

Test-retest reliability of a simplified questionnaire for screening adolescents with risk behaviours for eating disorders in epidemiologic studies

Abstract

Introduction: The number of people presenting abnormal eating behaviors based on the habit of going on strict diets and taking measures to compensate food consumption (use of laxatives, diuretics, and self-induced vomiting) is growing. There is a clear need for simple investigative tools in population-based studies. **Objective:** Evaluate the reliability of a simplified questionnaire to investigate risk behaviors for eating disorders (ED) among adolescents. **Methods:** The questionnaire had two questions: the second question was subdivided into four other ones, about the frequency of binge eating episodes, the use of laxatives and diuretics, self-induced vomiting and strict dieting. It was applied in two moments (test-retest), with an interval of 15 days between the two assessments, in 195 students (70 % girls) aged from 12 to 19.9 years, from a Niterói, RJ public school. The reliability of the questions was assessed through the weighted Kappa coefficient and their respective 95% confidence intervals (CI) and frequency adjusted kappa. **Results:** Frequency-adjusted kappa values for less frequent behaviors were: 0.93 for the use of laxatives, 0.97 for the use of diuretics and 0.92 for self-induced vomiting. For the more frequent risk behaviors for ED such as binge eating episodes and restrictive dieting, weighted Kappa coefficients were 0.50 and 0.61, respectively. **Conclusions:** The reliability of the questionnaire was good, with better results for investigating less frequent behaviors.

Keywords: Eating Disorders, adolescents, questionnaire reliability.

Julia Elba de Souza FERREIRA

Gloria Valeria da VEIGA

Instituto de Nutrição Josué de Castro, Universidade Federal do Rio de Janeiro - Av. Brigadeiro Trompowsky, s/nº, Ilha do Fundão, CCS, Bloco J 2º andar

Based on study on Risk Behaviors for Eating Disorders among Adolescents in Public Schools of Niterói-RJ (*Comportamentos de Risco para Transtornos Alimentares em adolescentes estudantes de escolas públicas de Niterói-RJ*), for the MSc dissertation, presented to the Graduate Program in Nutrition of the Instituto de Nutrição Josué de Castro of the Federal University of Rio de Janeiro.

No Funding source or conflict of interest

Forward mail to: Julia Elba de Souza Ferreira. Instituto de Nutrição Josué de Castro - Universidade Federal do Rio de Janeiro. Av. Brigadeiro Trompowsky, s/nº, Ilha do Fundão, CCS, Bloco J 2º andar. Rio de Janeiro, RJ. CEP 21941-590. E-mail: jesferreira@globo.com

Introduction

The number of adolescents presenting abnormal eating behaviors is increasing, stimulated by excessive concern over body weight and fear of fat. This phobia leads individuals to strict dieting or inappropriate methods to compensate overeating episodes, such as the use of laxatives and / or diuretics and self-induced vomiting¹.

These behaviors are part of the genesis and maintenance of eating disorders (ED), such as anorexia nervosa (AN), bulimia nervosa (BN) and binge eating disorder (BED). Depending on the intensity and frequency in which they occur, they are not diagnosed as ED but as non specified eating disorders or symptoms of risk behaviors for ED or abnormal eating behavior and may indicate a risk for the development of ED^{2,3,4}.

Abnormal eating behaviors have been observed in 15.8% of Italian girls⁵ and 53.3% of adolescent Latin girls¹, evaluated in the United States, with the habit of going on strict diets.

The epidemiological investigation of ED is scarce, probably on account of the difficulty for establishing a precise diagnosis in a short period of time. The most common form to approach the problem is the application of long and complex questionnaires, as analyzed by Freitas et al.⁶ in a broad review on the subject.

Individual interviews imply in the need for human resources and available time, which is not always feasible in epidemiological studies. Thus, the investigation of risk behaviors for ED, by means of simplified self-report questionnaires might be an alternative for contributing to the epidemiological investigation of ED. These questionnaires are considered efficient, economical and easy to use for a great number of individuals, besides allowing respondents to reveal a behavior they might be reluctant to mention in a face-to-face interview, for considering it shameful⁶.

In 1998, HAY⁷ used a simplified questionnaire to identify binge eating and inappropriate methods to compensate overeating

episodes, in a probabilistic sample of 3,000 Australians, in the 15-94 years age-range. This study has the objective of evaluating the reliability of a simplified questionnaire, adapted from the one proposed by Hay, and used for a sample of adolescent students from a public school in the city of Niterói, Rio de Janeiro, Brazil.

Methods

The present study was carried out with 195 adolescents in the 12-19.9 years age range, students of a state school, in the city of Niterói/RJ, in 2005. It is a convenience sample of all students enrolled in the grades comprised from the 2nd year of the Elementary level (1-8) to the 3rd year of the Intermediate level (1-3), who showed willingness to participate in this research and obtained signed consent of their legal guardians.

The questionnaire was applied in two distinct moments (test – retest) with a 15-day interval between the test application and the retest. For the first application of the questionnaire (test), 240 students were approached. Of these, 45 did not show on the day of the second questionnaire (re-test) application, with a loss of 18.7%. The questionnaire evaluated in this study (Annex) consists of an adaptation of the one used by Hay⁷ to verify the prevalence of ED in an Australian community. Hay⁷ based his questionnaire on a long script of the clinical interview known as Eating Disorder Examination - EDE, created by Fairburn⁸, and considered the “gold standard” for screening ED⁶. The version of the present study consisted in subdividing the second question of the original questionnaire into four sub-questions, aimed at providing better understanding to students.

The simplified questionnaire has 2 questions. The first question investigates the frequency of binge eating episodes and the second, divided into four inquiries, investigates the frequency of methods to control weight gain, or possible mechanis-

ms to compensate overeating episodes, such as purging (use of laxatives, diuretics and self-induced vomiting) and strict dieting or fasting. For each question, there were four categories of behavior frequency, ranging from no episodes to two or more times a week in the past three months.

The questionnaire was completed by students themselves in the classroom, in the presence of the researcher. When there was a question in any of the items, the researcher read the question to the student, without further explanation, aiming to avoid a differentiated influence in their response.

The anthropometrical evaluation was comprised weight and height. Weight was gauged by means of an electronic scale, with a capacity of up to 150 kg and a 50 g variation, and height was measured using a portable anthropometric measurer, with a 0.1 cm variation. Height was gauged twice, allowing for 0.5 cm maximum variation between the two measures, and the average value remained. The evaluation of the nutritional status was based on the body mass index ($BMI = \text{weight} / \text{height}^2$), according to the criteria of COLE et al.⁹ for the classification of overweight and obesity, and WHO's¹⁰ for classifying underweight.

Data analysis was carried out utilizing PEPI version 3.0¹¹ and Statistical Package for Social Sciences (SPSS), version 11.0 software. The Chi-Square test was used to compare frequencies. The reliability of the questions in the questionnaire was evaluated by weighted kappa (wk), using the absolute error as a weighting system, and the frequency adjusted kappa (ak) and their respective 95% confidence intervals. The proportion of concordant and discordant answers was also calculated for each question of the questionnaire¹². The criterion proposed by Landis & Koch *apud*: SILVA & PEREIRA¹³ was used for interpreting agreement and the $p < 0.05$ value for statistical significance.

The study was approved by the Ethics Committee on Research of the Instituto de Puericultura Martagão Gesteira (IPPMG) of the Federal University of Rio de Janeiro – Brazil.

Results

Among the 45 adolescents who did not participate in the second application of the questionnaire (losses), 28 (62.2%) were girls, 26 (57.8%) were 16 years old or above. These proportions did not differ significantly from those observed in the 195 adolescents who participated fully in the study (p value = 0.29 compared in gender and 0.39, in age range). It was possible to assess nutritional status, based on the BMI of 38 of these adolescents and differences were not significant: 89.5% were eutrophic ($p = 0.14$), 2.6% were underweight ($p = 0.32$), 5.3% overweight ($p = 0.11$), and 2.6% obese ($p = 0.64$).

Among the adolescents assessed, 29.7% ($n = 58$) were male and 70.3% ($n = 137$) female; 64.6% ($n = 126$) belonged to the 16-19.9 years age range, and 35.4% ($n = 69$) were younger, between 12-15.9 years old.

Regarding the nutritional status, 14.6% were overweight, 1.6% were obese, and 4.7% underweight. On account of the low prevalence of obesity, the overweight and obese categories were grouped to comprise the “overweight” category, corresponding to 16.2%. There was no significant difference related to gender and age range, with regard to nutritional status (Table 1).

Weighted kappa values for the questions concerning risk behaviors with low frequency, that is, those in which more than 90% of the group studied reported no episode, when adjusted to frequency, were very high ($k = 0.93$ for the use of laxatives; $k = 0.97$ for the use of diuretics, and $k = 0.92$ for self-induced vomiting). Whereas in the behaviors that presented a higher frequency, the pondered kappa values corresponded to moderate agreement ($k = 0.50$) for binge eating, and to a substantial one ($k = 0.61$) for strict dieting (Table 2).

In relation to the proportion of concordant answers between test and retest, 60% of agreement was verified for the question concerning the frequency of binge eating episodes, 94.9% for the use of laxatives, 97.4% for the use of diuretics, 94.3% for self-induced vomiting, and 74.3% for the habit

Table 1 - Distribution of adolescents regarding nutritional status, gender and age

Nutritional status**	Gender				p value*
	Male		Female		
	n	%	n	%	
Eutrophic	44	77.2	108	80.0	0.35
Underweight	5	8.8	4	3.0	
Overweight	7	12.3	21	15.5	
Obesity	1	1.7	2	1.5	
Total	57	100	135	100	

	Age range (years)				p value*
	< 16 years		≥ 16 years		
	n	%	n	%	
Eutrophic	54	78.3	98	79.7	0.98
Underweight	3	4.3	6	4.9	
Overweight	11	15.9	17	13.8	
Obesity	1	1.5	2	1.6	
Total	69	100	123	100	

* Chi-Square test; ** There was a loss of 3 adolescents in the nutritional evaluation

of going on strict diets or fasting.

No significant difference was verified in the weighted kappa values between gender and age range for the most frequent behaviors as binge eating and the habit of going on strict diets, nonetheless there was a tendency toward higher stability in the answers of females and among respondents above 15 years of age, especially toward binge eating (Table 3).

Discussion

This study evaluated the reliability of a simplified questionnaire to investigate risk behaviors for ED, adapted from another one used in the Australian population⁷. Evaluating reliability is the first step to know the accuracy of an instrument. One of the ways of evaluating the reliability of self-administered questionnaires is through test – retest. This technique allows researchers to evaluate if similar results are reproduced under the same methodological conditions, but at different moments^{12,14}.

The results showed that the agreement of the questions related to the most frequent

risk behaviors for ED, such as the habit of going on strict diets and / or fasting or binge eating episodes, was moderate. In counterpart, the questions in reference to less frequent behaviors, as the use of laxatives, diuretics, and / or self-induced vomiting presented kappa values, when adjusted for frequency, corresponding to a quasi-perfect agreement. This agreement, however, would be underestimated if weighted kappa test were used, given, in that case, values would be low (lower than 0.50), indicating poor agreement.

The kappa coefficient, simple or weighted, has been widely used and is considered an adequate statistical method for reliability technique studies and tools that generate categorical variables¹⁵. However, the limitations of this coefficient for low frequency outcomes have been identified^{12,13}. The high frequency of answers concentrated in a single category of the variable being evaluated (in the present study “not once”, which indicated low frequency of compensatory mechanisms) made the values of the weighted kappa to become close to zero. This would indicate, wrongly, low

Table 2 - Frequency of risk behaviors for eating disorders obtained in the 1st application (test) and 2nd application (retest) of the questionnaire, kappa values (and 95% confidence intervals).

Frequency a week	0*	Less than once	Once	Two or more times
Binge eating				
0*	77	8	2	2
Less than once	19	18	5	0
Once	10	11	8	4
Two or more times	6	5	6	14
Weighted Kappa (CI 95 %) = 0,50 (0,40-0,60); Adjusted Kappa for frequency = 0,47				
Use of laxatives				
0*	183	2	3	---
Less than once	2	2	1	---
Once	---	---	---	---
Two or more times	0	1	1	---
Weighted Kappa (CI 95 %) = 0.44 (0.21-0.67); Adjusted Kappa for frequency = 0.93				
Use of diuretics				
0*	189	---	1	0
Less than once	3	---	0	0
Once	---	---	---	---
Two or more times	1	---	0	1
Weighted Kappa (95 %CI)= 0.42 (-0.12-0.96); Adjusted Kappa for frequency = 0.97				
Self-induced vomiting				
0*	182	3	1	---
Less than once	4	2	1	---
Once	1	0	0	---
Two or more times	0	1	0	---
Weighted Kappa (95 %CI) = 0.34 (0.11-0.57); Adjusted Kappa for frequency = 0.92				
Strict dieting				
0*	119	3	4	2
Less than once	9	7	2	1
Once	4	2	5	2
Two or more times	6	4	11	14
Weighted Kappa (95 %CI) = 0.61(0.51-0.71); Adjusted Kappa for frequency = 0.66				

* No occurrence of behavior in the past 3 months

levels of reproducibility of the tool. The use of frequency-adjusted kappa corrects this and avoids mistaken conclusions about the quality of the tool analyzed¹². Indeed, for the variables use of laxatives, use of diuretics and self-induced vomiting, less frequent behaviors among the adolescents studied, the use of kappa statistics adjusted for frequency more than doubled kappa values

and underscored the quality of the questionnaires agreement in the investigation of these events.

The answers to the inquiries in the questionnaire revealed to be stable and the agreement varied between moderate and almost perfect. The question related to binge eating episodes was the only one that presented moderate reliability. This is probably due to

Table 3 - Weighted kappa values and 95% confidence intervals, stratified in age range and gender for binge eating and strict dieting.

Risk behaviors for ED	Gender		Age range	
	Males (n=58)	Females (n=137)	< 16 years (n=69)	≥ 16 years (n=126)
Binge Eating	0.39 (0.19-0.59)	0.54 (0.43-0.65)	0.37 (0.20-0.54)	0.57 (0.46-0.68)
Strict Dieting	0.60 (0.33-0.86)	0.58 (0.47-0.70)	0.60 (0.42-0.77)	0.61 (0.49-0.73)

Statistical significance of the presented weighted kappa values: $p < 0.001$

the fact that in spite of the question being in conformity with the description of a binge eating episode, according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)¹⁶, the consensus among professionals of the area is that it is very fragile for describing this phenomenon, because of the difficulty in quantifying “overeating episodes”. Considering the subjectivity of the question, it might be more appropriate to include other questions, which probably would improve the reliability of the questionnaire to investigate this behavior, especially among younger subjects.

Most of the studies that aimed at evaluating the reliability of tools for the investigation of ED^{17, 18, 19, 20} tested long and complex questionnaires applied to a young adult population, for example young university students whose level of comprehension is higher than that of the adolescents here evaluated, which limits the comparison to our results. Another limitation is the fact that the reliability of individual questions in the questionnaire was assessed for each risk behavior, and not with a diagnosis made on the basis of a sum of points (score), as is the case in most questionnaires. However, bearing the methodological differences in mind, some comparisons are relevant. Luce & Crowther²¹, while investigating the reliability of scores obtained through the Eating Disorder Examination Questionnaire (EDE-Q) in 139 American university students with an average age of 18.5 years, also using 15-day interval between test and retest, found good correlation (coefficient of Pearson) among the answers to the questionnaire items related to binge eating episodes ($r = 0.68$), self-induced vomiting ($r = 0.92$), use

of laxatives ($r = 0.65$) and use of diuretics ($r = 0.54$). These authors concluded that the questionnaire had good stability for tracking ED in the referred population.

Waadegaard et al.²², in order to evaluate the reliability of a questionnaire called Risk Behaviors Related to Eating Disorders (RiBED-8), applied it to 49 Danish adolescents in the age range of 14-21 years, found values of $k=0.71$ ($p=0.001$), and considered the tool to have very good stability for epidemiological studies. A similar agreement, with kappa values close to the ones found in the present study ($k=0.67-0.82$), was also reported by Moya et al.²³, in a research on the reliability of questionnaires present in the Session of Eating Disorders of the Development and Well-Being Assessment (DAWBA) validated by these authors for the diagnosis of ED in Brazilian adolescents in the 7-17 years age range.

The fact that the stability of the question on the habit of going on strict diets being a little better for males when compared to females is in agreement with Johnson et al.²⁴, when they studied the reliability of the Questionnaire on Eating and Weight Patterns - revised (QEWP), version for adolescents, in a sample of 106 American students of a public school, in the age range of 12-18 years. With the aim of evaluating the QEWP's reliability and of verifying if there was any gender difference, the authors drew the conclusion that this questionnaire presented good stability for males; however, it should be used with more caution for females.

The reliability of the questionnaire evaluated in the present study was good to track the frequency of strict dieting and to investigate infrequent behaviors such

as the use of diuretics, laxatives and self-induced vomiting, and moderate to assess the frequency of binge eating episodes. In conclusion, considering that this questionnaire is shorter and simpler than others, and even so, it covers issues on the main behaviors that indicate the presence of risk for the development of ED, this questionnai-

re can be useful in epidemiological studies to approach the problem in young people. Even without the objective of obtaining a complete diagnosis of ED, the early identification of risk behaviors could already help reduce the progression of such undesirable behaviors and its deleterious consequences on health.

References

1. Granillo BA, Jones-Rodriguez MPH, Carvajal SC. Prevalence of eating disorders in Latin adolescents: associations with substance use and other correlates. *J Adolesc Health* 2005; 36: 214- 220.
2. Cordás TA, Salzano FT, Rios SR. Os Transtornos Alimentares e a evolução no diagnóstico e no tratamento. In: Philippi ST, Alvarenga M. (org.) *Transtornos Alimentares*. Barueru: Editora Manole; 2004. p.39 – 62.
3. Nunes MA, Olinto MTA, Barros FC, Camey S. Influência da percepção do peso e do índice de massa corporal nos comportamentos alimentares anormais. *Rev Bras Psiquiatr* 2001; 23: 21-7.
4. Nunes MA, Barros FC, Anselmo OMT, Camey S, Mari JD. Prevalence of abnormal eating behaviors and inappropriate methods of weight control in young women from Brazil: a population-based study. *Eat Weight Disord* 2003; 8:100-6.
5. Miotto P, Coppi M, Frezza M, Preti A. The spectrum of eating disorders: prevalence in area of Northeast Italy. *Psychiatry Res* 2003; 119: 145-154.
6. Freitas S, Gorenstein C, Appolinario JC. Instrumentos para avaliação dos transtornos alimentares. *Rev Bras de Psiquiatria* 2002; 24: 34 – 38.
7. Hay P H. The Epidemiology of Eating Disorder Behaviors: An Australian Community-Based Survey. *J Eat Disord* 1998; 23: 371 – 382.
8. Fairburn CG, Beglin SJ. Assessment of eating disorders: interview or self-report questionnaire? *Int J Eat Dis* 1994; 16: 363 – 370.
9. Cole JT, Bellizzi CM, Flegal MK, Dietz HW. Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 2000; 320: 1240-1243.
10. World Health Organization Expert Committee. Physical status: the use and interpretation of anthropometry. WHO Technical Report Series, Geneva: WHO, 1995; 854: 263-344.
11. Abramson JH, Gahlinger PM. Computer programs for epidemiologists: PEPI, version 3. London: Brixton Books, 1999.
12. Campos MA, Leal MC, Souza PR, Cunha CB. Consistência entre fontes de dados e confiabilidade interobservador do Estudo da Morbi-mortalidade e Atenção Peri e Neonatal no Município do Rio de Janeiro. *Cad Saúde Pública* 2004; 20: 34 – 43.
13. Silva EF, Pereira MG. Avaliação das estruturas de concordância e discordância nos estudos de confiabilidade. *Rev Saúde Pública* 1998; 32: 383 – 93.
14. Griep RH, Chor D, Faerstein E, Lopes C. Confiabilidade teste-reteste de aspectos da rede social no Estudo Pró – Saúde. *Rev Saúde Pública* 2003; 37: 379 – 85.
15. Raggio Luiz R. Associação Estatística em epidemiologia. In: Medronho RA (org); Carvalho DM, Bloch KV, Raggio Luiz R, Werneck GL. *Epidemiologia*. Rio de Janeiro: Atheneu; 2004. p. 309 – 334.
16. DSM-IV – *Diagnostic and Statistical Manual of Mental Disorders* - Manual Diagnóstico e Estatístico de Transtornos Mentais. Trad. Dayse Batista, 4. Ed. Porto Alegre: Artes Médicas, 1995.
17. Rizvi SL, Peterson CB, Crow SJ, Agrad S. Test-Retest Reliability Of The Eating Disorder Examination. *Int J Eat Disord* 2000; 28: 311 – 316.
18. Perry L, Morgan J, Reid F. Screening for Symptoms of Eating disorder: reliability of the SCOFF Screening Tool with Written Compared to Oral Delivery. *Int J Eat Disord* 2002; 32: 466 – 472.
19. Ghaderi A, Scott B. The Preliminary Reliability a Validity of the Survey for Eating Disorders (SEDS): a Self-report Questionnaire of Diagnosing Eating Disorders. *Eur Eat Disorders Rev* 2002; 10: 61-76.
20. Bauer S, Winn S, Schmidt U, Kordy H. Construction, Scoring and Validation of the Short Evaluation of Eating Disorders (SEED). *Eur Eat Disorders Rev* 2005; 13: 191 – 200.

21. Luce KH, Crowther JH. The Reliability of the Eating Disorder Examination - Self-Report Questionnaire Version (EDE-Q). *Int J Eat Disord* 1999; 25: 349–351.
22. Waadegaard M, Thoning H, Petersson B. Validation of a Screening Instrument for Identifying Risk Behaviour Related to Eating Disorders. *Eur Eat Disorders Rev* 2003; 11: 433 – 455.
23. Moya T, Fleitlich-Bilyk B, Goodman R, Nogueira FC, Focchi PS, Nicoletti M, et al. The Eating Disorders Section of the Development and Well-Being Assessment (DAWBA): development and validation. *Rev Bras Psiquiatr* 2005; 27: 25-31.
24. Johnson WG, Kirk AA, Reed AA. Adolescent Version of the Questionnaire of Eating and Weight Patterns: Reliability and Gender Differences. *Int J Eat Disord* 2001; 29: 94–96.