & A. Jablensky (Eds.), *Alternative perspectives on psychiatric validation* (chapter 11). Oxford University Press.

Jenkins, P. E., Blalock, D. V., Duffy, A., Mehler, P. S., & Rienecke, R. D. (2024). Empirical Investigation of Different Factor Structures for the Eating Disorder Examination-Questionnaire in Adult Women With Anorexia Nervosa. *Assessment*, *31*(4), 920–932. https://doi.org/10.1177/10731911231198207

Keeton, J., Ofori, A., Booker, Q., Schneider, B., McAdams, C., & Messiah, S. E. (2020). Psychosocial Factors that Inform the Decision to Have Metabolic and Bariatric Surgery Utilization in Ethnically Diverse Patients. *Obesity surgery*, *30*(6), 2233–2242. https://doi.org/10.1007/s11695-020-04454-y

Kelly, N. R., Mitchell, K. S., Gow, R. W., Trace, S. E., Lydecker, J. A., Bair, C. E., & Mazzeo, S. (2012). An evaluation of the reliability and construct validity of eating disorder measures in white and black women. *Psychological assessment*, *24*(3), 608–617. https://doi.org/10.1037/a0026457

Kliem, S., Mößle, T., Zenger, M., Strauß, B., Brähler, E., & Hilbert, A. (2016). The eating disorder examination-questionnaire 8: A brief measure of eating disorder psychopathology (EDE-Q8). *International Journal of Eating Disorders*, *49*(6), 613–616. https://doi.org/10.1002/eat.22487

Klimek, P., Convertino, A. D., Pennesi, J. L., Gonzales, M., 4th, Roesch, S. C., Nagata, J. M., & Blashill, A. J. (2021). Confirmatory factor and measurement invariance analyses of the Eating Disorder Examination Questionnaire in sexual minority men and women. *The International journal of eating disorders*, *54*(5), 745–754. https://doi.org/10.1002/eat.23488

Knight, R., & Preston, C. (2023). Exploring the effects of gender and sexual orientation on disordered eating: an EFA to CFA study of the Eating Disorder Examination Questionnaire. *Journal of eating disorders*, *11*(1), 100. https://doi.org/10.1186/s40337-023-00821-z

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Köster E. P. (2009). Diversity in the determinants of food choice: A psychological perspective. *Food Quality and Preference*, *20*(2), 70-82.

Kudel, I., Pona, A., Cox, S., Szoka, N., Tabone, L., & Brode, C. (2019). Psychometric properties of NIH PROMIS® instruments in bariatric surgery candidates. *Health psychology: official journal of the Division of Health Psychology, American Psychological Association*, *38*(5), 359–368. https://doi.org/10.1037/hea0000697

Lac, A. (2016). Measurement Validity. In R. Levesque (Eds.). *Encyclopedia of Adolescence*. Springer. https://doi.org/10.1007/978-3-319-32132-5_781-1

Laskowski, N. M., Halbeisen, G., Braks, K., Huber, T. J., & Paslakis, G. (2023). Factor structure of the Eating Disorder Examination-Questionnaire (EDE-Q) in adult men with eating disorders. *Journal of eating disorders*, *11*(1), 34. https://doi.org/10.1186/s40337-023-00757-4

Lev-Ari, L., Bachner-Melman, R., & Zohar, A. H. (2021). Eating Disorder Examination Questionnaire (EDE-Q-13): expanding on the short form. *Journal of eating disorders*, *9*(1), 57. https://doi.org/10.1186/s40337-021-00403-x

Lindvall Dahlgren, C., Wisting, L., & Rø, Ø. (2017). Feeding and eating disorders in the DSM-5 era: a systematic review of prevalence rates in non-clinical male and female samples. *Journal of eating disorders*, *5*, 56. https://doi.org/10.1186/s40337-017-0186-7

Linna, M. S., Raevuori, A., Haukka, J., Suvisaari, J. M., Suokas, J. T., & Gissler, M. (2013). Reproductive health outcomes in eating disorders. *The International journal of eating disorders*, *46*(8), 826–833. https://doi.org/10.1002/eat.22179

de Lucena, M. C. M. D., de Souza, H. K. C., & Alchieri, J. C. (2012). Revisão de aspetos técnico-metodológicos da avaliação psicológica de candidatos à cirurgia bariátrica. *Boletim Academia Paulista de Psicologia, 32*(83), 408-423.

Machado, P. P. P., Grilo, C. M., Rodrigues, T. F., Vaz, A. R., & Crosby, R. D. (2020). Eating Disorder Examination - Questionnaire short forms: A comparison. *The International journal of eating disorders*, *53*(6), 937–944. https://doi.org/10.1002/eat.23275

Machado, P. P., Martins, C., Vaz, A. R., Conceição, E., Bastos, A. P., & Gonçalves, S. (2014). Eating disorder examination questionnaire: psychometric properties and norms for the Portuguese population. *European eating disorders review : the journal of the Eating Disorders Association*, *22*(6), 448–453. https://doi.org/10.1002/erv.2318

Mahmoodi, M., Moloodi, R., Ghaderi, A., Babai, Z., Saleh, Z., Alasti, H., Naghashian, F., & Mohammadpour, Z. (2016). The Persian Version of Eating Disorder Examination Questionnaire and Clinical Impairment Assessment: Norms and Psychometric Properties for Undergraduate Women. *Iranian journal of psychiatry*, *11*(2), 67–74.

Mayes, S. D., Fernandez-Mendoza, J., Baweja, R., Calhoun, S., Mahr, F., Aggarwal, R., & Arnold, M. (2014). Correlates of suicide ideation and attempts in children and adolescents with eating disorders. *Eating disorders*, *22*(4), 352–366. https://doi.org/10.1080/10640266.2014.915694

Matsumoto, D. (2003). Cross-cultural research. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology*. Oxford: Blackwell.

McEntee, M. L., Serier, K. N., Smith, J. M., & Smith, J. E. (2021). The sum is greater than its parts: Intersectionality and measurement validity of the Eating Disorder Examination

Questionnaire (EDE-Q) in Latinx undergraduates in the United States. *Sex Roles: A Journal of Research, 84*(1-2), 102–111. https://doi.org/10.1007/s11199-020-01149-7

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica, 22(3), 276-282.

McLean, C. P., Lavale, A., Kulkarni, J., & Sharp, G. (2022). Psychometric properties and factor structure of the Eating Disorder Examination-Questionnaire in omnivores, vegetarians, and vegans. *Body image*, *43*, 374–384. https://doi.org/10.1016/j.bodyim.2022.10.006

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127-143.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525-543.

Mitchell, J. E., King, W. C., Courcoulas, A., Dakin, G., Elder, K., Engel, S., Flum, D., Kalarchian, M., Khandelwal, S., Pender, J., Pories, W., & Wolfe, B. (2015). Eating behavior and eating disorders in adults before bariatric surgery. *The International journal of eating disorders*, *48*(2), 215–222. https://doi.org/10.1002/eat.22275

Mitchison, D., Hay, P., Slewa-Younan, S., & Mond, J. (2014). The changing demographic profile of eating disorder behaviors in the community. *BMC public health*, *14*, 943. https://doi.org/10.1186/1471-2458-14-943

Mitchison, D., & Mond, J. (2015). Epidemiology of eating disorders, eating disordered behaviour, and body image disturbance in males: a narrative review. *Journal of eating disorders*, *3*, 20. https://doi.org/10.1186/s40337-015-0058-y

Mokkink, L. B., Elsman, E., & Terwee, C. B. (2024). The COSMIN guideline systematic reviews of Patient-Reported Outcome Measures (PROMs). *Quality of life research*. https://doi.org/10.1007/s11136-024-03761-6.

Murray, S. B., Nagata, J. M., Griffiths, S., Calzo, J. P., Brown, T. A., Mitchison, D., Blashill, A. J., & Mond, J. M. (2017). The enigma of male eating disorders: A critical review and synthesis. *Clinical psychology review*, *57*, 1–11. https://doi.org/10.1016/j.cpr.2017.08.001

National Heart, Lung, and Blood Institute. (2022, May 18). Heart treatments. U.S. Department of Health & Human Services, National Institutes of Health.

Nunes, M. A., Appolinário, J. C., Galvão, A. L., & Coutinho, W. (2006). *Transtornos alimentares e obesidade* (2nd ed.). Artem.

de Oliveira, J. H. A., & Yoshida, E. M. P. (2008). Avaliação psicológica de obesos grau III antes e depois de cirurgia bariátrica. *Psicologia: Reflexão e Crítica, 22*(1), 12-19. https://doi.org/10.1590/S0102-79722009000100003

de Oliveira Júnior, M. L., Almeida, M., Santos, C. G., Brown, T. A., & de Carvalho, P. H. B. (2023). Psychometric properties of the Eating Disorder Examination Questionnaire among Brazilian cisgender gay and bisexual adult men. *The International journal of eating disorders*, *56*(4), 736–746. https://doi.org/10.1002/eat.23853

Otani, M., Hiraide, M., Horie, T., Mitsui, T., Yoshida, T., Takamiya, S., Sakuta, R., Usami, M., Komaki, G., & Yoshiuchi, K. (2021). Psychometric properties of the Eating Disorder Examination-Questionnaire and psychopathology in Japanese patients with eating disorders. *The International journal of eating disorders*, *54*(2), 203–211. https://doi.org/10.1002/eat.23452

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. [DOI: 10.1186/s13643-016-0384-4]

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*(71). https://doi.org/10.1136/bmj.n71

Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2015). Psychometric evaluation of disordered eating measures in bariatric surgery patients. *Eating behaviors*, *19*, 39–48. https://doi.org/10.1016/j.eatbeh.2015.05.007

Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2016). Psychometric Evaluation of Disordered Eating Measures in Bariatric Surgery Candidates. *Obesity surgery*, *26*(3), 563–575. https://doi.org/10.1007/s11695-015-1780-x

Passi, V. A., Bryson, S. W., & Lock, J. (2003). Assessment of eating disorders in adolescents with anorexia nervosa: self-report questionnaire versus interview. *The International journal of eating disorders*, *33*(1), 45–54. https://doi.org/10.1002/eat.10113

Pasternak, Y., Weintraub, A. Y., Shoham-Vardi, I., Sergienko, R., Guez, J., Wiznitzer, A., Shalev, H., & Sheiner, E. (2012). Obstetric and perinatal outcomes in women with eating disorders. *Journal of women's health (2002)*, *21*(1), 61–65. https://doi.org/10.1089/jwh.2011.2907

Peterson, C. B., Crosby, R. D., Wonderlich, S. A., Joiner, T., Crow, S. J., Mitchell, J. E., Bardone-Cone, A. M., Klein, M., & le Grange, D. (2007). Psychometric properties of the eating disorder examination-questionnaire: factor structure and internal consistency. *The International journal of eating disorders*, *40*(4), 386–389. https://doi.org/10.1002/eat.20373

Phillips, K. E., Jennings, K. M., & Gregas, M. (2018). Factor Structure of the Eating Disorder Examination-Questionnaire in a Clinical Sample of Adult Women With Anorexia

Nervosa. *Journal of psychosocial nursing and mental health services*, *56*(5), 33–39. https://doi.org/10.3928/02793695-20180108-03

Prnjak, K., & Jukic, I. (2021). Development and validation of the Croatian version of the Eating Disorder Examination Questionnaire in a community sample. *Eating and weight disorders : EWD*, *26*(3), 859–868. https://doi.org/10.1007/s40519-020-00915-6

Rand-Giovannetti, D., Cicero, D. C., Mond, J. M., & Latner, J. D. (2020). Psychometric Properties of the Eating Disorder Examination-Questionnaire (EDE-Q): A Confirmatory Factor Analysis and Assessment of Measurement Invariance by Sex. *Assessment*, *27*(1), 164–177. https://doi.org/10.1177/1073191117738046

Reas, D. L., Grilo, C. M., & Masheb, R. M. (2006). Reliability of the Eating Disorder Examination-Questionnaire in patients with binge eating disorder. *Behaviour research and therapy*, *44*(1), 43–51. https://doi.org/10.1016/j.brat.2005.01.004

Rica, R., Solar, M., Compte, E. J., & Sepúlveda, A. R. (2022). Establishing the optimal male cut-off point: confirmatory factor analysis of the eating disorder examination-questionnaire (EDE-Q) in a representative sample of Spanish university students. *Eating and weight disorders : EWD*, *27*(3), 969–977. https://doi.org/10.1007/s40519-021-01234-0

Rikani, A. A., Choudhry, Z., Choudhry, A. M., Ikram, H., Asghar, M. W., Kajal, D., Waheed, A., & Mobassarah, N. J. (2013). A critique of the literature on etiology of eating disorders. *Annals of neurosciences*, *20*(4), 157–161. https://doi.org/10.5214/ans.0972.7531.200409

Rø, Ø., Reas, D. L., & Lask, B. (2010). Norms for the Eating Disorder Examination Questionnaire among female university students in Norway. *Nordic journal of psychiatry*, *64*(6), 428–432. https://doi.org/10.3109/08039481003797235

Rø, Ø., Reas, D. L., & Stedal, K. (2015). Eating Disorder Examination Questionnaire (EDE-Q) in Norwegian Adults: Discrimination between Female Controls and Eating Disorder Patients. *European eating disorders review : the journal of the Eating Disorders Association*, *23*(5), 408–412. https://doi.org/10.1002/erv.2372

Roehrig, M., Masheb, R. M., White, M. A., & Grilo, C. M. (2009). The metabolic syndrome and behavioral correlates in obese patients with binge eating disorder. *Obesity (Silver Spring, Md.)*, *17*(3), 481–486. https://doi.org/10.1038/oby.2008.560

Rose, J. S., Vaewsorn, A., Rosselli-Navarra, F., Wilson, G. T., & Weissman, R. S. (2013). Test-retest reliability of the eating disorder examination-questionnaire (EDE-Q) in a college sample. *Journal of eating disorders*, *1*, 42. https://doi.org/10.1186/2050-2974-1-42

Sahlan, R. N., Akoury, L. M., Habashy, J., Culbert, K. M., & Warren, C. S. (2022). Sociocultural correlates of eating pathology in college women from US and Iran. *Frontiers in psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.966810

Schmidt, U., Adan, R., Böhm, I., Campbell, I. C., Dingemans, A., Ehrlich, S., Elzakkers, I., Favaro, A., Giel, K., Harrison, A., Himmerich, H., Hoek, H. W., Herpertz-Dahlmann, B., Kas, M. J., Seitz, J., Smeets, P., Sternheim, L., Tenconi, E., van Elburg, A., van Furth, E., & Zipfel, S. (2016). Eating disorders: the big issue. *The lancet. Psychiatry*, *3*(4), 313–315. https://doi.org/10.1016/S2215-0366(16)00081-X

Serier, K. N., Smith, J. E., & Yeater, E. A. (2018). Confirmatory factor analysis and measurement invariance of the Eating Disorder Examination Questionnaire (EDE-Q) in a non-clinical sample of non-Hispanic White and Hispanic women. *Eating behaviors*, *31*, 53–59. https://doi.org/10.1016/j.eatbeh.2018.08.004

Simone, M., Askew, A., Lust, K., Eisenberg, M. E., & Pisetsky, E. M. (2020). Disparities in self-reported eating disorders and academic impairment in sexual and gender minority

college students relative to their hetero-sexual and cisgender peers. *International Journal of Eating Disorders*, *53*(4), 513–524. https://doi.org/10.1002/eat.23226

Smaidi, K., & Gonçalves, M. (2016). Cirurgia Bariátrica e seus aspetos psiquiátricos. *Psychiatry On-Line Brazil*, *21*(1).

Smink, F. R., van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: incidence, prevalence and mortality rates. *Current psychiatry reports*, *14*(4), 406–414. https://doi.org/10.1007/s11920-012-0282-y

Smith, A. R., Zuromski, K. L., & Dodd, D. R. (2018). Eating disorders and suicidality: what we know, what we don't know, and suggestions for future research. *Current opinion in psychology*, *22*, 63–67. https://doi.org/10.1016/j.copsyc.2017.08.023

Stanford, S. C., & Lemberg, R. (2012). Measuring eating disorders in men: development of the eating disorder assessment for men (EDAM). *Eating disorders*, *20*(5), 427–436. https://doi.org/10.1080/10640266.2012.715522

Streiner D. L. (2003). Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *Journal of personality assessment*, *80*(3), 217–222. https://doi.org/10.1207/S15327752JPA8003_01

Strother, E., Lemberg, R., Stanford, S. C., & Turberville, D. (2012). Eating Disorders in Men: Underdiagnosed, Undertreated, and Misunderstood. *Eating Disorders*, *20*(5), 346–355. https://doi.org/10.1080/10640266.2012.715512

Sullivan, P. F. (1995). Mortality in anorexia nervosa. *The American Journal of Psychiatry, 152*(7), 1073–1074. https://doi.org/10.1176/ajp.152.7.1073

Swanson, S. A., Crow, S. J., Le Grange, D., Swendsen, J., & Merikangas, K. R. (2011). Prevalence and correlates of eating disorders in adolescents. Results from the national comorbidity survey replication adolescent supplement. *Archives of general psychiatry*, *68*(7), 714–723. https://doi.org/10.1001/archgenpsychiatry.2011.22

Taib, M. N., Abdul Khaiyom, J. H., & Fauzaman, J. (2021). Psychometric properties of the adapted Malay Eating Disorder Examination-Questionnaire 6.0 (EDE-Q 6.0) among university students in Malaysia. *Eating behaviors*, *42*, 101533. https://doi.org/10.1016/j.eatbeh.2021.101533

Tang, W., Cui, Y., & Babenko, O. (2014). Internal consistency: Do we really know what it is and how to assess it?. *Journal of Psychology and Behavioral Science*, *2*(2), 205–220.

Unikel Santoncini, C., Bojorquez Chapela, I., Díaz de León Vázquez, C., Vázquez Velázquez, V., Rivera Márquez, J. A., Galván Sánchez, G., & Rocha Velis, I. (2018). Validation of eating disorders examination questionnaire in Mexican women. *The International journal of eating disorders*, *51*(2), 146–154. https://doi.org/10.1002/eat.22819

van Hoeken, D., & Hoek, H. W. (2020). Review of the burden of eating disorders: mortality, disability, costs, quality of life, and family burden. *Current opinion in psychiatry*, *33*(6), 521–527. https://doi.org/10.1097/YCO.0000000000000641

Vaz, A. R., Conceição, E. M., & Machado, P. P. P. (2009). A abordagem cognitivo-comportamental no tratamento das perturbações do comportamento alimentar. *Análise Psicológica*, *27*(2), 189-197.

Walsh, B. T., Fairburn, C. G., Mickley, D., Sysko, R., & Parides, M. K. (2004). Treatment of Bulimia Nervosa in a Primary Care Setting. *The American Journal of Psychiatry, 161*(3), 556–561. https://doi.org/10.1176/appi.ajp.161.3.556

Weigel, A., Löwe, B., & Kohlmann, S. (2019). Severity of somatic symptoms in outpatients with anorexia and bulimia nervosa. *European eating disorders review : the journal of the Eating Disorders Association*, *27*(2), 195–204. https://doi.org/10.1002/erv.2643

Williamson, D. A., Anderson, D. A., Jackman, L. P., & Jackson, S. R. (1995). Assessment of eating disordered thoughts, feelings, and behaviors. In D. B. Allison (Ed.), *Handbook*

of assessment methods for eating behaviors and weight-related problems: Measures, theory, and research (pp. 347–386). Sage Publications, Inc.

Wonderlich, S. A., Joiner, T. E., Jr., Keel, P. K., Williamson, D. A., & Crosby, R. D. (2007). Eating disorder diagnoses: Empirical approaches to classification. *American Psychologist, 62*(3), 167–180. https://doi.org/10.1037/0003-066X.62.3.167

Wuensch, K. L. (2012). *A brief introduction to reliability, validity, and scaling*. East Carolina University.

Zaina, F., Pesenti, F., Persani, L., Capodaglio, P., Negrini, S., & Polli, N. (2018). Prevalence of idiopathic scoliosis in anorexia nervosa patients: results from a cross-sectional study. *European spine journal: official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, *27*(2), 293–297. https://doi.org/10.1007/s00586-017-5181-9

Zipfel, S., Sammet, I., Rapps, N., Herzog, W., Herpertz, S., & Martens, U. (2006). Gastrointestinal disturbances in eating disorders: clinical and neurobiological aspects. *Autonomic neuroscience: basic & clinical*, *129*(1-2), 99–106. https://doi.org/10.1016/j.autneu.2006.07.023

Zohar, A. H., Lev-Ari, L., & Bachner-Melman, R. (2017). The EDE-Q in Hebrew: Structural and Convergent/Divergent Validity in a Population Sample. *Israel journal of psychiatry*, *54*(3), 15–20.

Supplementary Material

File 1 - Search Query

SCOPUS

TITLE-ABS-KEY (("EDE-Q" OR "Eating Disorder Examination-Questionnaire") AND ("reliability" OR "validity" OR "validation" OR "psychometric properties" OR "psychometric validation" OR "factor structure" OR "structural validity" OR "factor analysis" OR "CFA" OR "EFA" OR "measurement invariance")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SRCTYPE , "j"))

PubMed

(("EDE-Q"[Title/Abstract] OR "Eating Disorder Examination-Questionnaire"[Title/Abstract]) AND ("reliability"[Title/Abstract] OR "validity"[Title/Abstract] OR "validation"[Title/Abstract] OR "psychometric properties"[Title/Abstract] OR "psychometric validation"[Title/Abstract] OR "factor structure"[Title/Abstract] OR "factor analysis"[Title/Abstract] OR "CFA"[Title/Abstract] OR "EFA"[Title/Abstract] OR "measurement invariance"[Title/Abstract]))

The script was completed with the following filters: “Journal article”

APA PsycNet

("EDE-Q" OR "Eating Disorder Examination-Questionnaire") AND (“reliability" OR "validity" OR "validation" OR "psychometric properties" OR "psychometric validation" OR "factor structure" OR "factor analysis" OR "CFA" OR "EFA” OR "measurement invariance”

The script was completed with the following filters: “in Abstract”, “Document type: Journal Articles”

File 2 - Quality assessment of each individual study included in the present systematic review

	Box 4. Internal consistency			
	Statistical methods			Other
Authors (year)	For continuous scores: Was Cronbach's alpha or omega calculated?	For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	Were there any other important flaws in the design or statistical methods of the study?
Aldubayan et al. (2023)	Very good	NA	NA	Very good
Jenkins et al. (2024)	Very good	NA	NA	Very good
de Oliveira Junior et al. (2023)	Very good	NA	NA	Very good
McEntee et al. (2021)	Very good	NA	NA	Adequate
Otani et al. (2021)	Very good	NA	NA	Very good
Klimek et al. (2021)	Very good	NA	NA	Very good
Taib et al. (2021)	Very good	NA	NA	Very good
Carey et al. (2019)	Very good	NA	NA	Very good
Friborg et al. (2013)	Very good	NA	NA	Very good
Contreras-Valdez et al. (2022)	Very good	NA	NA	Very good
Compte et al. (2023)	Very good	NA	NA	Very good
Knight & Preston (2023)	Very good	NA	NA	Very good
Laskowski et al. (2023)	Very good	NA	NA	Very good
Rica et al. (2022)	Very good	NA	NA	Very good
Parker et al. (2016)	Very good	NA	NA	Very good

Prnjak & Jukic (2021)	Very good	NA	NA	Very good
Philips et al. (2018)	Very good	NA	NA	Very good
Unikel Santoncini et al. (2018)	Very good	NA	NA	Very good
Isomaa et al. (2016)	Very good	NA	NA	Very good
Rø et al. (2010)	Very good	NA	NA	Very good
Rose et al. (2013)	Very good	NA	NA	Very good
Baceviciene et al. (2020)	Very good	NA	NA	Very good
McLean et al. (2022)	Very good	NA	NA	Very good
Mahmoodi et al. (2016)	Very good	NA	NA	Very good
Parker et al. (2015)	Very good	NA	NA	Very good

Box 5. Cross-cultural validity\measurement invariance

	Design requirements	**Statistical methods**		**Other**
Authors (year)	Were the samples similar for relevant characteristics except for the group variable?	Was an appropriate approach used to analyse the data?	Was the sample size included in the analysis adequate?	Were there any other important flaws in the design or statistical methods of the study?
Aldubayan et al. (2023)	Adequate	Doubtful	Very good	Doubtful
Habashy et al. (2023)	Very good	Very good	Very good	Very good
Jenkins et al. (2024)	Adequate	Very good	Very good	Very good
de Oliveira Junior et al. (2023)	Adequate	Doubtful	Adequate	Very good
McEntee et al. (2021)	Very good	Adequate	Very good	Very good
Otani et al. (2021)	Very good	Doubtful	Adequate	Doubtful

Klimek et al. (2021)	Very good	Very good	Very good	Very good
Taib et al. (2021)	very good	Doubtful	Very good	Doubtful
Zohar et al. (2017)	Adequate	Doubtful	Very good	Doubtful
Carey et al. (2019)	Adequate	Adequate	Very good	Very good
Rand-Giovannetti et al. (2020)	Adequate	Very good	Very good	Very good
Contreras-Valdez et al. (2022)	Very good	Very good	Very good	Very good
Compte et al. (2023)	Adequate	Very good	Very good	Very good
Knight & Preston (2023)	Very good	Very good	Very good	Very good
Allen et al. (2011)	Adequate	Adequate	Very good	Doubtful
Prnjak & Jukic (2021)	Adequate	Doubtful	Very good	Doubtful
Isomaa et al. (2016)	Adequate	Doubtful	Adequate	Doubtful
Baceviciene et al. (2020)	Adequate	Very good	Very good	Very good
Sahlan et al. (2022)	Adequate	Very good	Very good	Very good
Mahmoodi et al. (2016)	Adequate	Doubtful	Very good	Doubtful

Box 8. Criterion validity

	Statistical methods		**Other**
Authors (year)	For continuous scores: were correlations, or the AUC calculated?	For dichotomous scores: were sensitivity and specificity determined?	Were there any other important flaws in the design or statistical methods of the study?
Rø et al. (2015)	Very good	Very good	Very good

Box 9a. Hypotheses testing for construct validity: Comparison with other outcome measurement instruments (convergent validity)

	Design requirements	Statistical methods		Other
Authors (year)	Is it clear what the comparator instrument(s) measure(s)?	Were the measurement properties of the comparator instrument(s) sufficient?	Were statistical methods appropriate for the subgroups being compared?	Were there any other important flaws in the design or statistical methods of the study?
Otani et al. (2021)	Very good	Very good	Very good	Very good
McLean et al. (2022)	Very good	Very good	Very good	Very good
Mahmoodi et al. (2016)	Very good	Very good	Very good	Very good
Prnjak & Jukic (2021)	Very good	Very good	Very good	Very good

Box 9b. Hypotheses testing for construct validity: Comparison between subgroups (discriminative or known-groups validity)

	Design requirements	Statistical methods	Other
Authors (year)	Was an adequate description provided of important characteristics of the subgroups?	Were statistical methods appropriate for the subgroups being compared?	Were there any other important flaws in the design or statistical methods of the study?
Otani et al. (2021)	Very good	Very good	Very good
Baceviciene et al. (2020)	Adequate	Very good	Very good
Rø et al. (2015)	Very good	Adequate	Adequate
McLean et al. (2022)	Very good	Very good	Very good
Mahmoodi et al. (2016)	Very good	Very good	Very good

Box 6. Reliability

	Design requirements			Statistical methods				Other
Authors (year)	Were patients stable on the construct to be measured in the time between the repeated measurements?	Was the time interval between the repeated measurements appropriate?	Were the measurement conditions similar for the repeated measurements – except for the condition being evaluated?	For continuous scores: Was the appropriate intraclass correlation coefficient (ICC) calculated?	For dichotomous scores: was kappa calculated?	For nominal scores: was an unweighted kappa calculated?	For ordinal scores: was a weighted kappa calculated?	Were there any other important flaws in the design or statistical methods of the study?
de Oliveira Junior et al. (2023)	Very good	Very good	Very good	Very good	NA	NA	NA	Very good
Rø et al. (2010)	Very good	Very good	Adequate	Doubtful	NA	NA	NA	Adequate
Rose et al. (2013)	Very good	Very good	Adequate	Doubtful	NA	NA	NA	Adequate
Baceviciene et al. (2020)	Very good	Very good	Adequate	Very good	NA	NA	NA	Very good
McLean et al. (2022)	Very good	Very good	Adequate	Very good	NA	NA	NA	Very good