Evidence that the Clinical Impairment Assessment (CIA) subscales should not be scored: Bifactor modelling, reliability, and validity in clinical and community samples

Running Head: CLINICAL IMPARIMENT ASSESSMENT

Abstract

Aim: The Clinical Impairment Assessment (CIA 3.0; Bohn et al., 2008) is the most widely used instrument assessing psychosocial impairment secondary to eating disorder symptoms. However, there is conflicting advice regarding the dimensionality and optimal method of scoring the CIA. We sought to resolve this confusion by conducting a comprehensive factor analytic study of the CIA in in a community sample (N = 301) and clinical sample comprising patients with a diagnosed eating disorder (N = 209). Convergent and discriminant validity were also assessed. Method: The CIA and measures of eating disorder symptoms were administered to both samples. Results: Factor analyses indicated there is a general impairment factor underlying all items on the CIA that is reliably measured by the CIA Global score. CIA Global demonstrated good convergent and discriminant validity. Conclusions: CIA Global is a reliable and valid measure of psychosocial impairment secondary to eating disorder symptoms however subscale scores should not be computed.

Keywords: clinical impairment assessment; eating disorders; factor structure; validity

Evidence that the Clinical Impairment Assessment (CIA) subscales should not be scored: Bifactor modelling, reliability, and validity in clinical and community samples

The Clinical Impairment Assessment (CIA v. 3.0; Bohn et al., 2008) is the most widely used instrument assessing psychosocial impairment secondary to eating disorder cognitions and behaviours. The instrument asks respondents to rate the extent to which their eating habits, exercising, or feelings about eating, shape or weight have resulted in particular psychosocial consequences (e.g., "made you feel ashamed of yourself") over the past 28 days. Measuring impairment is important given that the presence of clinically significant impairment secondary to eating disorder symptoms is required to diagnose an eating disorder (Diagnostic and Statistical Manual for Mental Disorders, Fifth Edition; DSM-5; American Psychiatric Association (APA), 2013).

The CIA is appealing to researchers and clinicians due to its ability to capture impairment resulting directly from symptoms of the eating disorder. As such, it is frequently used in randomised controlled trials (e.g. Fairburn et al., 2015) and effectiveness studies (e.g., Turner, Marshall, Stopa, & Waller, 2015) as a key outcome variable assessing the degree to which psychosocial impairment reduces over treatment. The CIA has also been used in diverse community samples, including university students, adult men and women, adolescents, and women at-risk of developing an eating disorder. Norms for the CIA have now been established for the United Kingdom (Bohn et al., 2008; Jenkins, 2013) and the United States (Vannucci et al., 2012), and norms for translated versions of the instrument have been established for Norway (Reas et al., 2010; Ro, Bang, Reas, & Rosenvinge, 2012), Sweden (Welch et al., 2011), and Fiji (Becker et al., 2010), highlighting the wide appeal of the instrument cross-culturally. Therefore, it is important to establish that the psychometric properties of the CIA are sound and replicable.

CIA Dimensionality and Scoring

Despite wide use of the CIA, there is conflicting advice over what aspects of impairment the CIA measures, and how it should be scored. Bohn et al. (2008, p. 1105) stated that the CIA is designed to measure "impairment overall and in three specific domains (personal, cognitive, social)." Examples of items tapping the three domains are "made you feel like a failure" (personal impairment), "stopped you going out with others" (social impairment), and "made you forgetful" (cognitive impairment). Bohn et al. computed *subscale* scores for the three domains, and also averaged all items to form a *global* score. In contrast, Bohn and Fairburn (2008a) noted that "The purpose of the CIA is to provide a simple single index … to measure the *overall severity* of secondary psychosocial impairment" (p. 315-16, emphasis in original) and made no mention of subscale scores. In subsequent research, a variety of approaches have been used, such as describing and reporting the three domain scores only (e.g., Martin et al., 2015), describing the three domains but reporting only the global score (e.g., Hogdahl, Levallius, Bjorck, Norring, & Birgegard; 2016), and describing and reporting only the global score (e.g., DeJong et al., 2013). The inconsistency in how the CIA is scored appears to stem from a lack of clarity as to whether the instrument measures *several* distinct forms of impairment, or whether all items in the scale are indicators of *one* common underlying mechanism.

Factor analysis can help clarify how many forms of impairment the CIA measures, and provide guidance on how it should be scored. To date, six factor analytic studies investigating the CIA have been conducted (Bohn et al., 2008; Jenkins, 2013; Parker, Mitchell, O'Brien, & Brennan, 2015; 2016; Martin et al., 2015; Reas et al., 2010; Vannucci et al., 2012). A strength of confirmatory factor analysis (CFA) is that the statistical fit of several factor models can be compared in order to determine the optimal model. However, this feature of CFA has not been utilised in studies to date. Only a three-factor model has been considered and found to have adequate fit (e.g., Jenkins, 2013), leading the authors to

conclude that three factors underlie the CIA. This conclusion may be premature, given the three-factor model has not been shown to be better than alternatives such as a single-factor (unidimensional) model.

Another limitation of studies that have modelled CIA responses using three factors is that the size of the inter-factor correlations has never been reported. The magnitude of the inter-factor correlations has important implications for how the CIA should be scored. Weak correlations would imply that computing subscales is sound (as each would measure a distinct aspect of impairment), and also that calculating a global CIA score is unjustified (as it would involve combining items that measure three unrelated constructs). In contrast, large correlations would suggest the constructs being measured by each subscale overlap greatly, and that it would be most parsimonious to calculate only a global score.

While some authors have concluded the CIA measures three distinct factors (e.g., Jenkins, 2013; Reas et al., 2010), other evidence indicates the CIA may measure a single, general impairment factor. For example, a ratio of the first to second eigenvalues greater than 4:1 is often regarded as evidence that an instrument primarily measures a single construct (Reeve et al., 2007). Bohn et al. (2008) conducted a Principal Components Analysis and reported that the first two eigenvalues were 9.52 and 1.12, which is a ratio of 8.5:1. Furthermore, Bohn et al. modelled the CIA using a unidimensional Item Response Theory model and found that it provide a good fit to that data.

At first glance it seems contradictory that all CIA items load on a single impairment factor, yet analyses using a three-factor model also demonstrate adequate fit. However, such findings would be expected if the CIA has a *bifactor* structure. A bifactor model comprises a *general* factor that influences responses to every item, and *group* factors that influence responses to subsets of items (see Figure 1). Bifactor models have been found to provide an

excellent fit to many psychological instruments (Rodriguez, Reise, & Haviland, 2016a). To date, no factor analyses of the CIA have used bifactor models.

The CIA is a good candidate for bifactor modelling. Every item in the scale is designed to measure clinical impairment – thus there should be a *general impairment* factor on which every item loads. However the general impairment factor alone may not adequately model CIA responses due to the existence of item clusters that tap personal, social, and cognitive impairments. Items within each cluster should be more strongly correlated with each other than they are with items from other clusters. Modelling CIA responses using the general impairment factor and additional cognitive, social, and personal *group factors* may better account for associations between the items than using the general factor alone, or using only the three group factors.

An appealing feature of applying a bifactor model to the CIA is that the relative importance of the general and group factors can be quantified and used to determine whether global and/or subscale scores should be computed (Reise, Bonifay & Haviland, 2012). If the group factors explain little variance in CIA scores it would indicate that only a global score should be computed, whereas if they exert a strong influence on participants' responses it would suggest subscales should be computed. Several statistical indices can be calculated for bifactor models to determine whether global or subscale scores should be computed (Reise, Moore, & Haviland, 2010; Rodriguez, Reise, & Haviland, 2016a,b).

Aims of the Current Study

In the present study we sought to resolve confusion about the dimensionality and optimal method of scoring the CIA by conducting a comprehensive factor analytic study of the CIA in two samples – one clinical, and one community. We were particularly interested in determining whether subscale scores should be computed in addition to, or instead of, a global score. We also assessed the convergent and discriminant validity of the CIA. We

hypothesised that: 1) a bifactor model with a strong general factor would provide a better fit to the data than unidimensional or three-factor models; 2) CIA scores would be significantly positively correlated with scores on measures of eating disorder psychopathology in both samples; and 3) patients with a diagnosed eating disorder would have significantly higher CIA scores than individuals in the community.

Method

Participants and Procedures

Community and clinical samples were collected. The community sample consisted of university students while the clinical sample was recruited from a specialist eating disorders service and met diagnostic criteria for an eating disorder. Demographic characteristics are presented in Table 1.

Community sample.

Participants (N = 301) were university students who ranged in age from 17-54 years, with a mean age of 21.3 years (SD = 5.3 years), who were recruited via online advertisements placed on university websites. Participants first read an information sheet in which they were informed of the confidential nature of their responses and their right to withdraw without prejudice. After indicating consent, participants completed the online survey and received course credit in exchange for their participation. This study received approval from the Institution's Human Research Ethics Committee (Approval Number HR161/2014).

Clinical sample.

Participants (N = 209) were individuals with a diagnosed eating disorder (DSM-5; APA, 2013) who completed the CIA as part of their assessment at a specialist public mental health service with a dedicated eating disorders programme. Individuals with binge eating disorder (BED) are not treated at the clinic and are referred elsewhere. Individuals ranged in age from 16 to 73 years, with a mean age of 25.3 years (SD = 8.6 years). The clinical group

was recruited via two pathways. Some patients (*n* = 98) completed the CIA at assessment as part of routine clinical practice, comprising individuals with principal DSM-5 diagnoses of AN (16%), BN (54%), and OSFED (30%). This study received approval from the Institution's Human Research Ethics Committee (Approval Number QI 2014/21). The remaining patients (*n* = 111) completed the CIA at assessment as part of an anorexia nervosa (AN) randomised controlled trial (RCT), and all met DSM-5 criteria for AN. This study received ethics approval from multiple institutions (see Andony et al., 2014). Exclusion criteria across both clinical recruitment pathways included current psychosis, schizophrenia, or schizoaffective disorder, significant alcohol or substance abuse/dependence, and a BMI below 14 kg/m2. Only patients who provided written informed consent for their data to be used were included.

Measures

The Clinical Impairment Assessment 3.0 (CIA; Bohn & Fairburn, 2008a). The CIA is a 16-item self-report measure that assesses impairment experienced over the past 28 days secondary to eating disorder symptoms. Each item is rated on a 4-point Likert scale ranging from 0 ("not at all") to 3 ("a lot"), with CIA Global scores ranging from 0-48.

Eating Disorder Examination interview (EDE Version 12; Fairburn & Cooper, 1993). Clinicians administered the EDE to all clinical patients to assist in yielding a reliable eating disorder diagnosis. Clinicians were Clinical Psychologists trained in administration of the EDE and specializing in eating disorder treatment. The EDE Total score assessed the CIA's convergent validity in the clinical sample. The EDE has good convergent and concurrent validity, good inter-rater reliability, and discriminates well between groups with and without an eating disorder (Berg, Peterson, Frazier, & Crow, 2012; Fairburn & Cooper, 1993).

The Eating Disorder Examination-Questionnaire Version 5 (EDE-Q; Fairburn & Beglin, 1994). The EDE-Q was administered to all individuals in the community sample and

to clinical patients recruited via the community mental health clinic (n = 98). The EDE-Q is a self-report measure of eating disorder psychopathology that assesses symptoms experienced over the past 28 days. The EDE-Q was administered prior to the administration of the CIA, as recommended by Bohn and Fairburn (2008b), to ensure that eating disorder features are at the forefront of the respondents mind. The EDE-Q Global score assessed the CIA's convergent validity in the community and clinical samples. EDE-Q Global ranges from 0 to 6, with higher scores indicating more severe psychopathology. EDE-Q Global has acceptable reliability and validity (Berg, Peterson, Frazier, & Crow, 2012). Internal consistency of EDE-Q Global was high in the clinical (α = .87) and community (α = .92) samples.

Body Mass Index (BMI). In the community sample, BMI was computed based on self-reported height and weight collected using the EDE-Q. In the clinical sample, a clinician measured patients' height at assessment and weighed individuals (clothed but without shoes or outer garments) to calculate objective BMI.

Data Analysis

We conducted a series of analyses to evaluate the dimensionality of the CIA, using factor analytic techniques recommended in recent psychometric literature (Flora, LaBrish, & Chalmers. 2012; Rhemtulla, Brosseau-Lairrd & Savalei, 2012; Rodriguez, Reise, & Haviland, 2016a,b). We evaluated the statistical fit of four CFA models (see Figure 1) that were selected based on theory and findings of prior factor analyses of the CIA. These models were: a unidimensional model; uncorrelated and correlated three-factor models, and a bifactor model in which the three group factors corresponded to the Personal, Social, and Cognitive factors proposed by Bohn et al. (2008). Including correlated *and* uncorrelated three-factor models was important because prior researchers have not specified whether they allowed the factors to covary, and these two models have distinctly different implications for how the CIA should be scored. The CFAs were conducted using the *R* package 'lavaan' (Rosseel,

2012) using the WLSMV estimator. Models were identified by fixing factor variances to 1. Statistical fit was examined using the Comparative Fit Index (CFI ≥.95 indicative of good fit), Tucker-Lewis Index (TLI ≥.95 indicative of good fit), Root Mean Square Error of Approximation (RMSEA ≤.06 indicative of good fit) and Standardized Root Mean Residuals (SRMR ≤.08 indicative of good fit, Hu & Bentler, 1999)[1]. Differences in fit between nested models were tested using scaled Chi-Square difference tests.[2] Judgements about which factor model to retain were based on multiple criteria (global fit statistics, difference tests, local fit diagnostics, bifactor indices, and theoretical reasoning), as recommended by Schmitt (2011). Analyses were conducted separately for clinical and community samples.

For the bifactor model, we computed several indices recommended by Rodriguez, Reise, and Haviland (2016a,b). *Omega hierarchical* is a reliability coefficient that measures the proportion of variance in CIA global scores that is attributable to *each* common factor in a bifactor model. *Omega total* is proportion of variance in CIA global scores accounted for collectively by *all* of the factors. *Omega subscale* and *Omega subscale hierarchical* reliability coefficients were also computed for subscales. When Omega hierarchical is high (e.g., > .80), composite scores can be considered essentially unidimensional, because the vast majority of variation in scores is due to a single source (Rodriguez et al., 2016a).

The extent to which it is feasible to model bifactor data using a simple unidimensional model in SEM can be assessed through two indices – the *Explained Common Variance* (ECV) and the *Proportion of Uncontaminated Correlations* (PUC). ECV is the variance explained by the general factor, divided by the variance explained by the general *and* group factors. It provides an indication of 'how unidimensional' an instrument is. PUC is the number of correlations between items from different group factors, divided by the total number of correlations between items in the factor model. When ECV and PUC both exceed > .70, any bias in parameter estimates caused by ignoring the group factors will be minimal

(Rodriguez, 2016a). Another way of assessing the potential bias in parameter estimates from using a unidimensional rather than full bifactor model is to calculate the *mean relative bias* of the factor loadings. This involves comparing the loadings on the general factor in a bifactor model, against those from a unidimensional model. For example, imagine the standardized loading for an item was .80 on the general factor in the bifactor model, and .82 in the unidimensional model. The difference between the loadings (known as the *relative bias*) is .02 – a negligible amount. The mean relative bias is the average (absolute) difference in loadings for all items in a scale. Finally, *Coefficient H* is a reliability index that provides information about how well each latent variable in a bifactor model is measured by its indicators. Factors with *H* values below .70 are poorly defined by their indicators, and of questionable utility (Rodriguez, 2016b).

Other analyses were conducted with SPSS v. 18.0 (SPSS Inc., Chicago, Illinois). Pearson correlations between CIA Global and eating disorder psychopathology (EDE-Q Global and EDE Total) assessed the CIA's convergent validity. Independent samples *t*-tests were performed to compare CIA Global between community and clinical samples. Cohen's *d* effect sizes were computed using the online calculator at https://www.uccs.edu/~lbecker/.

Results

Confirmatory Factor Models

Robust fit statistics are presented in Table 2. The unidimensional model had acceptable CFI and TLI values, but the RMSEA and SRMR exceeded guidelines for good fitting models. The fit of the three uncorrelated factor model was extremely poor in both community and clinical samples, with low CFI and TLI values and very high RMSEA and SRMR values. The correlated three-factor model produced good fit statistics, except for the RMSEA. Factor intercorrelations were very high in both the community (r = .87, .86, and .71) and clinical samples (r = .78, .78, .74), suggesting an unmodelled general factor is

responsible for the high inter-correlations. The fit of the bifactor model was excellent in both samples. Scaled chi-square difference tests indicated that the observed data fit the bifactor model statistically significantly better ($p < .001$) than the unidimensional model and the three uncorrelated factor model, and that the fit of the bifactor and correlated traits model was not significantly different (clinical sample $p = .25$, community sample $p = .46$).

Bifactor Indices

Bifactor indices are presented in Table 3. The Explained Common Variance (ECV) values demonstrate that the general factor accounted for more than 90% of the explained common variance in both samples. This indicates that the CIA *primarily* measures a single, general impairment factor, despite *technically* being a multidimensional instrument. Omega reliability coefficients for the CIA global score and each subscale were very high (> .90). Omega-hierarchical values (.91 community, .88 clinical) were also high, whereas all Omega-subscale values were very low (< .06). These findings indicate that CIA Global is a reliable measure of the general impairment factor, but that the subscales are unreliable measures of the group factors.

Factor Loadings

Standardized factor loadings for the bifactor model are presented in Table 4. All items had high loadings on the general factor. As a comparison, loadings for the unidimensional solution are also presented. The discrepancy between these loadings is measured by the *mean relative bias* statistic, which was low (.07) for both samples, providing further evidence that the CIA primarily measures a general impairment factor.

Associations between CIA and eating disorder psychopathology

As predicted, positive correlations were observed between CIA Global and EDE-Q Global in both the community ($r = .86$, $p < .001$) and clinical ($r = .67$, $p < .001$) samples.

Positive correlations were also observed between CIA Global and EDE Total in the clinical sample (r = .54, p < .001).

Comparisons between clinical and community samples

Patients with a diagnosed eating disorder had significantly higher CIA Global (M = 34.0, SD = 10.6) than individuals in the community (M = 11.0, SD = 10.4), d = 2.2, $t(508)$ = 24.27, p < .001, and also demonstrated significantly higher EDE-Q Global and EDE Total scores (see Table 5).

Discussion

The present study sought to resolve confusion about the dimensionality and optimal method of scoring the CIA, and to assess the convergent and discriminant validity of the CIA in clinical and community samples. As predicted by hypothesis one, a bifactor model with a strong general factor provided a better fit to the data than viable alternative models. The present findings indicate that the CIA is *technically* multidimensional, but that it *primarily* measures a strong general impairment factor. This has several implications. First, CIA Global scores are a reliable measure of the overall severity of impairment due to eating disorder features, and should be routinely reported. Second, subscales should not be scored as they primarily measure the general factor rather than specific cognitive, personal, or social impairments. Third, because the general factor is strong and the group factors are weak, it is acceptable for researchers using the CIA in SEM analyses to ignore the group factors and model the CIA using a single factor (Reise, Scheines, Widaman, & Haviland, 2013). This is simpler than fitting a bifactor model, and should not substantively affect the results of SEM analyses. Future research could (i) assess the possibility that a shorter version of the CIA could reliably measure clinical impairment secondary to eating disorder symptoms, and (ii) test the factorial invariance of solutions obtained from different populations.[3]

As predicted by hypothesis 2, and consistent with previous research, convergent validity of the CIA was supported, as CIA Global exhibited moderate to strong positive associations with eating disorder severity (EDE-Q Global and EDE Total) in the community and clinical samples. As predicted by hypothesis 3, and consistent with previous research, CIA Global discriminated well between clinical and community samples. Notably, the mean CIA Global score for the community sample (M = 11.0) was higher than reported in previous community studies, including a sample of high-risk females (M = 8.3; Vannucci et al., 2012). This discrepancy may reflect the high proportion of individuals in the present community sample who endorsed having had a diagnosed eating disorder (3.2%) or who believed they may have had an undiagnosed eating disorder (14.4%). The proportion of individuals in the present sample community endorsing eating disorder symptoms is consistent with Jenkins (2013), who found that 14.8% of their university sample had spoken to a health professional about eating, shape, or weight concerns. It is also consistent with wider literature demonstrating high prevalence of eating disorders amongst university students, with 20% to 34% of individuals self-reporting an eating disorder diagnosis (e.g. National Eating Disorders Association, 2006; White, Reynolds-Malear, & Cordero, 2011). The mean CIA Global in the present clinical sample was equivalent to the highest mean reported in the literature (Welch et al., 2011), suggesting that individuals were highly impaired by their symptoms when they presented for treatment. The community clinic in this study is the only public outpatient service specialising in treatment of eating disorders in the state and, due to long waitlists, treatment may not be available for several months. Further research is needed to determine whether similar levels of clinical impairment are evident in individuals who seek treatment in other settings, such as inpatient or private practice settings where treatment may be accessed more promptly.

The present study has several strengths. It is the first time several factor models have been compared using CFA and the first study to use bifactor modelling. Findings from our clinical sample were replicated in a community sample. This is important, as the CIA is frequently used with non-clinical samples, despite originally being designed for use with clinical populations. Our clinical sample was larger and had a higher proportion of individuals diagnosed with Anorexia Nervosa than previous studies (e.g., Bohn et al, 2008). Furthermore, all individuals in the clinical sample were diagnosed with eating disorders, whereas some previous studies have included subthreshold cases (e.g., Vannucci et al., 2012), or focused on individuals with obesity (Parker et al., 2016).

The current study has notable limitations. The community sample comprised adult university students therefore results may not generalise to other community samples. The clinical sample did not include individuals with BED or individuals younger than 16 years of age. Further research is needed to examine the factor structure, reliability, and validity of the CIA in patients with BED and younger patients with eating disorders to determine the replicability of the present findings. The present study did not include an objective measure of impairment, although previous research has demonstrated strong positive correlations between CIA Global and clinician-rated impairment (Bohn et al., 2008). Bifactor models also have limitations, such as a tendency to overfit data (Binfay, Lane, & Reise, 2017). For this reason it is important to consider alternative lines of evidence, such as bifactor indices, when evaluating the fit of a bifactor model – as was done in this study. Another limitation of bifactor models is and that the emergence of a strong general impairment factor does not necessarily imply there is a single latent mechanism that causes impairment in people with eating disorders. A strong general factor may just indicate that the observable signs of impairment – as measured by CIA items – are so highly correlated that it is not possible (or necessary) to subdivide them into smaller, discrete categories (Bonifay et al , 2017). It is also

important to acknowledge that there may also be additional aspects of impairment not assessed by the CIA that are more separable from the common factor. Bifactor analysis can only speak to the structure of the particular *measure* under study, not to the underlying *construct*.

In sum, the present paper offers clear guidelines for scoring and interpretation of the CIA. We recommend that CIA Global is routinely reported in future studies as a valid and reliable measure of clinical impairment secondary to eating disorder symptoms. However, CIA subscale scores should not be computed.

Footnotes

[1] We acknowledge that these 'rules of thumb', while commonly used, have important limitations and should not be the sole way in which model fit is assessed (see Kline, 2015).

[2] This is equivalent to using DIFFTEST in Mplus.

[3] We attempted to conduct analyses testing measurement invariance for the bifactor model, but the metric and scalar invariance models would not converge. This is likely due to an interaction between the complexity of the bifactor model, and relatively small size of the clinical sample. Future studies examining measurement invariance are likely to require considerably larger samples than were employed in the current study.

Acknowledgements

The AN RCT was supported by the National Health and Medical Research Council 634434 and the McCusker Foundation.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC.

Andony, L.J., Tay, E., Allen, K.A., Wade, T.D., Hay, P., Touyz, S., McIntosh, V.V.W., Treasure, J., Schmidt, U.H., Fairburn, C.G., Erceg-Hurn, D.M., Fursland, A., Crosby, R.D., & Byrne, S.M, (2014). Therapist adherence in the strong without anorexia nervosa (SWAN) study: A randomized controlled trial of three treatments for adults with anorexia nervosa. *International Journal of Eating Disorders, 48*, 1170-1175. doi: 10.1002/eat.22455

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. *International Journal of Eating Disorders, 45,* 428-38.

Bohn, K., Doll, H.A., Cooper, Z., O'Connor, M.E., Palmer, R.L., & Fairburn, C.G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour Research and Therapy, 46*, 1105 -1110.

Bohn, K, & Fairburn, C.G. (2008a). Clinical Impairment Assessment Questionnaire (CIA 3.0), In Fairburn, C.B., *Cognitive Behaviour Therapy and Eating Disorders*. New York: Guildford Press.

Bohn, K, & Fairburn, C.G. (2008b). http://www.credo-oxford.com/pdfs/CIA_3.0_Instructions_for_users.pdf

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three Concerns With Applying a Bifactor Model as a Structure of Psychopathology. *Clinical Psychological Science*, *5*, 184–186. doi:10.1177/2167702616657069

Fairburn, C.G., Bailey-Straebler, S., Basden, S., Doll, H.A., Jones, R., Murphy, R., O'Connor, M.E., Cooper, Z. (2015). A transdiagnostic comparison of enhanced cognitive

behaviour therapy (CBT-E) and interpersonal psychotherapy in the treatment of eating disorders. *Behaviour Research and Therapy, 70,* 64-71.

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16,* 363-370.

Fairburn, C. G. & Cooper, Z. (1993). The Eating Disorder Examination. In Fairburn, C. G. & Wilson, G. T. (Eds), *Binge eating: Nature, assessment, and treatment.* New York: Guilford Press.

Flora, D.B., LaBrish, C., & Chalmers, R.P. (2012). Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Frontiers in Psychology, 3,* 1-21. http://dx.doi.org/10.3389/fpsyg.2012.00055

Hogdahl, L., Levallius, J., Bjorck, C., Norring, C., & Birgegard, A. (2016). Personality predicts drop-out from therapist-guided internet-based cognitive behavioural therapy for eating disorders. Results from a randomized controlled trial. *Internet Interventions, 5,* 44-50.

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

Jenkins, P.E. (2013). Psychometric validation of the Clinical Impairment Assessment in a UK eating disorder service. *Eating Behaviours, 14,* 241-243.

Kline, R. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: The Guilford Press.

Martin, J., Padierna, A., Unzurrunzaga, A., Gonzalez, N., Berjano, B., & Quintana, J. M. (2015). Adaptation and validation of the Spanish version of the clinical impairment assessment questionnaire. *Appetite, 91,* 20-27.

National Eating Disorders Association. (2013). Eating Disorders on the College Campus: A national survey of programs and resources. Retrieved from https://www.nationaleatingdisorders.org/sites/default/files/CollegeSurvey/CollegiateSurveyProject.pdf

Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2015). Psychometric evaluation of disordered eating measures in bariatric surgery candidates. *Eating Behaviours, 19,* 39-48.

Parker, K., Mitchell, S., O'Brien, P., & Brennan, L. (2016). Psychometric evaluation of disordered eating measures in bariatric surgery candidates. *Obesity Surgery, 26,* 563-575.

Reas, D.L., Ro, O., Kapstad, H., & Lask, B. (2010). Psychometric properties of the Clinical Impairment Assessment: Norms for young adult women. *International Journal of Eating Disorders, 43,* 72-76.

Reeve, BB., Hays, R.D., Bjorner, J.B., Cook, K.F., Crane, P.K., Teresi, J.A., …Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45,* S22-S31.

Reise, S.P., Bonifay, W. E., & Haviland, M.G. (2012). Scoring and Modeling Psychological Measures in the Presence of Multidimensionality. *Journal of Personality Assessment, 95,* 129-140.

Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scores. *Journal of Personality Assessment, 92,* 544-559. http://dx.doi.org/10.1080/00223891.2010.496477

Reise, S.P., Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and structural coefficient bias in structural equation modelling: A bifactor perspective. *Educational and Psychological Measurement, 73,* 5-26.

Rhemtulla, Brosseau-Liard, & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17,* 354-373.

Rodriguez, A., Reise, S. P., Haviland, M. G. (2016a). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment, 98,* 223-237.

Rodriguez, A., Reise, S. P., Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21,* 137-150. http://dx.doi.org/10.1037/met0000045

Rosseel, Y (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48,* 1-36. URL http://www.jstatsoft.org/v48/i02/

Schmitt, T. A. (2011). Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. *Journal of Psychoeducational Assessment, 29,* 304-321. doi: 10.1177/0734282911406653

Turner, H., Marshall, E., Stopa, L., & Waller, G. (2015). Cognitive-behavioural therapy for outpatients with eating disorders: Effectiveness for a transdiagnostic group in a routine clinical setting. *Behaviour Research and Therapy, 68,* 70-75.

Vannucci, A., Kass., A.E., Sinton, M.M., Aspen, V., Weisman, H., Bailey, J.O., Wilfley, D.E., & Taylor, C.B. (2012). An examination of the Clinical Impairment Assessment among women at high risk for eating disorder onset. *Behaviour Research and Therapy, 50,* 407-414.

White, S, Reynolds–Malear, J., Cordero, E. (2011). Disordered eating and the use of unhealthy weight control methods in college students: 1995, 2002 and 2008. *Eating Disorders: Journal of Treatment and Prevention, 19,* 323-334.

Table 1

Characteristics of the Clinical (N = 209) and Community (N = 301) Samples.

Baseline Variable	Clinical sample	Non-clinical sample
Age (years)		
M (SD), *range*	25.3 (8.6), *16-73*	21.3 (5.3), *17-54*
Gender, *n*		
Female: Male: Other	203:6:0	248:51:2
Body mass index[1]		
M (SD), *range*	18.8 (3.6), *13.8-37.0*	23.0 (4.7), *15.6-39.1*
Employment status, *n* (%)		
Full/part-time	135 (64.6%)	233 (77.4%)
Unemployed	65 (31.1%)	68 (22.6%)
Missing	*9 (4.3%)*	-
Ethnicity, *n* (%)		
Aust-Anglo/Europ	193 (92.3%)	225 (75%)
Aust-Asian	1 (0.5%)	37 (12.3%)
Other	14 (6.7%)	39 (13%)
Missing	*1 (0.5%)*	-
Education, *n* (%)		
Less than year 12	24 (11.4%)	-
Year 12 or equivalent	75 (35.9%)	-
Technical/trade	26 (12.4%)	-
Tertiary	61 (29.2%)	301 (100%)
Missing	*23 (11.0%)*	-
Relationship status, *n* (%)		
Single	140 (67.0%)	256 (84.2%)
Defacto/Married	53 (25.4%)	44 (14.6%)
Divorced/separated	8 (3.8%)	1 (0.3%)
Missing	8 (3.8%)	-
ED Diagnosis, *n* (%)		
AN	127 (60.8%)	1 (0.3%)
BN	53 (25.4%)	8 (2.6%)
OSFED	29 (13.9)	1 (0.3%)
Believe I have had an ED, never diagnosed	-	45 (14.5%)
ED Chronicity (years)		
M (SD), *range*	6.6 (6.7), 0.25-28	N/A

Note: ED = Eating Disorder; M = Mean; SD = Standard Deviation;
[1]BMI for community sample was computed from self-reported height and weight.
[2]ED diagnosis for community sample is based on self-report diagnosis from a health professional

Table 2.

Robust Fit Statistics for Confirmatory Factor Analysis Models

						90% CI		
Model	X	df	CFI	TLI	RMSEA	LL	UL	SRMR
Community Sample								
Unidimensional	615.8	104	.966	.961	.128	.118	.138	.105
Three Independent Factors	5800.0	104	.625	.567	.427	.418	.437	.508
Three Correlated Factors	220.2	101	.992	.991	.063	.051	.074	.044
Bifactor	147.0	88	.996	.995	.047	.033	.060	.034
Clinical Sample								
Unidimensional	504.2	104	.912	.898	.136	.124	.148	.095
Three Independent Factors	2426.6	104	.488	.409	.328	.316	.339	.417
Three Correlated Factors	230.4	101	.971	.966	.078	.065	.092	.057
Bifactor	135.5	88	.990	.986	.051	.033	.067	.041
Good Fit Guidelines			>.95	>.95	<.06			<.08

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Residuals (SRMR). LL and UL = Lower and Upper Limits of 90% Confidence Interval. Guidelines for good fitting models are from Hu & Bentler (1999). *P* values for all Chi-Square tests were < .001 except for the Bifactor model in the Clinical Sample where $p = .001$.

Table 3.

Bifactor Indices

	Community				Clinical			
Index	Gen	GR-P	GR-S	GR-C	Gen	GR-P	GR-S	GR-C
Omega	.98	.97	.96	.95	.97	.94	.93	.91
Omega H / Omega HS	.91	.05	< .01	.03	.88	.05	.01	.02
ECV	.92				.91			
PUC	.71				.71			
Mean relative bias	.07				.07			
H	.97	.66	.26	.62	.95	.67	.55	.61

Omega = Omega total / Omega subscale. Omega H/S = Omega Hierarchical (for general factor) and Omega hierarchical subscale (for group factors). ECV = Explained Common Variance. Percentage of uncontaminated correlations (PUC). H = Coefficient H Construct Reliability. Gen = General Factor. GR-P, GR-S, GR-C = Group Personal, Social and Cognitive Factors.

Table 4

Standardized Factor Loadings for Unidimensional and Bifactor Restricted (Confirmatory) Factor Solutions

		Community Sample					Clinical Sample				
		Uni	*Bifactor*				*Uni*	*Bifactor*			
	Item		Gen	GR-P	GR-S	GR-C		GEN	GR-P	GR-S	GR-C
1	….made it difficult to concentrate?	.83	.84			.30	.76	.78			.19
2	….made you more critical of yourself?	.90	.79	.47			.74	.71	.30		
3	….stopped you going out with others?	.80	.85		-.06		.78	.79		.20	
4	….affected your performance at work (if applicable)?	.71	.62			.53	.67	.65			.28
5	….made you forgetful?	.81	.69			.63	.78	.63			.70
6	….affected your ability to make everyday decisions?	.81	.76			.45	.73	.71			.30
7	….interfered with meals with family or friends?	.80	.82		.37		.76	.71		.44	
8	….made you upset?	.94	.80	.52			.81	.71	.48		
9	….made you feel ashamed about yourself?	.94	.81	.49			.82	.63	.71		
10	….made it difficult to eat out with others?	.83	.86		.34		.76	.69		.70	
11	….made you feel guilty?	.87	.74	.50			.87	.82	.38		
12	….interfered with your doing things you used to enjoy?	.84	.92		-.22		.78	.80		.16	
13	….made you absent-minded?	.85	.79			.45	.82	.73			.54
14	….made you feel a failure?	.91	.83	.41			.82	.74	.42		
15	….interfered with your relationship with others?	.89	.94		.03		.78	.82		.06	
16	….made you worry?	.88	.73	.54			.77	.70	.40		

Note. Uni = Unidimensional model loading. GEN = General Factor in Bifactor solution. GR-P = Group Personal Factor. GR-S = Group Social Factor. GR-C = Group Cognitive Factor

Table 5

Descriptives and Comparison of Community (N = 301) and Clinical (N = 209) Samples on CIA Global, EDE-Q Global, and EDE Total.

Baseline Variable	Clinical sample M (SD), *range*	Community sample M (SD), *range*	*t*-value	*p*-value	*d*
CIA Global	34.0 (10.6), *1-48*	11.0 (10.4), 0-45	24.28	< .001	2.19
EDE-Q Global	4.3 (1.2), *0.5-5.7*	1.9 (1.4), *0-5.8*	16.77	< .001	1.84
EDE Total	3.3 (1.4), *0.1-5.6*	-	-	-	

Note: Only patients from the community clinic completed the EDE-Q (*n* = 98)

Figure 1

Restricted (Confirmatory) Factor Models