





ORIGINAL ARTICLE

The Athletes' Relationships with Training scale (ART): A self-report measure of unhealthy training behaviors associated with eating disorders

Danielle A. N. Chapa¹  | Kelsey E. Hagan¹  | Kelsie T. Forbush¹  | Victoria L. Perko¹ |
Daria A. Sorokina² | Ahmed Y. Alasmar¹ | Carolyn B. Becker³  | Roberta T. Sherman⁴ |
Ron A. Thompson⁵ | Jennifer G. Farrell⁶ | Tiffany M. Stewart⁷

¹Department of Psychology, University of Kansas, Lawrence, Kansas

²Department of Educational Psychology, University of Kansas, Lawrence, Kansas

³Department of Psychology, Trinity University, San Antonio, Texas

⁴Bloomington, IN

⁵Department of Intercollegiate Athletics, Indiana University-Bloomington, Bloomington, Indiana

⁶Department of Kinesiology, University of North Carolina, Greensboro, North Carolina

⁷Pennington Biomedical Research Center, Baton Rouge, Louisiana

Correspondence

Kelsie T. Forbush, Department of Psychology, Fraser Hall, 1415 Jayhawk Blvd., Lawrence, KS 66045

Email: kforbush@ku.edu

Funding information

National Institutes of Mental Health (NIMH), Grant/Award Number: 1 RO1 MH094448-01

Abstract

Objective: Several studies indicate that eating-disorder (ED) psychopathology is elevated in athletes compared to non-athletes. The assessment of excessive exercise among athletes is a challenge because, compared to non-athletes, athletes are required to train at higher intensities and for longer periods of time. However, individuals participating in competitive sports are still susceptible to unhealthy physical-activity patterns. Most ED assessments were developed and normed in non-athlete samples and, therefore, do not capture the nuances of athletes' training experiences. The purpose of the current study was to develop and validate a clinically useful, self-report measure of unhealthy training behaviors and beliefs in athletes, the Athletes' Relationships with Training Scale (ART).

Method: The initial item pool was administered to $N = 267$ women collegiate athletes who were participating in an ED prevention program study and $N = 65$ women athletes who were in ED treatment.

Results: Factor analyses indicated the ART had a four-factor structure. Factorial and construct validity of the ART were demonstrated. ART scores significantly predicted health care utilization and differed between athletes with an ED versus athletes without an ED. For athletes in ED treatment, ART scores significantly decreased from treatment admission to discharge.

Discussion: The ART showed evidence of strong psychometric properties and clinical utility. The ART could be helpful for clinicians and athletic trainers to help gauge whether athletes are engaging in unhealthy training practices that may warrant clinical attention and for tracking clinical outcomes in athletes with EDs who are receiving treatment.

KEYWORDS

assessment, eating disorders, excessive exercise, scale development

1 | INTRODUCTION

Athletes may be at increased risk for the development of an eating disorder (ED) compared to their non-athlete counterparts (Byrne & McLean, 2002; Fortes, Kakeshita, Almeida, Gomes, & Ferreira, 2014; Holm-Denoma, Scaringi, Gordon, Van Orden, & Joiner, 2009; Martinson & Sundgot-Borgen, 2013; Rosendahl, Bormann, Aschenbrenner,

Aschenbrenner, & Strauss, 2009; Smolak, Murnen, & Ruble, 2000; Sundgot-Borgen, 1993; Sundgot-Borgen & Torstveit, 2004; Toro et al., 2005; Torstveit, Roseninge, & Sundgot-Borgen, 2008). EDs in sport can compromise an athlete's physical and psychological health (Mountjoy et al., 2015; Rauh, Nichols, & Barrack, 2010), which, can interfere with optimal sport performance and functioning in other life domains. To keep athletes healthy, it is important to properly assess

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. International Journal of Eating Disorders published by Wiley Periodicals, Inc.

and identify EDs when they occur. Some ED symptoms, however, are more challenging to assess in athletes compared to non-athletes. In particular, excessive exercise should be evaluated carefully in athletes because their exercise or training levels diverge considerably from non-athletes. To properly identify EDs in athletes, sport psychologists and sport medicine staff must have access to reliable and valid assessment tools that characterize an athlete's training behaviors and beliefs. Thus, the purpose of this study was to develop a measure of unhealthy training in a sample of athletes to signal when an athlete's relationship with training may be indicative of an ED.

Excessive exercise is an ED behavior that facilitates weight loss or is used to compensate for an episode of self-perceived binge eating. In non-athlete samples, excessive exercise is typically identified by the quantity (e.g., exercising for two or more hours per day), as well as the quality of exercise (e.g., feeling driven to exercise). When assessing excessive exercise in athletes, however, one must consider the unique qualities of an athlete's training. For example, athletes sometimes exercise >2 hr in a day (e.g., when coaches schedule two separate practices in a day) and have different motivations for exercise compared to non-athletes. One important motivation for exercise/training among athletes is to improve sport performance (Bompa, 1983). Although training to improve performance is likely the primary reason athletes exercise, several other reasons for engaging in physical activity exist. First, athletes may be motivated to train due to pressures from society and sport culture to have an "athletic" body type (Galli & Reel, 2009). Second, athletes (and coaches) frequently believe that a leaner, thinner athlete will perform better (Currie, 2010; Sundgot-Borgen & Garthe, 2011). For example, some people in sport believe that there is a biomechanical advantage to lower body weight and/or body fat in sports that emphasize running and jumping (Sundgot-Borgen & Garthe, 2011; Thompson & Sherman, 2010). Thus, training might also be viewed as an indirect means to enhance sport performance via weight or body fat loss. Third, many believe that there is a judging (or role) advantage in aesthetic sports (e.g., figure skating, gymnastics, classical ballet), in which individuals are reinforced with better scores (or roles) for having a thinner body (Thompson & Sherman, 2010). Finally, training might be motivated by the need to "make weight" in weight-class sports (e.g., wrestling), in which a weight limit is imposed and must be attained (and maintained) to compete.

Given that athletes have unique experiences with training, traditional screening instruments for EDs might not be appropriate for identifying unhealthy training behaviors and beliefs among athletes. Traditional ED assessment tools, including the Eating Disorder Inventory-Third Edition [EDI-3; (Garner, 2004)] and the Eating Disorder Examination Questionnaire [EDE-Q; (Fairburn & Beglin, 2008)], were not developed for athletes and do not fully assess exercise behaviors, even in non-athletes. For example, the EDI-3 does not include items that assess exercise behavior and the EDE-Q contains only one item that asks about the frequency of driven exercise in the past month. The Eating Pathology Symptoms Inventory [EPSI; (Forbush et al., 2013)], a more recent measure of ED psychopathology, includes an "Excessive Exercise" scale, yet this scale was not developed or validated in athletes.

Eating disorder screening instruments have been developed for athletes specifically, although past screening instruments for athletes

are limited by their item content and scale development procedures. When developing a new scale or measure, it is recommended that authors use both *rational and empirical* scale development methods. First, the initial item pool of a scale should be developed using *rational* methods in which item writing is guided by theory. Second, the initial item pool should be tested using *empirical* methods in which poor-performing items are identified and eliminated using statistics (Clark & Watson, 1995). The Brief Eating Disorder in Athletes Questionnaire [BEDA-Q; (Martinsen, Holme, Pensgaard, Torstveit, & Sundgot-Borgen, 2014)] is limited because it is a compilation of items from three EDI-3 subscales including Body Dissatisfaction, Drive for Thinness, and Purging, none of which examine exercise/training behaviors or beliefs. The Athlete Milieu Direct Questionnaire [AMDQ; (Nagel, Black, Leverenz, & Coster, 2000)] was designed to assess specific ED symptoms (e.g., binge eating, restricting, laxative use, etc.), including excessive training behaviors/beliefs, in women collegiate athletes. Because many of the AMDQ items include content specific to collegiate sports (e.g., scholarships for college) and the experience of women in sports (e.g., loss of menstrual cycle), the AMDQ cannot be used in men or professional sports. Moreover, the AMDQ was developed using rational methods without empirically testing individual item performance. The Female Athlete Screening Tool [FAST; (McNulty, Adams, Anderson, & Affenito, 2001)] items were developed rationally using theory, although they were not empirically evaluated to ensure that the items measured the latent construct of interest (i.e., factor analytic strategies were not used).

Exercise for athletes is multidimensional rather than unidimensional; thus, in addition to assessing the quantity (or amount) of training, the quality of training needs to be considered (Ackard, Brehm, & Steffen, 2002). Training quality includes attitudes (Seigel & Hetta, 2001), motivation (Beumont, Arthur, Russell, & Touyz, 1994), and emotions, such as guilt, experienced when one is unable to exercise (Mond & Calogero, 2009). Given the multidimensional nature of unhealthy training in athletes, an alternative definition of "exercise" must be considered that incorporates not only weight control, but also quality-related factors that maintain unhealthy training (Plateau et al., 2014).

To date, one instrument, the Compulsive Exercise Test [CET; (Meyer, Taranis, Goodwin, & Haycraft, 2011; Taranis, Touyz, & Meyer, 2011)], has been developed to assess multiple dimensions of exercise. Originally developed in a non-athlete sample, the CET assesses five factors related to exercise: Avoidance and Rule Driven Behavior, Weight Control Exercise, Mood Improvement, Exercise Rigidity, and Lack of Exercise Enjoyment. Plateau et al. (2014) tested the CET in a non-clinical sample of athletes and found that in contrast to the factor-structure of the CET in non-athletes, athletes exercised for: Weight Control, Avoidance of Negative Affect, and for Mood Improvement. As a result of the differential factor structure in athletes, the authors of the CET recommended an alternative 15-item, three-scale CET for use in athletes. In conclusion, although Plateau et al.'s (2014) study suggested that training is a multidimensional construct for athletes, no studies have developed a multidimensional training measure for athletes using up-to-date scale development procedures [i.e., combining both rational item writing with empirical testing of individual items; (Clark & Watson, 1995)].

1.1 | Current study

The purpose of the current study was to develop and validate the Athletes' Relationships with Training Scale (ART), which was designed to assess the quality and quantity of athletes' training. Such an instrument could be used by sport psychologists, athletic trainers, and clinicians to assess whether athletes' physical activity behaviors and beliefs are unhealthy and perhaps counterproductive to optimal sport performance. We aimed to create a reliable and valid self-report measure that would capture cognitive and behavioral (Meyer, Taranis, & Haycraft, 2011) aspects of unhealthy training in athletes.

2 | METHOD

2.1 | Initial item Pool development

Item construction was completed by R.S., R.T., J.F., and K.F. and was based on theory and past research suggesting that both qualitative and quantitative motivations for unhealthy training should be considered (Ackard et al., 2002). We focused on writing items that were specific to aspects of the sport environment (i.e., coaches, teammates, training loads, etc.). Items were written to reflect qualitative aspects of training, including: attitudes about exercise (Seigel & Hetta, 2001), exercise motivation (Beumont et al., 1994), and emotions that are often associated with unhealthy exercise [e.g., anxiety (Beumont et al., 1994) and guilt when unable to exercise (Mond & Calogero, 2009)]. Additional items were written to assess "exercise rigidity" (Meyer et al., 2011; Taranis et al., 2011), an inability to alter one's training regimen, including under conditions of illness, injury, or against medical advice (Powers & Thompson, 2008).

2.2 | Procedure and participants

2.3 | Female athlete body project sample

This sample was embedded within a larger parent study, The Female Athlete Body (FAB) Project [see (Stewart et al., 2017) for additional detail on the parent study]. The parent FAB trial consisted of a randomized-control trial of an ED prevention program for athletes with 3-week, 6-month, 12-month, and 18-month follow-ups. Women collegiate athletes were cluster randomized by sports team to either receive the FAB prevention program or a waitlist-brochure control condition. The present sample served as an ancillary study to the FAB parent study. After launching the FAB trial, we obtained Institutional Review Board approval to add the ART to the 12- and 18-month questionnaire-based assessment. The initial version of the ART was administered to participants in the FAB trial ($N = 267$) at follow-up after their written consent. Given that the current study sought to test the initial psychometric properties of the ART, those randomly assigned to the experimental prevention program and control conditions of the parent trial were collapsed into one group to maximize statistical power.

The initial version of the ART was administered to $N = 267$ women athletes from three universities who were participating in the

FAB Project (Stewart et al., 2017). The majority of the sample was European American (76.7%), followed by African American (16.9%), Asian American (1.9%), Native Hawaiian (0.8%), and Native American (0.8%). See Table 1 for additional demographic information. The mean EDE-Q global score for the FAB sample was 0.98, indicating, on average, this group had relatively low levels of ED psychopathology [note that past non-clinical samples of undergraduate women had mean EDE-Q global scores ranging from 1.54 to 1.74 with standard deviations ranging from 1.29 to 1.32 [(Kelly, Cotter, & Mazzeo, 2012; Luce, Crowther, & Pole, 2008; Mond, Hay, Rodgers, Owen, & Beumont, 2004)].

2.4 | ED patient sample

An independent sample of athletes with EDs were recruited from a specialized ED treatment program for athletes. Levels of care included residential treatment, partial hospitalization, intensive outpatient, and transitional living. Treatment included group therapy, skills training, mindfulness/meditation, yoga, art, and an opportunity to continue training for sport (e.g., weight conditioning) in a supervised environment. All patients with an ED were asked to provide signed informed consent indicating that assessments completed during treatment could be used for research purposes. The ART was administered to women athletes seeking treatment for an ED at admission ($N = 65$) as well as at discharge for a subset of athletes ($n = 33$). ED diagnoses were established by clinicians using the *Diagnostic and Statistical Manual of Mental Disorders- Fourth Edition [DSM-IV]*; (American Psychiatric Association, 1994). Diagnoses included anorexia nervosa ($n = 43$; 66.2%), bulimia nervosa ($n = 11$; 16.9%), and ED not otherwise specified ($n = 11$; 16.9%), which included binge ED. See Table 1 for additional demographic information.

2.5 | Measures

2.5.1 | Athletes' relationships with training scale

[ART]. The initial item pool of the ART consisted of 35 items to measure athletes' training behaviors and beliefs. Items were written to capture training amount and rigidity as well as athletes' motivations, emotions, and attitudes towards training. Items that assessed the frequency of training behaviors and beliefs were measured on a Likert scale from 1 (*Never*) to 5 (*Always*). Items that assessed the level of agreement that athletes had with statements about training were measured on a Likert scale from 1 (*Strongly Agree*) to 5 (*Strongly Disagree*).

2.5.2 | Eating disorder examination-questionnaire

[EDE-Q; (Fairburn & Beglin, 2008)]. The EDE-Q is a self-report version of the EDE semi-structured interview (Cooper & Fairburn, 1987). EDE-Q items assess four subscales of ED psychopathology including Restraint, Eating Concern, Shape Concern, and Weight Concern, which can be summed to obtain a global score. The EDE-Q has been tested in college-aged students and has demonstrated good internal consistency and discriminant validity from measures of depression and self-esteem (Grilo, Reas, Hopwood, & Crosby, 2015; Luce & Crowther, 1999). The EDE-Q has been used to detect ED- or

TABLE 1 Sample characteristics

	FAB sample (N = 267)		Treatment sample (N = 65)	
	Mean	SD	Mean	SD
Age (years)	19.07	0.95	19.35	3.20
Body mass index	22.29	2.66	19.00	3.67
Sport	n	%	n	%
Track and field/cross country	64	24.1%	28	40.0%
Swimming/diving	34	12.8%	3	4.6%
Soccer	32	12.0%	9	13.8%
Volleyball	29	10.9%	3	4.6%
Baseball/softball	21	7.9%	1	1.5%
Basketball	21	7.9%	1	1.5%
Cheerleading	17	6.4%	—	—
Tennis	14	5.3%	1	1.5%
Lacrosse	11	4.1%	1	1.5%
Gymnastics	9	3.4%	3	4.6%
Golf	8	3.0%	2	3.1%
Field hockey	6	2.3%	3	4.6%
Rowing	—	—	3	4.6%
Dance	—	—	4	6.2%
Bodybuilding/cross fit	—	—	1	1.5%
Squash	—	—	1	1.5%
Waterskiing	—	—	1	1.5%
Figure skating	—	—	1	1.5%
Equestrian	—	—	1	1.5%

Note. FAB = female athlete body project; SD = standard deviation. In the FAB sample, most athletes were participating in a Division I sport (61.7%), although some were participating in Division II (7.5%) or Division III (30.8%) athletics. In both the FAB sample and the ED patient sample, the self-reported sport indicates each athlete's primary sport.

disordered-eating psychopathology among athletes (Nichols, Rauh, Barrack, Barkai, & Pernick, 2007; Nichols, Rauh, Lawson, Ji, & Barkai, 2006; Pernick et al., 2006; Rauh et al., 2010), although there is limited research addressing whether the psychometric properties of the EDE-Q are comparable to non-athlete samples. Internal consistency reliability of the EDE-Q in the FAB parent study was adequate to good, with Cronbach's alpha ranging from 0.77 to 0.94 (Stewart et al., 2017). The EDE-Q was administered to the FAB sample and was included in the current study to assess whether the ART had convergent validity with a measure of ED psychopathology.

2.5.3 | Positive and negative affect scale-revised

[PANAS-X; (Watson & Clark, 1999)]. The PANAS-X is an extended version of the original PANAS (Watson, Clark, & Tellegen, 1988). It is comprised of 11 scales that measure two general dimensions of affect, including positive affect and negative affect. Participants in the FAB sample were asked to report the extent to which they experienced 17 emotions related to fear, guilt, and sadness during the past few weeks. The PANAS-X has been administered to undergraduate students and demonstrated convergent validity (Watson & Clark, 1999) with the Profile of Mood States Measure (McNair, Lorr, & Droppelman, 1971). Internal consistency reliability of the PANAS-X in the FAB parent study was excellent [$\alpha = 0.93$; (Stewart et al., 2017)]. The PANAS-X was administered to the FAB sample and used in the current study to assess discriminant validity with the ART.

2.5.4 | Health survey utilization scale

[HSUS; (Stice, Shaw, Burton, & Wade, 2006)]. The HSUS assesses how many hours individuals spent talking to a healthcare professional about various health domains. In the FAB sample, five health domains were assessed, including: physical health; mental health; weight problems; ED or body image concern; and any other personal problem. Participants were asked to indicate how often they talked to a primary-care doctor, psychiatrist, nurse, therapist, or support group about each domain in the past year. Internal consistency reliability for the HSUS in the parent study was good with Cronbach's alpha ranging from 0.77 to 0.87 (Stewart et al., 2017). In the current study, the HSUS was administered to the FAB sample to test the ART's criterion validity.

2.6 | Statistical analyses

Data were analyzed using SAS, SPSS, Mplus, and R. Prior to data analysis, we used mean-based, maximum likelihood multiple imputation (averaged over 11 imputations) in SAS, on a questionnaire-by-questionnaire basis, if 10% or less of data for a questionnaire were missing. The ART and the PANAS were the only scales imputed in the current study.

2.6.1 | Scale development

Data analyses proceeded in several steps: First, we conducted parallel analysis using the "paran" package in R. Parallel analysis is a form of

Monte-Carlo simulation that provides information to help gauge when the number of factors extracted is no longer statistically meaningful (O'Connor, 2000). Second, using information from parallel analysis, we conducted a series of exploratory factor analyses (EFA) to identify the preliminary structure of the ART using FAB 12-month follow-up data. EFAs were carried out in Mplus using the GEOMIN rotation and robust maximum likelihood estimation (WLSMV), which is an appropriate estimator for analyzing ordinal data. Consistent with expert recommendations for scale development, we eliminated items that loaded below |0.40| on the primary factor or had cross-loadings on other factors that were above |0.30| (Clark & Watson, 1995). Items were selected based on a sequential EFA procedure in which a new EFA was conducted each time an item was eliminated due to low item loading. Our analytic strategy allowed us to identify a structure that would ensure that each scale measured one (and only one) construct, rather than blending different constructs within the same scale.

2.6.2 | Construct and criterion-related validity

We used Pearson's *r* to test convergent validity with existing measures of ED psychopathology and excessive exercise using FAB 12-month follow-up data. We also tested discriminant validity by assessing correlations of the ART subscales with a measure of affect using FAB 12-month follow-up data. Construct validity was tested by inspecting the pattern of convergent and discriminant correlations; ART subscales were deemed to show evidence for construct validity if they had significant and strong correlations with similar constructs (i.e., measures of ED psychopathology) and moderate-to-low correlations with different constructs (i.e., measures of affect).

Next, we tested whether ART subscales demonstrated criterion-related validity for predicting healthcare visits among FAB athletes using multiple linear regression. Finally, to test concurrent validity, we computed independent samples *t*-tests to determine whether ART scores were significantly different in athletes without EDs (from the FAB sample) versus athletes who were receiving treatment for an ED (ED patient sample). To identify athletes without ED psychopathology, we included participants from the FAB sample who were within one standard deviation of EDE-Q global score norms in non-clinical undergraduate student samples (Luce & Crowther, 1999). The "compute.es" package in R was used to calculate effect sizes.

2.6.3 | Reliability, stability, and sensitivity to change

To test internal-consistency reliability, we computed Cronbach's alpha, McDonald's omega, and Average Interitem Correlations (AIC) in both the FAB and treatment-seeking samples. Cronbach's alpha

assumes the variance of true scores is constant across all items (tau-equivalency). McDonald's omega, however, makes less assumptions about true score variance. McDonald's omega was calculated in R using the procedure outlined in Dunn, Baguley, & Brunsden (2014). AIC was included because high alphas can be achieved by including redundant items within a subscale. Thus, testing Cronbach's alpha, McDonald's omega, and AIC enabled us to test whether high internal-consistency reliability was achieved only due to redundant content. We tested whether ART subscales were stable over a 6-month period in the FAB sample using intra-class correlations (ICC). We chose to use ICCs, instead of Pearson's *r*, to account for both the consistency of responses over time (i.e., within-person change) and group-level change (i.e., systematic change in mean scores). ICCs were calculated with two-way (two-time points) random-effects models (participants \times months) using absolute agreement for single measures. Paired *t*-tests were used to identify whether the ART subscales demonstrated evidence for sensitivity to change (pre-to-post treatment) in women athletes who received ED treatment. Finally, as an additional test of stability and replicability, we used confirmatory factor analysis (CFA) to evaluate whether subscales identified through EFA at FAB 12-month follow-up were replicated at 18-month follow-up.

3 | RESULTS

3.1 | Exploratory factor analysis

Parallel analyses indicated that a maximum of five factors could be extracted from the 35 items of the initial ART scale. Thus, we tested one-through-five factor solutions using EFA in the 12-month follow-up FAB data. After dropping low-loading and cross-loading items, 15 of the original 35 items were retained. Model-fit indices for all EFA solutions are presented in Table 2. A four-factor solution was the most interpretable and demonstrated an excellent fit to the data. The four factors derived included: (a) Affect-Driven Training; (b) Training Amount; (c) Training Against Medical Advice; and (d) Body Dissatisfaction. Factor loadings for the items from the four-factor EFA are presented in Table 3. The four factors derived from EFA loaded onto a higher-order factor and the fit was good, $\chi^2(86) = 221.246$, $p < .001$, $RMSEA = 0.077$ (90% Confidence Interval [CI] = 0.064, 0.089), $CFI = 0.987$, $TLI = 0.985$, using a WLSMV estimator. Thus, our data suggested that it is reasonable to calculate a total score by summing all 15 items on the ART (A bifactor CFA model was also fitted to the data to test whether the ART subscales explained any additional variance in item responses after the ART Total

TABLE 2 Exploratory factor analysis fit indices

Number of factors extracted	WLSMV χ^2	df	RMSEA	RMSEA 90%CI	CFI	TLI
1	1,305.918*	90	0.225	0.215–0.236	0.887	0.868
2	757.154*	76	0.184	0.172–0.196	0.936	0.912
3	413.785*	63	0.145	0.132–0.158	0.967	0.945
4	131.973*	51	0.077	0.061–0.094	0.992	0.984
5	78.312*	40	0.060	0.040–0.080	0.996	0.991

Note. WLSMV = weight least squares means and variance-adjusted estimation; *df* = degrees of freedom; RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker-Lewis index. Bolded text represents the final four-factor model that was extracted. * $p < .001$.

TABLE 3 Exploratory factor analysis factor loadings

ART© item	Affect-driven Training	Training amount	Training against medical advice	Body dissatisfaction
4	0.848*	-0.020	-0.012	0.049
5	0.859*	-0.004	0.048	-0.009
6	0.505*	0.059	0.198*	0.082
7	0.722*	0.107	0.138*	-0.033
14	0.515*	-0.001	-0.104	0.229*
1	0.011	0.857*	0.003	-0.028
9	0.076	0.878*	-0.002	-0.047
12	-0.249	0.461*	0.219*	0.170*
15	0.127*	0.502*	-0.047	0.180*
2	-0.034	0.218*	0.644*	0.017
3	0.087	0.009	0.787*	0.017
8	0.045	-0.113	0.721	-0.024
10	-0.007	0.007	0.082*	0.924*
11	0.058*	-0.034	0.007	0.978*
13	-0.019	0.050	-0.026	0.850*

Note. Oblique GEOMIN rotation was employed within the exploratory factor analysis. Bolded factor loadings indicate the factor that the corresponding item loaded on. Items 19, 20, 21, 31, 33, and 35 are reverse-coded.

* $p < .01$.

Score was accounted for. The bifactor CFA model was identified and demonstrated good model fit, $\chi^2(75) = 103.124$, $p = .017$, RMSEA = 0.044 (90% CI = 0.019–0.064), CFI = 0.978, TLI = 0.969, indicating that it is useful to keep subscale scores in addition to the ART Total Score. The final 15-item version of the ART along with information about scoring and interpretation of scores can be found at <https://kuscholarworks.ku.edu/handle/1808/26248>.

3.2 | Construct validity

The ART had excellent convergent validity with the EDE-Q; all four subscales and the ART Total Score were significantly (and strongly) correlated with the EDE-Q Restraint, Shape Concern, Weight Concern, and Eating Concern subscales and with the EDE-Q global score. The ART demonstrated excellent discriminant validity, as demonstrated by small-to-moderate correlations with the PANAS-X (see Table 4).

3.3 | Criterion-related validity

The Training Amount ($\beta = 0.126$, $p = .041$) and Training Against Medical Advice subscales ($\beta = 0.129$, $p = .035$) significantly predicted the number of hours athletes spent talking to a healthcare provider about

their physical-health issues within the past year. Athletes' Body Dissatisfaction scores significantly predicted the number of hours they spent talking to health care providers about their mental health concerns ($\beta = 0.177$, $p = .004$) and weight concerns ($\beta = 0.126$, $p = .040$) within the past year. Furthermore, the ART Total Score significantly predicted the number of hours athletes spent talking to a healthcare provider about their physical health ($\beta = 0.141$, $p = .022$) and weight concerns ($\beta = 0.149$, $p = .015$) within the past year.

Treatment-seeking athletes with an ED had significantly higher scores on Affect-Driven Training, Training Amount, Training Against Medical Advice, and the ART Total Score compared to athletes without EDs. Athletes with and without EDs did not significantly differ from one another on Body Dissatisfaction scores (see Table 5).

3.4 | Reliability, stability, and sensitivity to change

For most subscales, the ART demonstrated good internal consistency reliability (see Table 6). We carried out CFA as an additional test of stability to evaluate whether the factor structure of the ART identified through EFA was replicated at FAB 18-month follow-up. Our results showed that the four-factor (subscale) model demonstrated a good fit

TABLE 4 Convergent and discriminant validity results

Measure	Affect-driven training	Training amount	Training against medical advice	Body dissatisfaction	ART Total
PANAS-X	0.059	0.015	0.057	0.003	0.048
EDE-Q global score	0.412**	0.325**	0.171**	0.627**	0.547**
EDE-Q restraint	0.349**	0.254**	0.135*	0.463**	0.432**
EDE-Q shape concern	0.381**	0.308**	0.158*	0.648**	0.530**
EDE-Q weight concern	0.411**	0.338**	0.155*	0.646**	0.554**
EDE-Q eating concern	0.339**	0.272**	0.170**	0.470**	0.445**

Note. PANAS-X = positive and negative affect scale- revised; EDE-Q = eating disorder examination- questionnaire. Values are Pearson r correlations. The PANAS-X was included to assess discriminant validity. Discriminant validity would be demonstrated by low-to-moderate correlations with the ART. The EDE-Q was included to test convergent validity, which would be demonstrated by significant moderate-to-large correlations with the ART. Together, convergent and discriminant validity represent construct validity.

* p value $< .05$; ** p value $< .01$.

TABLE 5 Independent samples *T*-test comparing ART scores of athletes with eating disorders versus athletes without eating disorders

ART subscale	Athletes with EDs Mean (SD)	Athletes without EDs Mean (SD)	<i>T</i> -statistic	<i>p</i> value	Cohen's <i>d</i> [95% CI]
Affect-driven training	19.02 (4.49)	14.66 (4.43)	-7.00	<.001	-0.97 [-1.26, -0.69]
Training amount	13.63 (3.19)	10.18 (3.51)	-7.20	<.001	-1.00 [-1.28, -0.72]
Training against medical advice	8.68 (3.35)	5.82 (2.45)	-6.44	<.001	-0.90 [-1.18, -0.61]
Body dissatisfaction	10.18 (3.63)	10.43 (2.99)	0.50	.622	0.07 [-0.20, 0.34]
ART Total	51.62 (11.02)	41.08 (7.30)	-7.20	<.001	-1.00 [-1.29, -0.72]

Note. EDs = eating disorders; SD = standard deviation; CI = confidence interval. *N* = 65 for athletes with EDs and *N* = 251 for athletes without EDs. The athletes with EDs group was comprised from the ED patient sample, and the ART was administered at treatment admission. Athletes without EDs were sampled from the Female Athlete Body Project prevention program study and were included in these analyses if their Eating Disorder Examination- Questionnaire Global Score were within college student norms (scores \leq two standard deviations above college student norms).

to the data, $WLSMV \chi^2 (84) = 215.350, p < .001, RMSEA = 0.090$ (90% CI = 0.075, 0.105), $CFI = 0.986, TLI = 0.983$. A diagram of the CFA is presented in Figure 1.

Finally, we tested sensitivity-to-change in women athletes who received ED treatment. Paired samples *t*-tests indicated that the ART Total Score, as well as the four subscale scores, were significantly lower at discharge compared to admission (see Table 7).

4 | DISCUSSION

The purpose of this study was to develop and validate a measure of unhealthy training behaviors and beliefs. The Athletes' Relationships with Training Scale (ART) is significant as one of the first self-report tools that was developed in a sample of athletes to identify unhealthy training. The ART may be used to: (a) screen athletes for EDs/disordered eating within a sport context by sport medicine staff or sport psychologists, (b) monitor how unhealthy training behaviors and beliefs change over the course of ED treatment, and (c) research unhealthy training behaviors and beliefs in athletes.

Four subscales of the ART were identified: (a) Affect-Driven Training assesses negative affect associated with training or lack-of-training, (b) Training Amount evaluates training beyond scheduled practices and coach recommendations, (c) Training Against Medical Advice assesses training when injured or against medical recommendations, and (d) Body Dissatisfaction assesses training to acquire a

certain body type or to increase body satisfaction. Additionally, the four-factor structure of the ART was stable over time.

The ART demonstrated strong construct validity as evidenced by convergent validity with a well-validated measure of ED psychopathology and discriminant validity from a measure of affect. Most ART subscales showed evidence for acceptable internal consistency. Despite strong internal-consistency reliability, the Body Dissatisfaction subscale appeared to achieve high internal consistency through the inclusion of redundant item content. The Body Dissatisfaction subscale was reduced to three items due to a small initial item pool. All three items comprising the Body Dissatisfaction subscale begin with the phrase "No matter how hard I train, ..." which could explain the high AIC coefficient.

The ART demonstrated criterion validity by predicting how much time athletes spent talking to medical professionals about their physical health, mental health, and weight concerns. Almost all ART scores (with the exception of Body Dissatisfaction) were significantly higher for athletes with EDs compared to athletes without EDs suggesting that the ART is useful for identifying training attitudes and behaviors that are typically associated with an ED. ART scores were sensitive to change during treatment as indicated by patient scores being lower at discharge compared to admission.

Several findings, however, were unexpected. It was interesting that athletes with EDs had similar levels of Body Dissatisfaction compared to athletes without EDs. Thus, considering Body Dissatisfaction subscale scores alone was insufficient for identifying athletes with EDs in

TABLE 6 ART reliability and stability

	FAB sample					Treatment sample			
	α	ω_H	ω_H 95% CI	AIC	ICC	α	ω_H	ω_H 95% CI	AIC
Affect-driven training	0.83	0.84	[0.80, 0.87]	0.49	0.62	0.79	0.81	[0.66, 0.88]	0.42
Training amount	0.76	0.78	[0.73, 0.81]	0.44	0.67	0.54	0.77	[0.60, 1.0]	0.22
Training against medical advice	0.70	0.71	[0.63, 0.77]	0.44	0.58	0.84	0.84	[0.76, 0.90]	0.63
Body dissatisfaction	0.93	0.93	[0.91, 0.95]	0.81	0.69	0.91	0.92	[0.88, 0.95]	0.78
Total	0.86	0.86	[0.79, 0.90]	0.29	0.70	0.87	0.86	[0.77, 0.91]	0.31

Note. FAB = female athlete body project; α = Cronbach's alpha; ω_H = McDonald's Omega; CI = confidence interval; AIC = average interitem correlation; ICC = intra-class correlation. ICCs were calculated for each of the four ART subscales and the total score as a measure of stability between FAB 12-month follow-up and FAB 18-month follow-up. Omega and alpha values > 0.70 indicate acceptable internal consistency. Ideally, AIC values should range between 0.15–0.50 (Clark & Watson, 1995). ICCs ≤ 0.40 are considered poor, 0.40–0.59 are considered fair, 0.60–0.74 are considered good, and 0.75–1.00 are considered excellent (Cicchetti, 1994).

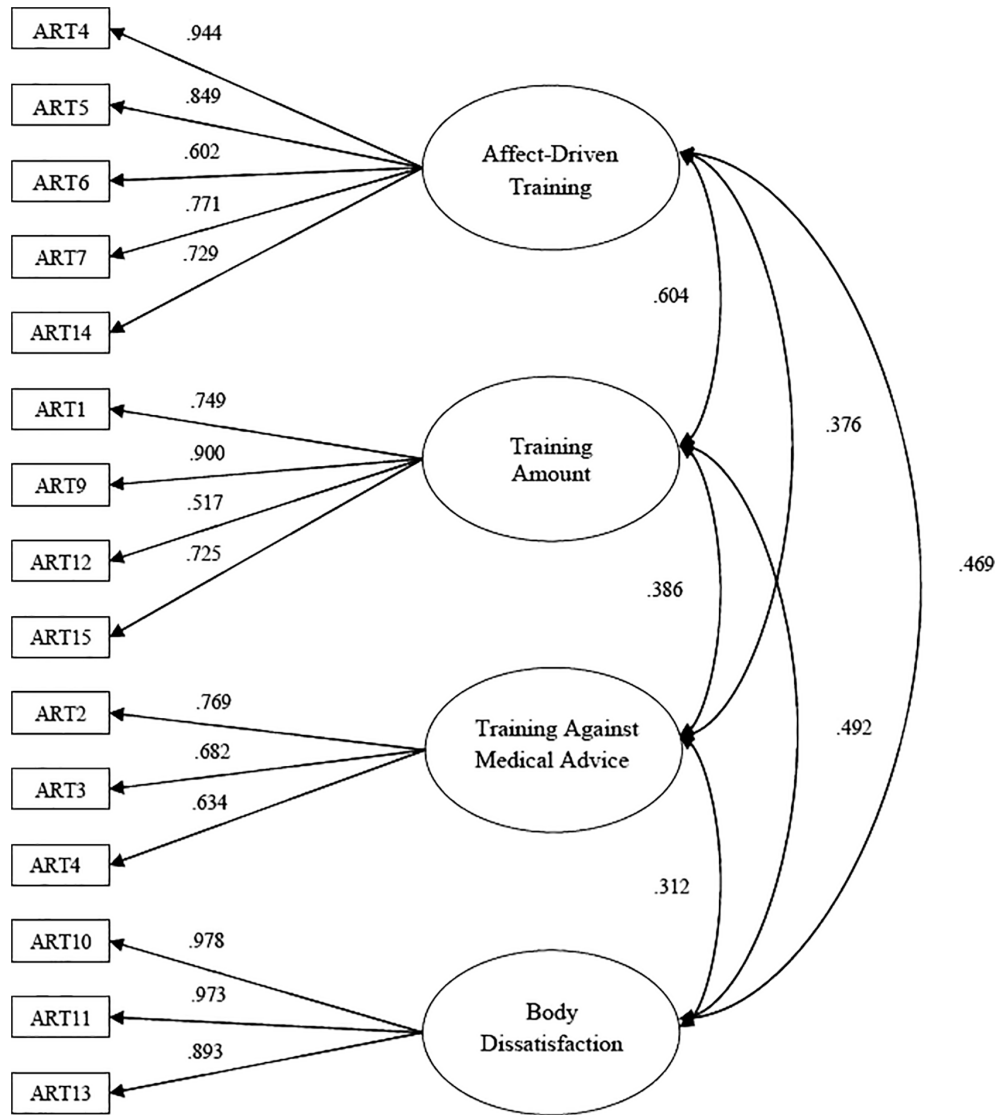


FIGURE 1 ART factor structure this figure displays factor loadings and factor correlations that were identified in the CFA model in 18-month follow-up FAB data. All parameters were significant at $p < .001$

the current study. Body Dissatisfaction scores were still important to consider, however, as high scores significantly predicted time spent discussing mental health and weight concerns with a health care provider.

One possible explanation for why there were no observed differences in Body Dissatisfaction scores in athletes with EDs versus athletes without EDs could be because the sample without EDs was comprised

TABLE 7 Paired sample t-tests of ART scores at treatment admission and treatment discharge

ART	Mean difference	T-statistic	df	p value	Cohen's d [95% CI]
Affect-driven training	-3.26	4.02	30	<.001	0.99 [0.47, 1.51]
Training amount	-1.91	2.82	32	.008	0.69 [0.19, 1.20]
Training against medical advice	-2.36	4.82	32	<.001	1.19 [0.65, 1.72]
Body dissatisfaction	-1.33	2.27	32	.030	0.56 [0.06, 1.06]
ART Total	-8.81	4.30	30	<.001	1.06 [0.53, 1.58]

Note. df = degrees of freedom. CI = confidence interval. Paired sample t-test was used to determine whether athletes who were seeking treatment for an ED (N = 33) had significantly lower ART scores at treatment discharge compared to their ART scores at treatment admission. Mean Difference represents the mean of each ART subscale at discharge subtracted by the mean of each ART scale at admission. Thus, negative mean difference scores indicate that the athletes had lower ART scores (less unhealthy training behaviors) at discharge compared to when they began treatment.

of collegiate athletes whose experiences may not generalize to all athletes. Future studies should assess group differences in Body Dissatisfaction scores among different samples of athletes.

Some limitations are worth noting. First, although the ART demonstrated consistent factor structure, strong construct validity, and criterion validity in two samples of women athletes, this study did not include the Compulsive Exercise Test [CET; (Plateau et al., 2014)]. Inclusion of the CET would have made it possible to compare the ART with another measure of unhealthy training/exercise to test convergent validity as well as incremental validity (i.e., testing whether the ART could predict ED symptoms and health care utilization compared to the CET). In athletes, the CET has three factors including: Weight Control, Avoidance of Negative Affect, and Mood Improvement (Plateau et al., 2014). Affect-Driven Training may measure a similar construct as the CET's Avoidance of Negative Affect and Mood Improvement scales. Body Dissatisfaction may be similar to the Weight Control scale of the CET. Two ART subscales emerged, however, that expanded upon the CET, including: Training Against Medical Advice and Training Amount. Future studies should directly compare the three-factor version of the CET with the ART. Second, the factor structure of the ART was tested in only one sample of women, the FAB sample. Future studies should assess the robustness of the ART's factor structure, reliability, and validity in additional samples including men. Third, we were not able to test how athletes in different sports may vary in their self-reports on the ART due to a small *n* for most sports.

The ART is novel because it is one of the first assessment tools developed to measure exercise attitudes, beliefs, and motivation in athletes. Although other tools have been developed to assess excessive or compulsive exercise (e.g., the CET), the ART is one of the first measures to be developed using a combination of rational and empirical scale development techniques in a sample of athletes.

In conclusion, the current study developed a tool to assess and identify unhealthy training beliefs and behaviors that are likely to result in a physical health concern or ED. The ART demonstrated strong psychometric properties in two groups of women with and without EDs. Thus, the ART is a useful tool to understand when an athlete's training behaviors and beliefs may put them at risk for needing medical attention or developing an ED. The ART could be administered by sports medicine staff and sports psychologists. High scores on the ART Total Score and the Training Amount, Training Against Medical Advice, and Affect-Driven Training subscales indicate that an athlete may be engaging in behaviors that are associated with an ED. With this information, sport medicine staff and sport psychologists may be able to identify unhealthy training behavior early, which may speed up the process of treatment referrals and ultimately reduce the number of injuries experienced by athletes (see <https://kusolarworks.ku.edu/handle/1808/26248> for interpretation guidelines). In conclusion, the ART could provide useful clinical information on the relationship between the athlete's sport participation and EDs, which could be helpful in determining how to best treat the ED and to determine when (or if) the athlete should return to sport training and competition.

ACKNOWLEDGMENTS

This study was funded by a National Institutes of Mental Health (NIMH) Grant: 1 RO1 MH094448-01 (PI: Stewart and Co-PI: Becker).

ORCID

Danielle A. N. Chapa  <https://orcid.org/0000-0001-6946-4194>

Kelsey E. Hagan  <https://orcid.org/0000-0001-5723-3591>

Kelsie T. Forbush  <https://orcid.org/0000-0002-5900-4204>

Carolyn B. Becker  <https://orcid.org/0000-0002-6369-2185>

REFERENCES

- Ackard, D. M., Brehm, B. J., & Steffen, J. J. (2002). Exercise and eating disorders in college-aged women: Profiling excessive exercisers. *Eating Disorders*, 10(1), 31–47. <https://doi.org/10.1080/106402602753573540>
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Beumont, P. J. V., Arthur, B., Russell, J. D., & Touyz, S. W. (1994). Excessive physical activity in dieting disorder patients: Proposals for a supervised exercise program. *International Journal of Eating Disorders*, 15(1), 21–36. [https://doi.org/10.1002/1098-108X\(199401\)15:1<21::AID-EAT2260150104>3.0.CO;2-K](https://doi.org/10.1002/1098-108X(199401)15:1<21::AID-EAT2260150104>3.0.CO;2-K)
- Bompa, T. O. (1983). *Theory and methodology of training: The key to athletic performance*. Toronto: Human Kinetics.
- Byrne, S., & McLean, N. (2002). Elite athletes: Effects of the pressure to be thin. *Journal of Science and Medicine in Sport*, 5(2), 80–94. [https://doi.org/10.1016/S1440-2440\(02\)80029-9](https://doi.org/10.1016/S1440-2440(02)80029-9)
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Cooper, Z., & Fairburn, C. (1987). The eating disorder examination: A semi-structured interview for the assessment of the specific psychopathology of eating disorders. *International Journal of Eating Disorders*, 6(1), 1–8. [https://doi.org/10.1002/1098-108X\(198701\)6:1<1::AID-EAT2260060102>3.0.CO;2-9](https://doi.org/10.1002/1098-108X(198701)6:1<1::AID-EAT2260060102>3.0.CO;2-9)
- Currie, A. (2010). Sport and eating disorders—understanding and managing the risks. *Asian Journal of Sports Medicine*, 1(2), 63–68.
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Fairburn, C. G., & Beglin, S. (2008). Eating disorder examination questionnaire (EDE-Q 6.0). An overview. In C. G. Fairburn (Ed.), *Cognitive behavior therapy and Eating Disorders* (pp. 309–313). New York, NY: Guilford Press.
- Forbush, K. T., Wildes, J. E., Pollack, L. O., Dunbar, D., Luo, J., Patterson, K., ... Watson, D. (2013). Development and validation of the eating pathology symptoms inventory (EPSI). *Psychological Assessment*, 25(3), 859–878. <https://doi.org/10.1037/a0032639>
- Fortes, L. D. S., Kakeshita, I. S., Almeida, S. S., Gomes, A. R., & Ferreira, M. E. C. (2014). Eating behaviours in youths: A comparison between female and male athletes and non-athletes. *Scandinavian Journal of Medicine and Science in Sports*, 24(1), e62–e68. <https://doi.org/10.1111/sms.12098>
- Galli, N., & Reel, J. J. (2009). Adonis or hephaestus? Exploring body image in male athletes. *Psychology of Men and Masculinity*, 10(2), 95–108. <https://doi.org/10.1037/a0014005>
- Garner, D. M. (2004). *Eating disorder inventory-3 professional manual*. Lutz, FL: Psychological Assessment Resources.
- Grilo, C. M., Reas, D. L., Hopwood, C. J., & Crosby, R. D. (2015). Factor structure and construct validity of the eating disorder examination-questionnaire in college students: Further support for a modified brief version. *International Journal of Eating Disorders*, 48(3), 284–289. <https://doi.org/10.1002/eat.22358>

- Holm-Denoma, J. M., Scaringi, V., Gordon, K. H., Van Orden, K. A., & Joiner, T. E. (2009). Eating disorder symptoms among undergraduate varsity athletes, club athletes, independent exercisers, and nonexercisers. *International Journal of Eating Disorders*, 42(1), 47–53. <https://doi.org/10.1002/eat.20560>
- Kelly, N. R., Cottler, E. W., & Mazzeo, S. E. (2012). Eating disorder examination questionnaire (EDE-Q): Norms for black women. *Eating Behaviors*, 13(4), 429–432. <https://doi.org/10.1016/j.eatbeh.2012.09.001>
- Luce, K. H., & Crowther, J. H. (1999). The reliability of the eating disorder examination—Self-report questionnaire version (EDE-Q). *International Journal of Eating Disorders*, 25(3), 349–351. [https://doi.org/10.1002/\(SICI\)1098-108X\(199904\)25:3<349::AID-EAT15>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1098-108X(199904)25:3<349::AID-EAT15>3.0.CO;2-M)
- Luce, K. H., Crowther, J. H., & Pole, M. (2008). Eating disorder examination questionnaire (EDE-Q): Norms for undergraduate women. *International Journal of Eating Disorders*, 41(3), 273–276. <https://doi.org/10.1002/eat.20504>
- Martinsen, M., Holme, I., Pensgaard, A. M., Torstveit, M. K., & Sundgot-Borgen, J. (2014). The development of the brief eating disorder in athletes questionnaire. *Medicine and Science in Sports and Exercise*, 46(8), 1666–1675. <https://doi.org/10.1249/MSS.0000000000000276>
- Martinsen, M., & Sundgot-Borgen, J. (2013). Higher prevalence of eating disorders among adolescent elite athletes than controls. *Medicine and Science in Sports and Exercise*, 45(6), 1188–1197. <https://doi.org/10.1249/MSS.0b013e318281a939>
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the profile of mood states*. San Diego, CA: Educational and Industrial Testing Services.
- McNulty, K. Y., Adams, C. H., Anderson, J. M., & Affenito, S. G. (2001). Development and validation of a screening tool to identify eating disorders in female athletes. *Journal of the American Dietetic Association*, 101(8), 886–892.
- Meyer, C., Taranis, L., Goodwin, H., & Haycraft, E. (2011). Compulsive exercise and eating disorders. *European Eating Disorders Review*, 19(3), 174–189. <https://doi.org/10.1002/erv.1122>
- Mond, J. M., & Calogero, R. M. (2009). Excessive exercise in eating disorder patients and in healthy women. *Australian and New Zealand Journal of Psychiatry*, 43(3), 227–234. <https://doi.org/10.1080/00048670802653323>
- Mond, J. M., Hay, P. J., Rodgers, B., Owen, C., & Beumont, P. J. V. (2004). Validity of the eating disorder examination questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, 42(5), 551–567. [https://doi.org/10.1016/S0005-7967\(03\)00161-X](https://doi.org/10.1016/S0005-7967(03)00161-X)
- Mountjoy, M., Sundgot-Borgen, J., Burke, L., Carter, S., Constantini, N., Lebrun, C., ... Ljungqvist, A. (2015). Relative energy deficiency in sport (RED-S) clinical assessment tool (CAT). *British Journal of Sports Medicine*, 49(7), 421–423. <https://doi.org/10.1136/bjsports-2014-094559>
- Nagel, D. L., Black, D. R., Leverenz, L. J., & Coster, D. C. (2000). Evaluation of a screening test for female college athletes with eating disorders and disordered eating. *Journal of Athletic Training*, 35(4), 431–440.
- Nichols, J. F., Rauh, M. J., Barrack, M. T., Barkai, H.-S., & Pernick, Y. (2007). Disordered eating and menstrual irregularity in high school athletes in lean-build and nonlean-build sports. *International Journal of Sport Nutrition and Exercise Metabolism*, 17(4), 364–377. <https://doi.org/10.1123/ijsnem.17.4.364>
- Nichols, J. F., Rauh, M. J., Lawson, M. J., Ji, M., & Barkai, H.-S. (2006). Prevalence of the female athlete triad syndrome among high school athletes. *Archives of Pediatrics and Adolescent Medicine*, 160(2), 137–142. <https://doi.org/10.1001/archpedi.160.2.137>
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, 32(3), 396–402. <https://doi.org/10.3758/BF03200807>
- Pernick, Y., Nichols, J. F., Rauh, M. J., Kern, M., Ji, M., Lawson, M. J., & Wilfley, D. (2006). Disordered eating among a multi-racial/ethnic sample of female high-school athletes. *Journal of Adolescent Health*, 38(6), 689–695. <https://doi.org/10.1016/j.jadohealth.2005.07.003>
- Plateau, C. R., Shanmugam, V., Duckham, R. L., Goodwin, H., Jowett, S., Brooke-Wavell, K. S. F., ... Meyer, C. (2014). Use of the compulsive exercise test with athletes: Norms and links with eating psychopathology. *Journal of Applied Sport Psychology*, 26(3), 287–301. <https://doi.org/10.1080/10413200.2013.867911>
- Powers, P., & Thompson, R. (2008). *The exercise balance: What's too much, what's too little, and what's just right for you*. Carlsbad, CA: Gürze Books.
- Rauh, M. J., Nichols, J. F., & Barrack, M. T. (2010). Relationships among injury and disordered eating, menstrual dysfunction, and low bone mineral density in high school athletes: A prospective study. *Journal of Athletic Training*, 45(3), 243–252. <https://doi.org/10.4085/1062-6050-45.3.243>
- Rosendahl, J., Bormann, B., Aschenbrenner, K., Aschenbrenner, F., & Strauss, B. (2009). Dieting and disordered eating in German high school athletes and non-athletes. *Scandinavian Journal of Medicine and Science in Sports*, 19(5), 731–739. <https://doi.org/10.1111/j.1600-0838.2008.00821.x>
- Seigel, K., & Hetta, J. (2001). Exercise and eating disorder symptoms among young females. *Eating and Weight Disorders: Studies on Anorexia, Bulimia and Obesity*, 6(1), 32–39. <https://doi.org/10.1007/BF03339749>
- Smolak, L., Murnen, S. K., & Ruble, A. E. (2000). Female athletes and eating problems: A meta-analysis. *International Journal of Eating Disorders*, 27(4), 371–380. [https://doi.org/10.1002/\(SICI\)1098-108X\(200005\)27:4<371::AID-EAT1>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1098-108X(200005)27:4<371::AID-EAT1>3.0.CO;2-Y)
- Stewart, T. M., Pollard, T., Hildebrandt, T., Beyl, R., Wesley, N., Kilpela, L. S., & Becker, C. B. (2017). The female athlete body (FAB) study: Rationale, design, and baseline characteristics. *Contemporary Clinical Trials*, 60, 63–71. <https://doi.org/10.1016/j.cct.2017.06.005>
- Stice, E., Shaw, H., Burton, E., & Wade, E. (2006). Dissonance and healthy weight eating disorder prevention programs: A randomized efficacy trial. *Journal of Consulting and Clinical Psychology*, 74(2), 263–275. <https://doi.org/10.1037/0022-006X.74.2.263>
- Sundgot-Borgen, J. (1993). Prevalence of eating disorders in elite female athletes. *International Journal of Sport Nutrition*, 3(1), 29–40. <https://doi.org/10.1123/ijsn.3.1.29>
- Sundgot-Borgen, J., & Garthe, I. (2011). Elite athletes in aesthetic and Olympic weight-class sports and the challenge of body weight and body compositions. *Journal of Sports Sciences*, 29(suppl 1), S101–S114. <https://doi.org/10.1080/02640414.2011.565783>
- Sundgot-Borgen, J., & Torstveit, M. K. (2004). Prevalence of eating disorders in elite athletes is higher than in the general population. *Clinical Journal of Sport Medicine*, 14(1), 25–32.
- Taranis, L., Touyz, S., & Meyer, C. (2011). Disordered eating and exercise: Development and preliminary validation of the compulsive exercise test (CET). *European Eating Disorders Review*, 19(3), 256–268. <https://doi.org/10.1002/erv.1108>
- Thompson, R. A., & Sherman, R. T. (2010). *Eating disorders in sport*. New York, NY: Routledge.
- Toro, J., Galilea, B., Martínez-Mallén, E., Salamero, M., Capdevila, L., Mari, J., ... Toro, E. (2005). Eating disorders in Spanish female athletes. *International Journal of Sports Medicine*, 26(8), 693–700. <https://doi.org/10.1055/s-2004-830378>
- Torstveit, M. K., Rosenvinge, J. H., & Sundgot-Borgen, J. (2008). Prevalence of eating disorders and the predictive power of risk models in female elite athletes: A controlled study. *Scandinavian Journal of Medicine & Science in Sports*, 18(1), 108–118. <https://doi.org/10.1111/j.1600-0838.2007.00657.x>
- Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form. Retrieved from https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1011&context=psychology_pubs
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.

How to cite this article: Chapa DAN, Hagan KE, Forbush KT, et al. The Athletes' Relationships with Training scale (ART): A self-report measure of unhealthy training behaviors associated with eating disorders. *Int J Eat Disord*. 2018;51:1080–1089. <https://doi.org/10.1002/eat.22960>