J Abnorm Child Psychol (2013) 41:775–788
DOI 10.1007/s10802-013-9718-z

The Eating Attitudes Test-26 Revisited using Exploratory Structural Equation Modeling

Christophe Maïano · Alexandre J. S. Morin · Marie-Christine Lanfranchi · Pierre Therme

Published online: 24 January 2013

Abstract Most previous studies have failed to replicate the original factor structure of the 26-item version of the Eating Attitudes Test (EAT-26) among community samples of adolescents. The main objective of the present series of four studies (n=2178) was to revisit the factor structure of this instrument among mixed gender community samples of adolescents using both exploratory structural equation modeling (ESEM) and confirmatory factor analysis (CFA). First, results from the ESEM analyses provided satisfactory goodness-of-fit statistics and reliability coefficients for a six-factor model of the EAT with 18 items (EAT-18) closely corresponding to the original seven-factor structure proposed for the 40-item version of the EAT. Second, these analyses were satisfactorily replicated among a new sample of community adolescents using CFA. The results confirmed the factor loading and intercept invariance of this model across gender and age groups (i.e., early and late adolescence), as well as the complete invariance of the EAT-18 measurement model between ethnicities (i.e., European versus African origins) and across weight categories (i.e., underweight, normal weight and overweight). Finally, the last study provided support for convergent validity of the EAT-18 with the Eating Disorder Inventory and with instruments measuring global self-esteem, physical appearance, social physique anxiety and fear of negative appearance evaluation.

Keywords Disordered eating attitudes and behaviors · Eating attitudes test · EAT · Measurement · CFA · ESEM · Exploratory structural equation modeling · Measurement invariance · Norms · Reliability

Electronic supplementary material The online version of this article (doi:10.1007/s10802-013-9718-z) contains supplementary material, which is available to authorized users.

Christophe Maïano and Alexandre J. S. Morin contributed equally to this article and their order was determined at random: both should be considered first authors. Part of this manuscript was prepared while A.J.S.M was a visiting scholar at the Aix-Marseille University.

C. Maïano
Cyberpsychology Laboratory, Département de Psychoéducation et de Psychologie, Université du Québec en Outaouais (UQO), Saint-Jérôme, QC, Canada

A. J. S. Morin
Centre for Positive Psychology and Education, University of Western Sydney, Sydney, Australia

M.-C. Lanfranchi · P. Therme
UMR 7287, Institute of Movement Sciences, Etienne-Jules Marey, Aix-Marseille Univ & CNRS, Marseille, France

M.-C. Lanfranchi
Direction Régionale de la Jeunesse, des Sports et de la Cohésion Sociale - Provence-Alpes Côte-d'Azur, Marseille, France

C. Maïano (✉)
Département de Psychoéducation et de Psychologie, Université du Québec en Outaouais, Campus de Saint-Jérôme, 5 rue St-Joseph, Saint-Jérôme, QC J7Z 0B7, Canada
e-mail: christophe.maiano@uqo.ca

Disordered eating attitudes and behaviors (DEAB) are highly prevalent and some of them (i.e., weight and shape concerns, dieting, and fear of fatness) represent significant risk factors for threshold eating disorders that typically emerge in adolescence (e.g., Chamay-Weber et al. 2005; Jacobi et al. 2004; Smink et al. 2012; Stice 2002). DEAB thus represent a significant public health concern and their early screening has become an important part of efficient public health prevention programs (e.g., Golden et al. 2003).

Among the instruments that can be used to screen for DEAB among youth (Anderson et al. 2004; Micali and House 2011), the Eating Attitudes Test is one of the most widely used (EAT; Garfinkel and Newman 2001; Garner and Garfinkel 1979). The EAT is short, simple to administer and freely available (http://www.eat-26.com/). In addition, "although the EAT was designed to identify individuals with anorexia nervosa-like symptoms, it is best conceptualized as

a measure of general eating disorder pathology" (Anderson et al. 2004, p. 769) and has been shown to be efficient at detecting individuals at risk for threshold eating disorders (e.g., Carter and Moss 1984; Raciti and Norcross 1987).

The original 40-item version of the EAT was validated in a mixed community and clinical sample of Canadian women with and without anorexia nervosa (n=158) and includes seven subscales: (a) Food Preoccupation, (b) Body Image for Thinness, (c) Vomiting and Laxative Abuse, (d) Dieting, (e) Slow Eating, (f) Clandestine Eating and (g) Perceived Social Pressure to Gain Weight. Items are rated on a six-point scale ranging from *always* (6) to *never* (1), but answers are recoded into a four-point scale (0 to 3), by combining the two lowest response options.

The psychometric properties of the EAT-40 were re-examined in a second study (Garner et al. 1982) conducted among a mixed community and clinical sample of Canadian women with and without anorexia nervosa (n=300). The results from an oblique principal component analysis provided a satisfactory solution for a 26-item version of the EAT (EAT-26) with three factors: (a) Dieting (13 items); (b) Bulimia and Food Preoccupation (six items) and (c) Oral Control (seven items). Results also confirmed the convergent validity of the EAT-26 with the EAT-40 and related measures.

Since then, the EAT-26 has been extensively validated across other clinical and non-clinical subgroups from various cultural backgrounds (Eastern/Western Europe, South America, Middle East, Asia) (Garfinkel and Newman 2001; Nasser 1997). These studies mostly replicated the three-factor structure of the EAT-26 in clinical samples of adults and youths, but yielded inconsistent results among non-clinical samples of adolescents. Indeed, a literature review showed that 60 % of these studies failed to replicate the three-factor structure of the EAT-26 (see the online supplements), converging on solutions with three or four factors that differ from the original and a number of items ranging from 16 to 25. It is obvious from these studies that the original factor structure and screening properties of the EAT-26 among *community* samples of youth are questionable. This limitation is quite serious for screening practice as it means that the nature of the constructs that is assessed by this instrument remains undefined. In other words, decisions related to the identification of individuals at risk for DEAB based on their pattern of scores on the EAT-26 are hard to justify since the exact nature of these subscales remains unclear.

Previous studies of the EAT-26 in community samples of adolescents also have other important limitations. They all examined the EAT-26 based on transformed scores (ranging from 0 to 3). Yet, previous studies (Schoemaker et al. 1994; van Strien and Ouwens 2003) showed that converting six-point scales to four-point scales seriously damages instrument validity and integrity in non-clinical populations. It is thus preferable to anchor analyses and screening procedures on the six-point scales, which also provides more precise information regarding DEAB severity (e.g., Mintz and O'Halloran 2000). In practice, the four-point scales can have serious implications for the use of the instrument as a screening or clinical tool as it would reduce its ability to detect finer differences in severity levels that could be indicative of threshold or subthreshold levels of DEAB that may be worthy of clinical attention.

Most previous studies relied on single-gender samples of adolescent girls or adult women, precluding measurement equivalence tests of the EAT-26 factorial structure (i.e., measurement invariance) across genders. Similarly, none of the previous studies of the EAT-26 in community samples of adolescents provided evidence of measurement invariance according to age, ethnicity or weight categories. It is thus unknown whether the factor structure of the EAT-26 holds across genders, age groups (early vs. late adolescents), ethnicities/cultures (White/European vs. Black/African origin) and weight categories (underweight, normal weight and overweight-obese). This is worrisome since the EAT-26 is frequently used in epidemiological studies to identify youths at risk for threshold eating disorders among these various subsamples (Anderson et al. 2004; Garfinkel and Newman 2001). However, doing so requires a preliminary verification that the EAT-26 does measure DEAB, in the same manner, across these subpopulations (Vandenberg and Lance 2000). Indeed, measurement invariance tests allow us to verify whether higher DEAB scores commonly found among girls [(vs. boys) (e.g., Neumark-Sztainer et al. 2002)], older adolescents [(vs. early adolescents) (e.g., Smink et al. 2012)], White/European [(vs. Black/African) (e.g., Wildes et al. 2001)] and underweight or overweight-obese persons [(vs. normal weight) (e.g., Neumark-Sztainer et al. 2002)] truly reflect differences in DEAB levels and not differences in the measurement of the construct (Meredith 1993). Therefore, researchers and clinicians using the EAT-26 to screen for DEAB in community samples cannot be assured that this instrument is psychometrically equivalent across these populations (Meredith 1993).

A lack of invariance could lead to rather serious consequences, including erroneous treatment referrals for members of certain groups because the construct or the severity of DEAB symptoms are assessed differently in these groups. Each of the specific measurement invariance tests has serious practical implications for the estimation of possible measurement biases when applied to specific subgroups. First, the non-invariance of the factor loadings is the most severe case of measurement biases and indicates that the instrument simply does not measure the same constructs in the various subgroups. Second, non-invariance of the intercepts of the items means that, when their real levels on the construct of interest (e.g., DEAB) are identical, participants

from the various subgroups still tend to score higher or lower on the different items, thus giving the false impression of having higher or lower levels of DEAB—which could result in erroneous clinical decisions. Finally, non-invariance of the uniquenesses of the items indicates that the constructs are assessed with different levels of measurement error and precision in the various subgroups and that clinical decisions will thus be made with more or less certainty in these groups.

Moreover, all of the previous studies of the EAT-26 in non-clinical samples of youths relied on model modifications to find a well-fitting model. In such cases, a cross-validation sample is necessary to ensure that the resulting model reflects the true structure and not an overfitting of the model to the idiosyncratic characteristics of the sample and capitalization on chance (e.g., MacCallum et al. 1992).

Additionally, none of the previous studies of the EAT-26 in community samples of adolescents examined its convergent validity. It is thus unknown whether the EAT-26 relates properly to constructs commonly related to DEAB: global self-esteem, physical appearance, social physique anxiety and fear of negative appearance evaluation (Karpowicz et al. 2009; Krane et al. 2001; Maïano, Morin, Eklund et al. 2010). Regarding these last two limitations, it is clear that any use of an instrument for screening individuals at risk for DEAB in non-clinical samples needs to be based on the assurance that the instrument in question has stable properties in different community's samples and subpopulations and that it assesses what it should assess and thus relates as expected to other constructs.

Finally, all of the previous studies of the EAT-26 in community samples of youths relied on principal component analysis (PCA), rather than exploratory (EFA) or confirmatory (CFA) factor analysis. This is suboptimal when the aim is to probe for the underlying structure of psychological constructs. To reproduce total variance, including measurement error, factor analysis focuses on covariance, reflecting what is shared among the items (e.g., Fabrigar et al. 1999). However, in the past, PCA and EFA have presented major limitations when compared to CFA as it was impossible to conduct systematic measurement invariance tests with EFA and thus to systematically investigate whether the EAT-26 assessed DEAB equivalently across meaningfully different subgroups of participants (Meredith 1993). This may explain why no previous studies conducted such tests. In fact, many of the apparent advantages of CFA over EFA resulted from the integration of CFA within the structural equation modeling (SEM) framework (e.g., structural relations between latent constructs adjusted for measurement error, latent mean structures, goodness-of-fit). Fortunately, EFA was recently integrated with CFA/SEM into a global exploratory SEM (ESEM) framework (Asparouhov and Muthén 2009; Marsh et al. 2009), making the advances traditionally associated with CFA available in EFA. ESEM makes it possible to conduct all types of psychometric tests typically reserved for CFA within EFA. It is thus naturally suited to investigate the factor structure of the EAT-26. Given the greater flexibility of EFA and the observation that constraining small cross-loadings to zero in the context of CFA may result in biased factor correlations (Asparouhov and Muthén 2009), it has been argued that psychometric studies should always start from an ESEM model (Marsh et al. 2009; Morin et al. in press). Although the practical implications of this last methodological limitation are not obvious at first sight, this point is directly related to the ability, or lack thereof, of previous research that relied on suboptimal PCAs to correctly identify a stable, and invariant, factor structure for the EAT-26.

Overview of the Studies

The objective of the *first study* was to examine the factor structure and reliability of the EAT-26 using ESEM and the measurement invariance of the retained model across gender and age groups. The *second study* aimed to verify whether the results obtained in the first study were replicated in a new sample. This study also verified whether the final structure obtained in the first study truly required an ESEM model, or whether a more parsimonious CFA model could be sufficient. Afterwards, a combined sample (to maximize sample size) from the first and second studies was used to verify the measurement invariance across ethnic groups and weight categories, and to estimate point and interval estimates of percentile ranks for raw scores on the EAT subscales. Finally, the *third study* was performed to verify the convergent validity of the measurement model of the EAT with (a) another measure of DEAB (Eating Disorders Inventory; Garner et al. 1983); and (b) measures of global self-esteem, physical appearance, social physique anxiety and fear of negative appearance evaluation. The Eating Disorders Inventory was selected because it was known to be significantly correlated with the EAT (Berland et al. 1986; Garner et al. 1983; Raciti and Norcross 1987). The other constructs were chosen because current conceptions of DEAB emphasize that concern with body shape-weight and self-esteem are core aspects (Fairburn et al. 2003) that lead affected individuals to become anxious in situations where others may observe or judge their physical appearance. For all studies, participants and their parents gave written informed consent.

First Study: Exploratory Structural Equation Modeling of the EAT-26

Method

A sample of 1035 youths (aged 11–18 years; M_{age}= 14.58 years, SD_{age}=2.03) was recruited from 14 middle

and high schools in Southern France. This sample included: (a) 494 boys and 541 girls; (b) 753 participants of European origin and 282 of other origins (mostly African, but also Asian, North and South American, Middle Eastern and Oceanian, with categories too small to conduct separate measurement invariance tests); and (c) 109 participants classified as underweight, 811 as normal weight and 115 as overweight-obese. Weight categories were determined based on participants' Body Mass Index [*BMI* = weight/height*height (self-reported)] cut-off scores adjusted for gender and age provided by Cole et al. (2000) and Cole et al. (2007). We elected to rely on Cole et al.'s cut-off scores, as they are the ones most often used in previous research on overweight-obesity and thinness groups (Cole and Lobstein 2012). Participants completed the French version of the EAT-26 (Garner, et al. 1982; Leichner et al. 1994). They answered each item (see the online supplements) on a six-point scale ranging from *always* to *never*. Participants also self-reported their height, weight, gender, age and ethnic origin, so the groups could be defined in measurement invariance tests.

Data Analysis

All analyses in the first study were performed with Mplus 6.0 (Muthén and Muthén 2010), using the maximum likelihood robust estimator with full information estimation to account for the small amounts of missing data at the EAT-26 item level (0.10 % to 2.03 %; M_{missing}=0.65 %; SD_{missing}= 0.41 %; see Enders 2010). In the first and second stages, six ESEM models with three to eight correlated factors were tested (Models 1 to 6). These models were estimated with an oblique geomin rotation and an epsilon value of 0.5 (Marsh et al. 2009; Morin and Maïano 2011a; Morin et al. in press).

In the third and fourth stages, the measurement invariance of the factor model retained in the second stage (Model 7) was tested across gender and age categories. Age was dichotomized into two groups: 11–14 years old (early adolescence; n=485) and 15–18 years old (late adolescence; n= 550). ESEM models were first estimated separately (Models 8–1 and 9–1) in all gender- and age-related subsamples and then across gender and age groups (Models 8–2 and 9–2) as recommended by Meredith (1993): (a) configural invariance; (b) weak invariance (loadings); (c) strong invariance (loadings, intercepts); (d) strict invariance (loadings, intercepts, uniquenesses); (e) variance-covariance invariance; and (f) latent means invariance. In each invariance sequence, the preceding model served as a referent.

Chi-square tests (χ^2) are known to be overly sensitive to sample size and to minor deviations from multivariate normality. Therefore, applied research typically focuses on sample-size-independent indices to assess the fit for the models (e.g., Marsh, Hau et al. 2005), particularly the comparative fit index (CFI), the Tucker-Lewis index (TLI) and the root mean square error of approximation (RMSEA). CFIs and TLIs greater than 0.90 are considered to reflect adequate model fit, although values greater than 0.95 are preferable (Hu and Bentler 1999). RMSEAs smaller than 0.08 or 0.06 support, respectively, acceptable and good model fit (Hu and Bentler 1999). Scale score reliability was computed using McDonald's (1970) $\omega = (\Sigma|\lambda i|)^2 \Big/ \left([\Sigma|\lambda i|]^2 + \Sigma\delta ii\right)$ where λi were the loadings and δii, the uniquenesses.

Measurement invariance tests were evaluated through the examination of robust χ^2 difference tests ($\Delta\ \chi^2$), which take into account the scaling correction composite (Muthén and Muthén 2010; Satorra and Bentler 1999). Recent studies recommend complementing this information with changes in CFIs and RMSEAs (Chen 2007; Cheung and Rensvold 2002), which tend to be more trustworthy than $\Delta\ \chi^2$, which present the same limitations as the χ^2. Here, $\Delta\ \chi^2$ are reported but changes in fit indices are given priority. A ΔCFI of 0.01 or less and a ΔRMSEA of 0.015 or less between a more restricted model and the preceding one indicate that the invariance hypothesis should not be rejected. It should also be noted that for indices incorporating a penalty for lack of parsimony, such as the TLI and RMSEA, a more restrictive model could result in a better fit than a less restricted model; therefore, changes in the TLI should also be inspected (Marsh, Hau et al. 2005). Indeed, fit indices that correct for parsimony (TLI, RMSEA) may be particularly important in ESEM given the large number of parameters estimated (Marsh et al. 2009). However, since there are still few ESEM applications, these cut-off scores should be considered rough guidelines (Marsh et al. 2009).

Results and Discussion

First and Second Stages Six a priori EFA models (Models 1 to 6) with three to eight factors were examined, and their goodness-of-fit statistics are displayed in Table 1. A parallel analysis was also conducted on these data to help determine the correct number of factors in the data and showed that the real data line crossed the random data line between the five- and six-factor solutions (Fabrigar et al. 1999). Results showed unsatisfactory goodness-of-fit indices (CFI and TLI < 0.90) for ESEM models with three and four factors (Models 1 and 2). However, acceptable goodness-of-fit indices (CFI and TLI > 0.95; RMSEA < 0.05) were observed for ESEM models with five to eight factors (Models 3 to 6). Comparison of the five- and six-factor models showed the superiority of the six-factor model according to the significant $\Delta\chi^2$, and to the ΔCFIs and ΔTLIs over the recommended 0.01 increase. Comparison of the five- and six-factor solutions showed the six-factor solution to have a greater level of interpretability. None of the remaining models provided ΔTLIs, ΔCFIs or ΔRMSEAs above the

Table 1 Goodness of fit indices of EAT-26 and EAT-18 models

	N°	Descriptions	χ^2	df	CFI	TLI	RMSEA	90 % CI
Study 1, Stage 1	1	3 first-order factors - 26-item	1374.69**	250	0.840	0.791	0.066	0.063–0.069
	2	4 first-order factors - 26-item	939.47**	227	0.898	0.854	0.055	0.051–0.059
	3	5 first-order factors - 26-item	542.12**	205	0.952	0.924	0.040	0.036–0.044
	4	6 first-order factors - 26-item	343.08**	184	0.977	0.960	0.029	0.024–0.034
	5	7 first-order factors - 26-item	287.51**	164	0.982	0.965	0.027	0.022-0.032
	6	8 first-order factors - 26-item	280.69**	145	0.981	0.957	0.030	0.025–0.035
Stage 2	7	6 first-order factors - 18-item	138.46**	60	0.984	0.959	0.036	0.028–0.043
Stage 3	8–1	Boys (n=494)	78.44**	60	0.990	0.975	0.025	0.000–0.039
		Girls (n=541)	104.26**	60	0.985	0.962	0.037	0.025–0.049
Stage 4	9–1	11–14 years (n=485)	96.06**	60	0.984	0.959	0.035	0.021–0.048
		15–18 years (n=550)	107.08**	60	0.982	0.955	0.038	0.026–0.049
Study 2, Stage 1	10	6 first-order factors – 18-item	404.987**	123	0.946	0.925	0.056	0.050–0.062
Stage 2	11–1	Boys (n=279)	251.23**	123	0.936	0.911	0.061	0.050–0.072
		Girls (n=465)	288.65**	123	0.949	0.928	0.054	0.046–0.062
Stage 3	12–1	11–14 years (n=177)	189.13**	123	0.951	0.932	0.055	0.039–0.070
		15–18 years (n=567)	328.82**	123	0.933	0.907	0.061	0.054–0.068
Stage 4	13–1	Sample 1 (n=1035)	557.87**	123	0.935	0.909	0.058	0.054–0.063
		Sample 2 (n=744)	404.99**	123	0.946	0.925	0.056	0.050–0.062
Stage 5	14–1	European (n=1380)	661.90**	123	0.942	0.919	0.056	0.052–0.061
		African (n=333)	255.26**	123	0.935	0.910	0.057	0.047–0.067
Stage 6	15–1	Underweight (n=190)	201.22**	123	0.940	0.917	0.058	0.043–0.072
		Normal weight (n=1403)	644.22**	123	0.943	0.921	0.055	0.051–0.059
		Overweight (n=186)	243.52**	123	0.904	0.867	0.073	0.059–0.086

CFI comparative fit index; *df* degrees of freedom; *RMSEA* root mean square error of approximation; *TLI* Tucker-Lewis index; *90% CI* 90 % confidence interval for the RMSEA point estimate; *p<0.05; **p<0.01

recommended criteria. The six-factor model was thus retained for subsequent analyses.

Examination of the standardized loadings of the 26 items of the retained six-factor EFA model (detailed results reported in the online supplements) showed that eight items (2–4, 15, 19 and 23–25) were problematic (i.e., main loadings < 0.400, often with substantial cross-loadings) and should be deleted. The resulting 18-item six-factor EFA model of the EAT (EAT-18) was re-estimated in the second stage and provided acceptable goodness-of-fit indices (see Table 2). The parameter estimates from this model (Model 7) are presented in Table 2. The first factor comprises four items with significant main loadings (i.e., items 1, 11, 12 and 14), all reflecting participants' Fear of Getting Fat (e.g., I am terrified about being overweight). The second factor contains three items with significant main loadings (i.e., items 8, 13 and 20), all reflecting Felt Social Pressure to Gain Weight (e.g., I feel that others would prefer if I ate more). The third factor comprises two items with significant main loadings (i.e., items 9 and 26), both assessing Vomiting-Purging Behaviors (e.g., I vomit after I have eaten). The fourth factor contains five items with significant main loadings (i.e., items 5, 6, 7, 16 and 17), all reflecting Eating-Related Control (e.g., I avoid foods with sugar in them). The fifth factor comprises two items with significant main loadings (i.e., items 18 and 21), both reflecting participants' Food Preoccupation (e.g., I give too much time and thought to food). Finally, the sixth factor contains four items (i.e., items 5, 10, 11 and 22) with significant loadings, of which only two exceed 0.400 (items 10 and 22). These items were still retained because they fit the factor, which reflects Eating-Related Guilt (e.g., I feel uncomfortable after eating sweets).

All factors presented modest to acceptable reliability coefficients (ω=0.69 for Eating-Related Control to ω=0.83 for Fear of Getting Fat; see Table 2, left panel). The latent variable correlations (Table 2) were small to moderate, confirming the relative independence of the factors. Still, most of these correlations were statistically significant (except the correlations between Social Pressure to Gain Weight and both Eating-Related Control and Eating-Related Guilt), confirming that most factors are related to a common DEAB core. All of the estimated cross-loadings were small and mostly non-significant (|0.000 to 0.317|, $M_{|\lambda|}$=0.055; $SD_{|\lambda|}$ =0.074; only 6/91 higher than 0.200). This suggests that this model provides a reasonable approximation of simple

Table 2 ESEM and CFA solutions for the EAT-18

Item	ESEM (Study 1, model 7)							CFA (Study 2, model 10)						
	FGF (λ)	SP (λ)	VPB (λ)	ERC (λ)	FP (λ)	ERG (λ)	δ	FGF (λ)	SP (λ)	VPB (λ)	ERC (λ)	FP (λ)	ERG (λ)	δ
1	**0.761**	0.039†	−0.013†	−0.025†	0.047†	−0.033†	0.442	**0.748**						0.440
5	0.060†	0.080†	0.102	**0.377**	0.053†	−0.241	0.826				**0.305**			0.907
6	0.032†	0.059†	−0.017†	**0.502**	0.005†	−0.023†	0.736				**0.538**			0.711
7	0.021†	0.042†	0.177	**0.492**	−0.074†	0.047†	0.666				**0.621**			0.614
8	0.046†	**0.907**	−0.007†	−0.014†	−0.102	0.017†	0.208		**0.842**					0.291
9	−0.003†	−0.026†	**0.936**	0.006†	0.003†	−0.015†	0.139†			**0.855**				0.269
10	0.171†	0.015†	0.202†	0.001†	0.026†	**0.616**	0.320						**0.878**	0.228
11	**0.724**	−0.030†	0.023†	0.020†	−0.025†	0.210	0.255	**0.875**						0.234
12	**0.408**	−0.051†	0.014†	0.271	0.086†	−0.011†	0.602	**0.642**						0.588
13	−0.088	**0.622**	0.052†	−0.009†	0.035†	−0.072†	0.572		**0.635**					0.597
14	**0.836**	−0.002†	−0.015†	0.024†	0.003†	0.044†	0.240	**0.896**						0.197
16	0.013†	−0.024†	−0.027†	**0.806**	−0.080†	0.000†	0.375				**0.729**			0.469
17	−0.104	−0.060	−0.009†	**0.562**	0.071†	0.062†	0.692				**0.618**			0.618
18	0.058†	−0.017†	0.010†	0.060†	**0.648**	0.055†	0.474					**0.682**		0.534
20	−0.029†	**0.693**	−0.022†	0.038†	0.145	0.035†	0.460		**0.728**					0.470
21	0.008†	0.026†	0.010†	−0.038	**0.790**	0.005†	0.369					**0.726**		0.472
22	0.016†	0.014†	−0.005†	0.317	0.051†	**0.617**	0.313						**0.820**	0.328
26	−0.009†	0.024†	**0.598**	−0.001†	0.010†	0.237†	0.492			**0.723**				0.478
Correlations	FGF	SP	VPB	ERC	FP	ERG		FGF	SP	VPB	ERC	FP	ERG	
FGF	1.00							1.00						
SP	−0.108*	1.00						−0.060	1.00					
VPB	0.215*	0.201*	1.00					0.337*	0.232*	1.00				
ERC	0.498*	0.077	0.223*	1.00				0.599*	0.088	0.453*	1.00			
FP	0.423*	0.145*	0.260*	0.242*	1.00			0.439*	0.238*	0.316*	0.286*	1.00		
ERG	0.540*	0.055	0.305*	0.383*	0.484*	1.00		0.770*	0.051	0.460*	0.674*	0.582*	1.00	
ω	0.83	0.80	0.79	0.69	0.71	0.71		0.87	0.78	0.77	0.70	0.66	0.84	

The main loadings of the items onto their a priori factor are in bold, cross loadings are in regular font. *CFA* confirmatory factor analysis; *ESEM* exploratory structural equation modeling; *ERC* eating-related control; *ERG* eating-related guilt; *FGF* fear of getting fat; *FP* food preoccupation; *SP* social pressure to gain weight; *VPB* vomiting-purging behavior; λ factor loading; δ uniquenesses. † non-significant parameter estimate. *$p<0.01$ for the latent factor correlation

structure as defined by Thurnstone (1947) and would likely not result in biased correlations if estimated within a classical CFA framework.

Third Stage The results from Models 8–1 and 9–1 are reported in Table 1 and show that the six-factor EFA model of the EAT-18 performed relatively well in the various gender and age subsamples (i.e., CFI and TLI > 0.95 and RMSEA < 0.05). Measurement invariance tests were then conducted across gender groups (see Model 8–2 in the online supplements). Across the full sequence of measurement invariance tests (corresponding to steps one to six of Model 8.2: configural, weak, strong, strict, variance-covariance and latent means), the results were highly similar and showed that (a) all of the χ^2 were significant, as were most of the $\Delta\chi^2$; (b) the CFI, TLI and RMSEA indicated adequate model fit at all steps; (c) the ΔRMSEA never showed an increase greater than 0.015; (d) the ΔCFI and ΔTLI remained below 0.01 for the first three steps, confirming the configural, weak and strong invariance of the model. However, the ΔCFI and ΔTLI were both greater than 0.01 at the fourth step, suggesting a lack of complete invariance for the uniquenesses of the items. Further probing of these results showed that this was because girls presented higher uniquenesses (i.e., more measurement errors) than boys for items 7, 10, 11 and 17. When the invariance constraints were relaxed for these four specific items, the results supported the partial invariance of the uniquenesses of the items. The ΔCFI and ΔTLI were also greater than 0.01 for the last two steps of the invariance sequence, showing that the latent variance-covariance matrix and the latent means differed according to gender. These results are consistent with known gender differences in DEAB, with girls presenting higher mean levels and more variability on the EAT-18 subscales. They confirm that these differences are not biased, since they were computed from an invariant measurement model.

Fourth Stage Similarly, the measurement invariance tests (Model 9–2; see the online supplements) across age categories yielded highly similar results, showing that (a) all of the χ^2 were significant and few of the $\Delta\chi^2$ were significant; (b) the CFI, TLI and RMSEA indicated adequate model fit at all steps; (c) the ΔRMSEA never showed an increase greater than 0.015; (d) the ΔCFI and ΔTLI never showed a decrease greater than 0.01, except for the fourth step, suggesting a lack of complete invariance of the uniquenesses of the items. Examination of this last result showed that the 11–14 year old group had slightly higher measurement errors on items 1, 7 and 9 compared to the 15–18 year old group. The measurement errors for these items were thus freely estimated in a subsequent model, keeping the invariance constraints on the remaining uniquenesses. The results from this additional model confirmed the partial strict invariance across age categories. Starting from this model, the next tests confirmed that the latent variance-covariance matrix and latent means were fully invariant (equal) across age categories. It should be noted that the 0.024 improvement in the TLI between the first and second steps can occur in ESEM because the TLI incorporates a penalty for lack of parsimony and because the number of freely estimated loadings and cross-loadings is very high in ESEM in the first (configural) step of the measurement invariance test. Therefore, constraining all of these loadings and cross-loadings to be invariant results in a substantial increase in model parsimony (Morin and Maïano 2011a).

Second Study: Confirmatory Factor Analyses of the EAT-18

Method

A new independent sample of 744 youths (aged 11–18 years; M_{age}=15.75 years, SD_{age}=1.68) was recruited in eight new schools (four middle and four high) in Southern France. This sample includes: (a) 279 boys and 465 girls; (b) 627 of European origin and 117 of other origins (mostly African, but also Asian, North American and Middle Eastern, with categories too small to conduct separate measurement invariance tests); and (c) 81 participants classified as underweight, 592 as normal weight and 71 as overweight-obese. Participants completed the EAT-18 developed in the first study in the same conditions.

To conduct further measurement invariance tests across weight categories, participants from the first and second studies were aggregated to obtain a total sample of 1779 youths (773 boys, 1006 girls; M_{age}=15.07 years, SD_{age}=1.98) including 190 participants classified as underweight, 1403 as normal weight and 186 as overweight-obese. However, participants of some ethnic origins (n=66; Asian, North and South American, Middle Eastern and Oceanian) still formed an insufficient subsample and were excluded. Consequently, measurement invariance tests across ethnic groups were conducted on a total sample of 1713 youths, including 1380 of European origin and 333 of African origin.

Data Analysis CFAs were performed on the final model identified in the first study using bootstrapped maximum likelihood estimation with AMOS 7.0. Full-information estimation was used to correct for the small amounts of missing data at the EAT-18 item level [second study subsample (n= 744): 0.13 %–1.21 %; $M_{missing}$=0.47 %; $SD_{missing}$=0.37 %].

Three latent variables (Vomiting-Purging Behavior, Food Preoccupation, Eating-Related Control) were based on two indicators, which is suboptimal, as it creates locally underidentified

constructs (although the overall model remains overidentified). Therefore, these constructs were locally identified using essentially tau-equivalent constraints (Little et al. 1999), which involves placing equality constraints on the loadings of both indicators to help locate the construct at the true intersection of the indicators. Little et al. (1999) noted that this procedure may result in a slight decrease in the fit of the models, which should not overly concern researchers if it does not become dramatic (i.e., $\Delta > 0.05$). Measurement invariance tests were adapted to these constraints following Morin and Maïano's (2011b) recommendations. Therefore, the configural and loading invariance models were first estimated without essentially tau-equivalent constraints, which were then imposed on the invariant loadings.

First, the model retained in the first study was estimated using CFA (Model 10). Second, this CFA model was estimated separately in the gender (Model 11–1) and age (Model 12–1) groups from the second study sample. Third, the measurement invariance of this model was systematically tested across gender (Model 11–2) and age (Model 12–2) groups. In the fourth stage, the retained CFA model was first estimated separately among the samples from the first and second studies (Model 13–1) to investigate how well it could replicate the ESEM results from the first study. This was even further examined with rigorous measurement invariance tests across the samples (Model 13–2). In the fifth and sixth stages, the retained CFA model was first estimated in the various ethnic and weight-related (Models 14–1 and 15–1) groups. Then, measurement invariance tests were systematically conducted across ethnic and weight-related groups (Models 14–2 and 15–2).

Finally, point and interval estimates of percentile ranks for raw scores on the EAT-18 scales were estimated for the total (n=1779) and for the gender subsamples using Crawford et al. (2009) program. The percentile ranks were obtained using the following formula: $(m+0.5k/N)*100$, where m was the number of participants scoring below a given score, k was the number obtaining the given score, and N was the total sample size (Crawford et al. 2009). The 95 % interval estimates of these percentile rank points were estimated using a Bayesian approach (Crawford et al. 2009).

Results and Discussion

First Stage The results from the six-factor CFA model of the EAT-18 are reported in Tables 1 and 2. This model presents acceptable goodness-of-fit indices (CFI and TLI > 0.90; RMSEA < 0.06.), significant and satisfactory standardized loadings (i.e., from 0.305 to 0.896) and modest to acceptable reliability coefficients (ω=0.66 for the Food Preoccupation scale to ω=0.87 for the Fear of Getting Fat scale; see Table 2, right panel). These results are highly similar to the EFA results reported in the first study and confirm that CFA assumptions hold for the EAT-18 measurement model. Furthermore, the latent variable correlations, also reported in Table 2, show no sign of being severely inflated ($Mr_{\text{sample 2}}$=0.41) when compared to those obtained in the first study ($Mr_{\text{sample 1}}$=0.26). This suggests that the gains in parsimony and clearer construct definitions outweighed the gains in precision linked to cross-loadings in this specific case. Most of them were again statistically significant and of moderate magnitude, except for the correlations between (a) Fear of Getting Fat and Social Pressure to Gain Weight, (b) Social Pressure to Gain Weight and Eating-Related Control, and (c) Social Pressure to Gain Weight and Eating-Related Guilt.

Second and Third Stages The results from Models 11–1 and 12–1 (Table 1) show that the six-factor CFA model of the EAT-18 performed relatively well in the separate gender and age groups (i.e., CFI and TLI > 0.90 and RMSEA ≤ 0.06). Measurement invariance tests (Models 11–2 and 12–2) across gender and age groups showed that (a) all of the χ^2 were significant; (b) most of the χ^2 difference tests were significant for the age-based measurement invariance tests, but none of them were significant for the gender-based measurement invariance tests; (c) the CFI, TLI and RMSEA indicated adequate model fit at all steps; (d) the ΔRMSEA never showed an increase greater than 0.015; and (e) the ΔCFI and ΔTLI never showed a decrease greater than 0.01, except when essentially tau-equivalent constraints were added to the gender models for the ΔTLI, and for the age-based measurement invariance tests in the fifth step, suggesting a lack of invariance of the factor variance-covariance matrix. When the results were investigated to locate the non-invariant parameters responsible for this substantial and significant decrease, they revealed that the lack of invariance was due to higher associations between Vomiting-Purging Behavior and Food Preoccupation, and between Vomiting-Purging Behavior and Eating-Related Guilt, in the younger (11–14 years old) group than in the older (15–18 years old) group. Overall, these results confirm the full strict measurement invariance of the EAT-18 model when based on CFA and fail to replicate the gender- and age-related differences noted in the first study. This difference in results may be due to the smaller sample size in the present study (resulting in less power to detect group-based differences), to the greater clarity of the constructs defined based on a CFA model without cross-loadings, or to idiosyncrasies of the samples used in this series of studies.

Fourth to Sixth Stages The results from Models 13–1 (Table 1) and 13–2 (see the online supplements) show that the six-factor CFA model of the EAT-18 performed relatively well (i.e., CFI and TLI > 0.90, and RMSEA ≤ 0.06) in both samples (Model 13–1) and was fully invariant (Model

13–2) across the samples used in the first and second studies, supporting our decision to merge these samples for further measurement invariance tests between ethnicities and across weight categories. The results from Models 14–1 and 15–1 (Table 1) further show that the six-factor CFA model of the EAT-18 performed relatively well in all subsamples formed on the basis of ethnicities and weight categories (i.e., with the sole exception of the small overweight group, all CFIs and TLIs > 0.90 and RMSEAs < 0.06). The measurement invariance tests (Models 14–2 and 15–2; see the online supplements) between ethnicities and across weight categories were highly similar and showed that (a) all of the χ^2 were significant and most of the $\Delta\chi^2$ were significant; (b) the CFI, TLI and RMSEA indicated adequate model fit at all steps; (c) the ΔRMSEA never showed an increase greater than 0.015; and (d) the ΔCFI and ΔTLI never showed a decrease greater than 0.01, except for the fifth step of Model 15–2, suggesting a lack of invariance of the factor variance-covariance matrix across weight categories. Further probing of this result showed weaker associations between (a) Eating-Related Control and Food Preoccupation, (b) Fear of Getting Fat and Eating-Related Guilt and (c) Social Pressure to Gain Weight and Food Preoccupation in the underweight groups than in the other weight groups.

Seventh Stage The point and interval estimates of percentile ranks for raw scores on the various EAT-18 scales were estimated for the total sample and the gender-based subsamples. They are reported in the online supplements.

Third Study: Convergent Validity of the EAT-18 with Criterion Measures

Samples

A new sample of 114 youths (aged 11–18 years; M_{age}=14.32 years, SD_{age}=1.51; M_{BMI}=18.90, SD_{BMI}=2.32) was recruited in four new schools (two middle and two high) in Southern France. This sample includes: (a) 47 boys and 67 girls and (b) 100 participants of European origin and 14 of African, Asian and North American origin. The participants completed, in the same standardized conditions, the EAT-18 and the short French version of the Eating Disorders Inventory adapted for adolescents (EDI-A-24; Garner et al. 1983; Maïano et al. 2009).

A second new sample of 285 youths aged between 11 and 18 years (M_{age}=14.96 years, SD_{age}=1.88; M_{BMI}=19.65, SD_{BMI}=2.54) was recruited in six of the same schools (three middle and three high). This sample includes: (a) 133 boys and 152 girls and (b) 237 participants of European origin and 48 of African origin. All adolescents completed, in the same aforementioned conditions, the EAT-18 and the French versions of (a) the short form of the Physical Self-Inventory (Maïano et al. 2008; Morin and Maïano 2011a), (b) the Social Physique Anxiety Scale (Hart et al. 1989) and (c) the Fear of Negative Appearance Evaluation Scale (Lundgren et al. 2004).

Measures

The EAT-18 and the EDI-A-24 were used to assess DEAB. The EDI-A-24 comprises 24 items assessing three DEAB scales and five related psychological characteristics (three items per dimension). To focus on DEAB as assessed by the EAT-18, only the three DEAB scales of the EDI-A-24 were used in this study: (a) Body Dissatisfaction (e.g., I like the shape of my buttocks); (b) Bulimia (e.g., I stuff myself with food); and (c) Drive for Thinness (e.g., I feel like going on a diet). The EDI-A-24 (Maïano et al. 2009) was validated on a mixed (males and females) sample of adolescents in a series of five studies (n=1323). These studies supported the original factor model across two independent samples, revealed acceptable internal consistency (α=0.73 to 0.96) and test-retest reliability (r=0.75 to 0.91) coefficients and confirmed the convergent validity of the EDI-A-24.

Participants' Global Self-Esteem (three items, e.g., I have a good opinion of myself) and Perceived Physical Appearance (three items, e.g., I have a nice body to look at) were assessed with subscales from the Physical Self-Inventory – Short (PSI-S; Maïano et al. 2008; Morin and Maïano 2011a). This instrument was validated and cross-validated on a total sample of 3047 French adolescents. These studies supported the factor structure, internal consistency (α=0.64 to 0.87), test-retest reliability (r=0.74 to 0.84) and convergent validity of the PSI-S. Items are answered on a six-point scale ranging from *not at all* (1) to *entirely* (6).

The French version of the Social Physique Anxiety Scale (SPAS; Hart et al. 1989; Maïano, Morin, Eklund et al. 2010) was used to assess participants' social physique anxiety. The seven items (e.g., I worry about wearing clothes that might make me look too thin or overweight) are rated on a five-point Likert-type scale [*not at all* (1) to *extremely* (5)]. The French version of the SPAS (Maïano, Morin, Eklund et al. 2010) was validated on a total sample of 1563 French adolescents. These studies supported the factor structure, internal consistency (α=0.81–0.87), test-retest reliability (r=0.78) and convergent validity of the SPAS.

The French version of the Fear of Negative Appearance Evaluation Scale (FNAES: Lundgren et al. 2004; Maïano, Morin, Monthuy-Blanc et al. 2010) was used to evaluate participants' fear of having their physical appearance negatively evaluated by others. The five items (e.g., I am concerned about what other people think of my appearance) are rated on a five-

point Likert-type scale [*not at all* (1) to *extremely* (5)]. The French FNAES (Maïano, Morin, Monthuy-Blanc et al. 2010) was validated on a sample of 684 adolescents in a series of three studies which supported its factor structure, internal consistency (α=0.83), test-retest reliability (r=0.77) and convergent validity.

Analyses, Results and Discussion

Pearson correlation analyses were performed to verify the convergent validity of the scale scores of the EAT-18 with the other measures. A Bonferroni correction was applied and the alpha error was set at 0.02 (0.05/3) for relations between the EAT-18 and the EDI-A-24, and at 0.01 (0.05/4) for the relations between the EAT-18 and the other instruments.

As shown in Table 3, the internal consistency coefficients for the EAT-18 and EDI-A-24 are modest to acceptable (α= 0.60–0.85). The correlations between both instruments are reported in Table 3 and show that: (a) the Fear of Getting Fat, Food Preoccupation and Eating-Related Guilt scales of the EAT-18 are significantly and positively related to the Drive for Thinness and Bulimia scales of the EDI-A-24; (b) the Eating-Related Control scale of the EAT-18 is significantly and positively related to the Drive for Thinness scale of the EDI-A-24; and (c) the Fear of Getting Fat scale of the EAT-18 is positively and significantly related to the Body Dissatisfaction scale of the EDI-A-24. Social Pressure to Gain Weight is the only EAT-18 scale that did not show a significant association with the EDI-A-24 scales.

As shown in Table 3, the Cronbach's alpha of the Global Self-Esteem and Physical Appearance scales of the PSI-S, the SPAS and the FNAES are modest to acceptable (α=0.62 to 0.89). The results from the correlation analyses are also reported in Table 3 and reveal that the Fear of Getting Fat, Eating-Related Control and Eating Related Guilt scales are (a) significantly and negatively related to the Global Self-Esteem and Physical Appearance scales of the PSI-S and (b) significantly and positively related to the SPAS and the FNAES. Additional results also showed that the Food Preoccupation and Vomiting-Purging Behavior scales are (a) significantly and negatively related to the Global Self-Esteem scale and (b) significantly and positively related to the SPAS and the FNAES. Finally, the Social Pressure to Gain Weight scale is significantly and positively related to the FNAES.

General Discussion

The purpose of the first study was to explore the factor structure, reliability and measurement invariance across gender and age categories of the EAT using ESEM. The results showed that neither the original three-factor structure of the EAT-26 nor the various models identified in previous studies could be satisfactorily replicated. Indeed, our results clearly supported a revised six-factor solution based on 18 of the original items. Although the resulting EAT version was slightly shorter than the original version, our objective was not to develop a short version of the EAT (although we did so more systematically than previous attempts to do so, e.g., Belon et al. 2011), but rather to systematically validate a version based on the most appropriate items from the EAT-26.

Interestingly, this six-factor model was highly similar to the original seven-factor structure initially proposed by Garner and Garfinkel (1979) for the EAT-40: (1) Food Preoccupation, (2) Body Image for Thinness, (3) Vomiting and Laxative Abuse, (4) Dieting, (5) Slow Eating, (6) Clandestine Eating and (7) Perceived Social Pressure to Gain Weight. The present six-factor solution apparently recovers most of this initial solution: (1) Fear of Getting Fat (four items, similar to the second EAT-40 factor), (2) Social Pressure to Gain Weight (three items, similar to the seventh EAT-40 factor), (3) Vomiting-Purging Behavior (two items, similar to the third EAT-40 factor), (4) Eating-

Table 3 Convergent validity of the EAT-18 items with criterion measures

EAT-18	DT[a]	BU[a]	BD[a]	GSE[b]	PA[b]	SPA[b]	FNAE[b]	Cronbach α
FGF	0.62**	0.28**	0.30**	−0.52**	−0.31**	0.56**	0.52**	0.85[a]/0.84[b]
SP	−0.02	0.05	−0.04	−0.03	0.10	0.09	0.13*	0.72/0.74
VPB	0.17	0.20*	−0.02	−0.18**	−0.02	0.26**	0.28**	0.78/0.69
ERC	0.39**	0.06	0.07	−0.21**	−0.18**	0.30**	0.27**	0.60/0.62
FP	0.44**	0.32**	0.10	−0.17**	−0.06	0.17**	0.16**	0.61/0.78
ERG	0.41**	0.28**	0.18	−0.41**	−0.28**	0.49**	0.36**	0.70/0.66
Cronbach α	0.75	0.74	0.75	0.70	0.71	0.76	0.89	–

BD body dissatisfaction; *BU* bulimia; *DT* drive for thinness; *EAT* eating attitudes test; *ERC* eating-related control; *ERG* eating-related guilt; *FNAES* fear of negative appearance evaluation; *FGF* fear of getting fat; *FP* food preoccupation; *GSE* global self-esteem; *PA* perceived physical appearance; *SP* social pressure to gain weight; *SPA* social physique anxiety; *VPB* vomiting-purging behavior. [a] based on the 114 participants from Study 3. [b] based on the 285 participants from Study 3. *p<0.05; **p<0.01

Related Control (five items, similar to a combination of the fourth and fifth EAT-40 factors), (5) Food Preoccupation (two items, similar to the first EAT-40 factor) and (6) Eating-Related Guilt (two items, similar to the sixth EAT-40 factor). This suggests that the failure to replicate Garner and Garfinkel's (1979) initial proposal may have been due to the inclusion of too many suboptimal items in the various versions of the EAT, reliance on suboptimal principal component analyses and selection criteria (i.e., scree test, Kaiser), and improper recoding of the answers into four categories. This last element is particularly noteworthy since it has been well documented that Likert-type scales including less than five answer categories, particularly when assessing abnormal phenomena like DEAB, were better modeled as ordered categories (e.g., Finney and DiStefano 2006). So, in addition to losing information and power linked to the arbitrary collapsing of the answer scales into a lower number of categories (MacCallum et al. 2002), previous studies also improperly modeled these recoded items relying on continuous normal-data assumptions. Using the full range of six answer categories allowed us to observe finer distinctions in the definition of the EAT subscales.

Further analysis of the EAT-18 showed that these six EAT-18 subscales were moderately correlated with one another, confirming that they had tapped into a core DEAB theme, while still reflecting relatively distinct subconstructs. Still, the non-significant correlations between the Social Pressure to Gain Weight scale and both the Eating-Related Control and Eating-Related Guilt scales confirm that both sources of pressure (social versus internal) were relatively independent. This independence may be explained by the participants' weight status. Indeed, only a relatively small subgroup of participants from this study was classified as underweight (~11 %) and, realistically, only these youths may be exposed to Social Pressure to Gain Weight. Therefore, for underweight youths, at least those not presenting problematic levels of DEAB, eating control and eating-related guilt may be more strongly related to Social Pressure to Gain Weight. Conversely, among normal-weight and overweight-obese youths, food control and guilt may be more strongly determined by their Fear of Getting Fat. Unfortunately, the sample size available in the third study (n=114) is too small to allow us to properly verify this possibility separately within each weight-related category. This issue should be clearly investigated in future research.

Subsequent analyses also showed adequate internal consistency coefficients for these subscales (ranging from 0.69 to 0.83). These coefficients were noteworthy since half of the scales included only two items and that internal consistency coefficients are known to be strongly and positively affected by the number of items in a scale (Streiner 2003). Additionally, the results from the first study confirmed the strong (i.e., loadings and intercepts) invariance of this measurement model across gender and age categories, as well as its partial strict measurement invariance (uniquenesses). This last result may, however, be attributed to the slightly larger sample size used in the first study (resulting in more power to detect differences) or to the ESEM approach including cross-loadings, since the second study failed to replicate this partial lack of invariance using CFA and rather confirmed the complete strict measurement invariance (i.e., loadings, intercepts and uniquenesses) of the EAT-18 across gender and age categories. However, this possible lack of invariance for the uniquenesses of at least some of the items suggests that scholars may rely on latent variable methodologies when comparing the EAT-18 subscale scores across gender or age categories. Indeed, latent variable methodologies explicitly control for measurement errors when estimating the constructs of interest. Therefore, the fact that the construct is measured with more or less errors in the various subgroups is explicitly taken into account, and fully controlled for, in the context of latent variable methodologies, which thus allows us to obtain estimates of group difference or treatment effects unbiased by measurement errors.

As expected (e.g., Chamay-Weber et al. 2005; Smink et al. 2012), the first study showed that boys presented lower scores than girls on most of the EAT-18 scales (except for the Vomiting-Purging Behavior scale). Not surprisingly, girls also presented a higher level of variability on most subscales. However, these results were not replicated in the second study, calling into question the equivalence of the samples, or perhaps reflecting the lower statistical power of the second study, which was based on a slightly reduced sample size. Indeed, the CFA models in the second study were found to be perfectly replicated and fully invariant when compared across the samples used in the first and second studies, supporting the idea that both samples were comparable, at least in regard to the EAT-18 measurement model. Since this study was the first to investigate the measurement, latent means and variance-covariance invariance of the EAT-18 across gender and age groups, these promising results deserve further investigation. It would be interesting to verify whether the EAT-18 lacks sensitivity to well-documented mean-level differences in the prevalence of DEAB symptoms across genders, or whether these differences are truly less pronounced in community samples.

The second study showed that the EAT-18 measurement model, as well as all other results from the first study (internal consistency, measurement invariance) could be successfully replicated using a more parsimonious CFA approach, without resulting in severely inflated latent factor correlations. Following the observation that this CFA model was fully replicated and completely invariant across the samples from the first and second studies, these two samples

were combined for further measurement invariance tests between ethnicities and across weight categories. These additional tests confirmed the strict measurement invariance of the EAT-18 CFA model (a) between youths of European or African origin and (b) between youths classified as underweight, normal weight or overweight-obese. Therefore, the results from these two initial studies support the psychometric properties of the EAT-18 and show that scholars can confidently use it among community samples of French youths from either a European or African background.

The third and fourth studies were performed to verify the convergent validity of the EAT-18 with another measure of DEAB and additional criterion measures. The results from these studies revealed that answers to the six scales of the EAT-18 were moderately correlated with the core symptom scales (i.e., Drive for Thinness, Bulimia, Body Dissatisfaction) of the EDI-A-24 and measures of global self-esteem, physical appearance, social physique anxiety and fear of negative appearance evaluation. These findings were in the expected direction and, in fact, were very similar to those from previous DEAB research (Berland et al. 1986; Karpowicz et al. 2009; Krane et al. 2001; Maïano, Morin, Eklund et al. 2010; Maïano, Morin, Monthuy-Blanc et al. 2010; Raciti and Norcross 1987). Only the perceived Social Pressure to Gain Weight subscale showed fewer correlations with the criterion measures than the other subscales. This subscale covers a facet of DEAB not generally reflected in the criterion measures, except perhaps the FNAES, with which it was moderately correlated. These findings support the convergent validity of the EAT-18, except perhaps for the Social Pressure to Gain Weight subscale, which merits further investigation.

These results have important clinical implications in terms of using the EAT as a screening instrument for DEAB and in clinical practice. Indeed, the practical usefulness of an instrument is directly related to the fact that the instrument can be shown to assess clearly defined constructs that can be reliably assessed in a stable manner across samples. Previous studies on the EAT-26 generally failed to converge on a well-replicated solution, calling into question the real confidence that could be ascribed to this instrument. Here, we obtained a clear and well-replicated factor structure, closely related to the proposed factor structure of the longer EAT-40. In addition, the results from our measurement invariance tests showed that the measured constructs did not change in meaning across meaningful subgroups of participants (loadings), yielded assessments of severity levels that are directly comparable across meaningful subgroups of participants (intercepts) and provided assessments of comparable precision across meaningful subgroups of participants (uniquenesses). Indeed, these results showed that the EAT-18 could be used to assess DEAB without showing evidence of measurement biases across gender or age groups. But more importantly, our results supported the ability of the EAT-18 to screen for DEAB and thus for threshold eating disorders for which they represent known risk factors (e.g., Chamay-Weber et al. 2005; Jacobi et al. 2004; Smink et al. 2012; Stice 2002). In fact, half of the EAT-18 subscales (i.e., Fear of Getting Fat, Food Preoccupation, Eating-Related Guilt) are directly related to core symptoms of both anorexia nervosa and bulimia nervosa. Similarly, other subscales of the EAT-18 may be used to screen for symptoms specific to bulimia (i.e., Vomiting-Purging Behavior) and anorexia (i.e., Eating-Related Control) nervosa, and thus can likely be used to screen for these specific eating disorders.

Four main limitations of these results must be considered. First, all analyses were based on community samples of French adolescents. Consequently, the use of the EAT-18 should be limited to samples comprising similar characteristics, but can be generalized to various gender and age categories (i.e., 11–18 years), ethnicities (i.e., European and African) and weight categories (i.e., underweight, normal weight and overweight-obese). Future studies should also investigate whether this factor structure holds with younger children, in clinical samples and in other linguistic and cultural groups. Second, this study did not assess the temporal stability (test-retest reliability) of the EAT-18, which should be verified to obtain a fuller picture of the psychometric properties of this instrument. Third, this instrument comprises a small number of items per factor (i.e., two to five) and it is currently unknown whether this is sufficient to adequately cover the content of each subconstruct. This issue should be clearly addressed in future research with the EAT-18 and the EAT-40. However, it should be noted that the current instrument was developed following recent recommendations for developing short scales (Marsh, Ellis et al. 2005; Smith et al. 2000) and that measures developed according to these criteria are generally found to provide a degree of construct coverage that is comparable to longer versions of the same instruments (Marsh, Ellis et al. 2005; Maïano et al. 2008) so that more is not always better. The results from the convergent validity tests conducted here seem to confirm this interpretation. Fourth, the discriminant validity of the EAT-18 was not tested in these studies. It is thus unknown whether the EAT-18 allows proper differentiation of youths with and without clinical levels of eating disorders or between youths with anorexia nervosa or bulimia nervosa.

In sum, the present results validated and cross-validated an 18-item version of the EAT covering six distinct DEAB dimensions. These findings suggest that this shorter version of the EAT-26 presents acceptable psychometric properties and may be used to screen for DEAB among community samples of French adolescents. Nevertheless, given the aforementioned limitations of these studies, it would be

premature at this time to recommend using this version with younger children and in other cultural or linguistic groups.

References

Anderson, D. A., Lundgren, J. D., Shapiro, J. R., & Paulosky, C. A. (2004). Assessment of eating disorders: review and recommendations for clinical use. *Behavior Modification, 28*, 763–782.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438.

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*, 317–320.

Berland, N. W., Thompson, J. K., & Linton, P. H. (1986). Correlation between the EAT-26 and the EAT-40, the eating disorders inventory, and the restrained eating inventory. *International Journal of Eating Disorders, 5*, 569–574.

Carter, P. I., & Moss, R. A. (1984). Screening for anorexia and bulimia nervosa in a college population: problems and limitations. *Addictive Behaviors, 9*, 417–419.

Chamay-Weber, C., Narring, F., & Michaud, P.-A. (2005). Partial eating disorders among adolescents: a review. *Journal of Adolescent Health, 37*, 417–427.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling, 14*, 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

Cole, T. J., & Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity, 7*, 284–294.

Cole, T. J., Bellizzi, M., Flegal, K., & Dietz, W. (2000). Overweight and obesity worldwide: international establishing a standard definition for child survey. *British Medical Journal, 320*, 1240–1243.

Cole, T. J., Flegal, K., Nicholls, D., & Jackson, A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *British Medical Journal, 335*, 194–197.

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: proposed reporting standards and methods for quantifying uncertainty over percentile ranks of test scores. *The Clinical Neuropsychologist, 23*, 1173–1195.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.

Fabrigar, L. R., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological reserach. *Psychological Methods, 4*, 272–299.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509–528.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 269–314). Greenwich: IAP.

Garfinkel, P. E., & Newman, A. (2001). The eating attitudes test: twenty-five years later. *Eating and Weight Disorders, 6*, 1–24.

Garner, D. M., & Garfinkel, P. E. (1979). Eating attitudes test: an index of the symptoms of anorexia nervosa. *Psychological Medicine, 9*, 273–279.

Garner, D. M., Olmstead, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitude test: psychometric features and clinical correlates. *Psychological Medicine, 12*, 871–878.

Garner, D. M., Olmstead, M. A., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2*, 15–34.

Golden, N. H., Katzman, D. K., Kreipe, R. E., Stevens, S. L., Sawyer, S. M., Rees, J., et al. (2003). Eating disorders in adolescents: position paper of the society for adolescent medicine. *Journal of Adolescent Health, 33*, 496–503.

Hart, E. H., Leary, M. R., & Rejeski, W. J. (1989). The measurement of social physique anxiety. *Journal of Sport & Exercise Psychology, 11*, 94–104.

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

Jacobi, C., Hayward, C., de Zwaan, M., Kraemer, H. C., & Agras, W. S. (2004). Coming to terms with risk factors for eating disorders: application of risk terminology and suggestions for a general taxonomy. *Psychological Bulletin, 130*, 19–65.

Karpowicz, E., Skärsäter, I., & Nevonen, L. (2009). Self-esteem in patients treated for anorexia nervosa. *International Journal of Mental Health Nursing, 18*, 318–325.

Krane, V., Stiles-Shipley, J. A., Waldron, J., & Michalenok, J. (2001). Relationships among body satisfaction, social physique anxiety, and eating behaviors in female athletes and exercisers. *Journal of Sport Behavior, 24*, 247–264.

Leichner, P., Steiger, H., Puentes-Neuman, G., Perreault, M., & Gottheil, N. (1994). Validation d'une échelle d'attitudes alimentaires auprès d'une population québécoise francophone. *Revue Canadienne de Psychiatrie, 39*, 49–54.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: when "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192–211.

Lundgren, J. D., Anderson, D. A., & Thompson, J. K. (2004). Fear of negative appearance evaluation: development and evaluation of a new construct for risk factor work in the field of eating disorders. *Eating Behaviors, 5*, 75–84.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin, 111*, 490–504.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Maïano, C., Morin, A. J. S., Ninot, G., Monthuy-Blanc, J., Stephan, Y., & Vallée, P. (2008). A short and very short form of the physical self-inventory for adolescents: development and factor validity. *Psychology of Sport & Exercise, 9*, 830–847.

Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., Garbarino, J.-M., & Stephan, Y. (2009). Eating disorders inventory: assessment of its construct validity in a non-clinical French sample. *Journal of Psychopathology and Behavioral Assessment, 31*, 387–404.

Maïano, C., Morin, A. J. S., Eklund, R. C., Monthuy-Blanc, J., Garbarino, J.-M., & Stephan, Y. (2010). Construct validity of the social physique anxiety scale in a French adolescent sample. *Journal of Personality Assessment, 92*, 1–10.

Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., & Garbarino, J.-M. (2010). Construct validity of the fear of negative appearance evaluation scale in a French adolescent's community sample. *European Journal of Psychological Assessment, 26*, 19–27.

Marsh, H. W., Ellis, L., Parada, R., Richards, G., & Heubeck, B. (2005). A short version of the self description questionnaire ii: operationalizing criteria for short-form with new applications of confirmatory factor analyses. *Psychological Assessment, 17*, 81–102.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Psychometrics. A Festschrift to Roderick P. McDonald*. Hillsdale: Erlbaum.

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439–476.

McDonald, R. P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1–21.

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

Micali, N., & House, J. (2011). Assessment measures for child and adolescent eating disorders: a review. *Child & Adol;escent Mental Health, 16*, 122–127.

Mintz, L., & O'Halloran, S. E. (2000). The eating attitudes test: validation with DSM-IV eating disorder criteria. *Journal of Personality Assessment, 74*, 489–503.

Morin, A. J. S., & Maïano, C. (2011a). Cross-Validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport & Exercise, 12*, 540–554.

Morin, A. J. S., & Maïano, C. (2011b). Cross-validation of the very short form of the Physical Self-Inventory (PSI-VS): invariance across genders, age groups, ethnicities and weight statuses. *Body Image, 8*, 404–410.

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (in press). Exploratory structural equation modeling: An introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.). Greewich: IAP.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide*. Los Angeles: Muthén & Muthén.

Nasser, M. (1997). The EAT speaks many languages: review of the used of the EAT in eating disorders research. *Eating and Weight Disorders, 2*, 174–181.

Neumark-Sztainer, D., Story, M., Hannan, P. J., Perry, C. L., & Irving, L. M. (2002). Weight-related concerns and behaviors among overweight and nonoverweight adolescents: implications for preventing weight-related disorders. *Archives of Pediatrics & Adolescent Medicine, 156*, 171–178.

Raciti, M. C., & Norcross, J. C. (1987). The EAT and EDI: screening, interrelationships, and psychometrics. *International Journal of Eating Disorders, 6*, 579–586.

Satorra, A., & Bentler, P. (1999). A scaled difference Chi-square test statistic for moment structure analysis. Technical Report University of California, Los Angeles. http://preprints.stat.ucla.edu/260/chisquare.pdf.

Schoemaker, C., Van Strien, T., & Van der Staak, C. (1994). Validation of the eating disorder inventory in a nonclinical population using transformed and untransformed responses. *International Journal of Eating Disorders, 15*, 387–393.

Smink, F. R. E., van Hoeken, D., & Hoek, H. W. (2012). Epidemiology of eating disorders: Incidence, prevalence and mortality rates. *Current Psychiatry Reports*. Advance online publication. doi: 10.1007/s11920-012-0282-y

Smith, G., McCarthy, D., & Anderson, K. (2000). On the sins of short-form development. *Psychological Assessment, 12*, 102–111.

Stice, E. (2002). Risk and maintenance factors for eating pathology: A meta-analytic review. *Psychological Bulletin, 128*, 825–848.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99–103.

Thurnstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

van Strien, T., & Ouwens, M. (2003). Validation of the Dutch EDI-2 in one clinical and two nonclinical populations. *European Journal of Psychological Assessment, 19*, 66–84.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Wildes, J. E., Emery, R. E., & Simons, A. D. (2001). The roles of ethnicity and culture in the development of eating disturbance and body dissatisfaction: a meta-analytic review. *Clinical Psychology Review, 21*, 521–551.