Journal of Athletic Training 2000;35(4):431–440

www.journalofathletictraining.org

Evaluation of a Screening Test for Female College Athletes with Eating Disorders and Disordered Eating

Deborah L. Nagel, PhD, RD, EdS*; David R. Black, PhD, MPH, CHES*; Larry J. Leverenz, PhD, ATC*; Daniel C. Coster, PhD†

*Purdue University, West Lafayette, IN; †Utah State University, Logan, UT

Objective: To develop a screening test to detect female college athletes with eating disorders/disordered eating (ED/DE). No validated eating disorder screening tests specifically for athletes have been available.

Design and Setting: In this cross-sectional study, subjects from a large midwestern university completed 3 objective tests and a structured diagnostic interview.

Measurements: A new test, developed and pilot tested by the researchers (Athletic Milieu Direct Questionnaire, AMDQ), and 2 tests normed for the general population (Eating Disorder Inventory-2, Bulimia Test-Revised) were used to identify ED/DE athletes. A structured, validated, diagnostic interview (Eating Disorder Examination, version 12.0D) was used to determine which test was most effective in screening female college athletes.

Subjects: Subjects included 149 female athletes, ages 18 to 25 years, from 11 Division I and select club sports.

Results: ED/DE subjects (35%) were found in almost every sport. Of the ED/DE subjects, 65% exhibited disordered eating, 25% were bulimic, 8% were classified as eating disordered not otherwise specified (NOS), and 2% were anorexic. The AMDQ more accurately identified ED/DE than any test or combination of items. The AMDQ produced superior results on 7 of 9 epidemiologic analyses; sensitivity was 80% and specificity was 77%, meaning that it correctly classified approximately 4 of every 5 persons who were truly exhibiting an eating disorder or disordered eating.

Conclusions: We recommend that the AMDQ subsets, which met statistical criteria, be used to screen for ED/DE to enable early identification of athletes at the disordered eating or NOS stage and to initiate interventions before the disorder progresses.

Key Words: anorexia, bulimia, dieting behaviors, Eating Disorder Examination

Disordered eating is one of 3 components of a serious syndrome called the female athlete triad. Its interrelated components include disordered eating, amenorrhea, and osteoporosis.[1] Research has shown that sports emphasizing low body weight pressure female athletes to achieve and maintain extremely unrealistically low body weights and body fat percentages.[1] Athletes are 2 to 3 times more likely than nonathletes (ie, general population and college students) to manifest characteristics of eating disorders.[2] In 1 study of 22 colleges and universities (n = 695 athletes), approximately 3% of the athletes met the medical criteria for anorexia nervosa, and 21% met the criteria for bulimia nervosa.[2] A substantial number of athletes (as high as 62%) practice pathogenic weight-control behaviors.[3–6] Despite serious medical complications (eg, bradycardia, electrolyte abnormalities, dehydration, dental erosion, hypotension)[7] associated with eating disorders and disordered eating, no screening test has been developed specifically for an athlete population. Current screening tests such as the Eating Disorders Inventory-2 (EDI-2)[8,9] (Psychological Assessment Resources, Inc, Odessa, FL) and the Bulimia Test-Revised (BULIT-R)[10,11] (M.H. Thelen, Columbia, MO) have not been validated with athletes, and the sensitivity and specificity of these tests in athletes are questionable. Wilmore,[12] for example, described 1 study that used the EDI to assess 14 female distance runners. The EDI identified only 3 of these athletes as having possible problems but not clear eating disorders. Seven runners, however, were subsequently diagnosed as having an eating disorder that required inpatient or outpatient treatment, or both. Wilmore[12] also noted similar conclusions in another study in which the Eating Attitudes Test (EAT) was administered to 110 elite female athletes. Based on the EAT, none of the athletes scored in the eating-disordered range, yet 18 (16.4%) received either inpatient or outpatient treatment for eating disorders in the subsequent 2-year period. O'Connor et al[13] also concluded that the EDI-2 can be easily faked and that response bias should be accounted for when using the EDI-2. These studies illustrate the diagnostic problems associated with the use of current, commercial, validated screening tests for eating disorders for the general population when they are used in female athletes.

The purpose of our study was to develop a screening test for eating disorders/disordered eating (ED/DE) specifically for female college athletes. The study was conducted as part of an initiative to develop a screening test for widescale distribution. The following research questions were addressed in this study. (1) Because conventional diagnostic and screening tools are

Address correspondence to David R. Black, PhD, MPH, CHES, Department of Health, Kinesiology, and Leisure Studies, Purdue University, 1362 Lambert Building, West Lafayette, IN 47907-1362. E-mail address: blackran@gte.net

not normed for athletes, are they the most effective tools to use in an athlete population? (2) Will specific items from the Athletic Milieu Direct Questionnaire (AMDQ), which we developed, or the 2 commercial screening tests not normed for athletes be useful in the development of a screening test specifically designed for female college athletes? We hypothesized that the AMDQ would more accurately assess the presence of ED/DE among female college athletes than the EDI-2 or BULIT-R, as verified by a systematic, psychometrically validated, diagnostic interview. Alternatively, we hypothesized that some combination of items from the AMDQ, EDI-2, and BULIT-R would more accurately assess the presence of ED/DE than any of these tests independently. Our extended purpose was to develop a new ED/DE screening test for female college athletes using items from the test pool (ie, the AMDQ, EDI-2, and BULIT-R) that best discriminated athletes with ED/DE from athletes without these disorders. The end product, therefore, would be a shorter test, specific for female college athletes, which could be used to screen for ED/DE.

METHODS

Subjects

The subjects' (n = 149) mean age was 20 years. Self-reported and observed weights were 61.24 kg (135 lb) and 62.14 kg (137 lb), respectively. Mean self-reported and observed heights were both 167.64 cm (66 in), body mass index (BMI, kg/m^2) was 22, and body fat percentage was 18. Subjects were athletes from a large midwestern Division I university and were recruited from all sports (n = 11) involving female athletes (ie, basketball, cheerleading, dance company, modern dance, golf, gymnastics, softball, swimming, tennis, track and cross-country, and volleyball). A census-selection procedure was used because the total number of participants was manageable in terms of the human and economic resources available to conduct the study. The institutional office of the Committee on the Use of Human Research Subjects approved the study and the procedures used for data collection.

Procedures

Test Administration. Data were collected in 2 sessions. In session 1, all subjects completed 3 tests in the order AMDQ, EDI-2, and BULIT-R. The time of completion of all 3 tests at 1 sitting was approximately 1 hour. The AMDQ was constructed because published eating disorder screening tests do not focus on athletes or the effect of the athletic environment on the athlete. Items for the AMDQ were developed based on Black[14]; Brownell et al[15]; Holliman[16]; Thompson and Sherman[17]; the research literature; and the diagnostic criteria for an eating disorder found in the *Diagnostic and Statistical Manual of Mental Disorders,* 3rd edition (*DSM-III*),[18] and the *Diagnostic and Statistical Manual of Mental Disorders,* 4th edition (*DSM-IV*).[19] Additional items were included pertaining to the effect of the athletic milieu on weight management and eating behaviors. Several demographic questions also were included. No attempt was made to disguise the purpose or intention of the objective test, and the response format was multiple choice. Extensive pilot testing of the AMDQ (described in the next column) was completed before this study to establish its psychometric properties (ie, readability, response bias, content validity, test-retest reliability, and criterion validity).

Pilot testing of the AMDQ was conducted in 4 stages over a 2-year period using another 175 female collegiate athletes not included in our study sample. In stage 1, we assessed the readability of the AMDQ (5.27 grade level) using the Gunning Fog Formula.[20] Subjects' written ratings and feedback about the clarity and wording of items and the adequacy of response options also were evaluated. In stage 2, we examined response bias by asking subjects whether or not each item would result in an honest answer or an answer that conformed to socially acceptable norms. The AMDQ was then revised based on stage 1 and 2 results. In stage 3, we assessed content validity by soliciting feedback from 3 experts in the areas of eating disorders and athletics using a structured evaluation form. Based on the stage 3 results, further revisions were made. Stage 4 involved evaluation of test-retest reliability and criterion validity. Test-retest reliability was determined by administering the tests to the same group of female athletes on 2 separate occasions, 2 to 4 weeks apart. Criterion validity was estimated by comparing 3 self-report items on the test (ie, height, weight, body fat percentage) with observed data measured by the researchers. The results for each stage of the pilot test clearly met psychometric standards for test construction, and detailed results can be obtained by contacting the authors.

A second session was scheduled to interview subjects to confirm written test results and to obtain physiologic data. We used Fairburn and Wilson's[21] interview questions and procedures (Eating Disorder Examination [EDE], Diagnostic Version, Edition 12.0D) to diagnose an eating disorder. The EDE is a psychometrically valid, systematic, structured interview. Interviewers (n = 2) were trained to conduct the EDE interviews by a licensed clinical psychologist who has extensive experience with eating disorders and athletes. Any athletes who met diagnostic criteria or were experiencing psychological or physical problems related to nutrition or weight management were referred to an appropriate health care provider (eg, registered dietitian, psychologist, physician, or a combination of these). Coaches were excluded from the recruitment and testing sessions to avoid response bias (ie, influencing athlete participation or responses). Many experts working with athletes with eating disorders[17,22,23] have recommended excluding coaches.

Physiologic Measurements. For purposes of subject classification with the EDE, we measured each subject's height, weight, and body fat percentage. Height was measured to the nearest 0.64 cm (0.25 in) and weight to the nearest 0.23 kg (0.5 lb) using a Detecto balance-beam scale (Detecto Scale Co, Webb City, MO). Body composition was calculated using Jackson and Pollock's formula[24] and measured using Harpenden (Burgess Hill, West Sussex, England) skinfold calipers. The body sites selected for assessment were the triceps, ilium, abdomen, and thigh. The same experienced laboratory technician took 3 skinfold measurements at each site. BMI was computed from height and weight measurements as another indicator of body composition.

Subject Classification. Subjects were classified according to *DSM-III-R*[18] and *DSM-IV*[19] criteria as having anorexia nervosa, bulimia nervosa, or an eating disorder not otherwise specified (NOS) based on responses to key EDE questions (Table 1). For anorexia nervosa, subjects had to meet 5 criteria. To satisfy criteria 1 and 2 for anorexia (fear of weight gain and feelings of fatness), they had to score 4, 5, or 6 on these EDE

Table 1. Subject Classification According to *DSM[a]-III-R*[13] and *DSM-IV*[14] Diagnostic Criteria for Eating Disorders

Criteria	Required Rating					
	Anorexia Nervosa	Bulimia Nervosa	NOS-Anorexic[b]	NOS-Bulimic[c]	Disordered Eating Major Criteria[d]	Disordered Eating Minor Criteria[e]
Fear of Weight Gain	4, 5, or 6[f] (for 3 mo)		3 (for 3 mo)			3 (for 3 mo)
Feelings of Fatness	4, 5, or 6 (for 3 mo)		3 (for 3 mo)			3 (for 3 mo)
Maintained Low Weight	1[g] (for 3 mo)		1 (for 3 mo)			2 (for 3 mo)
Menstruation (absence of)	0 or 7[h] (for 12 mo)		1–8 (for 12 mo)		9 (for 12 mo)	
Body Composition[i]						
1. Body Mass Index (BMI)	<18		18.0–19.9		18.0–19.9	
2. Body fat %	<14		14–17		14–17	
Purging Methods[j]						
1. Vomiting (self-induced)		>2×/wk (for 3 mo)		<1×/wk (for 3 mo)	<1×/wk (for 3 mo)	
2. Laxative misuse		>2×/wk (for 3 mo)		<1×/wk (for 3 mo)	<1×/wk (for 3 mo)	
3. Diuretic misuse		>2×/wk (for 3 mo)		<1×/wk (for 3 mo)	<1×/wk (for 3 mo)	
4. Intense exercise[k]		>3×/wk (for 3 mo)		1–2×/wk (for 3 mo)	1–2×/wk (for 3 mo)	
Objective Bulimic Episodes[l]		>2×/wk (for 3 mo)		<2×/wk (for 3 mo)	<2×/wk (for 3 mo)	
Subjective Bulimic Episodes[m]				<2×/wk (for 3 mo)	<2×/wk (for 3 mo)	
Importance of Shape/Weight		4, 5, or 6 (for 3 mo)		3 (for 3 mo)		3 (for 3 mo)

[a]*DSM, Diagnostic and Statistical Manual of Mental Disorders.*
[b]NOS-anorexic (not otherwise specified) subjects meet any 4 of the 5 criteria for anorexia but with lower scores.
[c]NOS-bulimic subjects meet all 3 criteria for bulimia but to a lesser degree or severity. Subjects classified as NOS-both meet NOS criteria for anorexia and bulimia.
[d]Disordered eating subjects meet 2 of the 4 major criteria: absence of menstruation (amenorrhea), body composition, purging methods, or bulimic episodes.
[e]Disordered eating subjects meet 2 of the 4 minor criteria.
[f]EDE scores can be between 0–6. Scores of 0–2 indicate more "normal" responses (for most items), a score of 3 indicate disordered eating, and scores of 4–6 indicate greater eating disorder symptomatology.
[g]1 = Attempts to lose weight or avoid weight gain because of weight or shape; 2 = attempts to lose weight or avoid weight gain for other reasons.
[h]0 = Absence of menses for 12 months. 7 = subject is taking birth control pills.
[i]Anorexic, NOS-anorexic, and disordered eating subjects satisfy 1 of the 2 criteria (BMI or body fat percentage) for body composition.
[j]Only 1 of the 4 purging methods listed must be used by the subject to meet classification criteria.
[k]Defined as ≥3×/wk for 30 minutes over and above scheduled team practices or competitions.
[l]Defined by the consumption of a large amount of food (eg, 1000 calories or more) in a very short time, with the subject experiencing a loss of control over eating. Bulimic subjects exhibit objective episodes. NOS-bulimic or disordered eating subjects experience either objective or subjective episodes and with less frequency.
[m]Defined as episodes in which the subject "feels" he or she ate too much but really ate a "normal" amount of food. Loss of control over eating is exhibited.

questions. EDE scores can be from 0 to 6; scores of 0–2 indicated more "normal" responses (for most items), while a score of 3 indicated disordered eating, and scores of 4 to 6 indicated greater eating disorder symptoms. For criterion 3, dieting behavior (ie, tried to "maintain low weight"), a score of 1 was required (meaning the subject was trying to lose or maintain weight because of shape or weight concerns). To satisfy criterion 4, amenorrhea, subjects had to score a 0 or 7, meaning they had no menstrual periods in 12 months or that they were on birth control pills (score of 7). To satisfy criterion 5, body composition, subjects had to be less than 14% body fat or have a BMI less than 18.

For bulimia nervosa, subjects had to meet 3 criteria (Table 1). To satisfy criterion 1, subjects had to report use of 1 of 4 purging methods (ie, self-induced vomiting, laxative misuse, diuretic misuse, or intense exercising that was defined as more than 3 times per week for 30 minutes over and above scheduled team practices or competitions). Purging methods had to be used at least twice per week for the past 3 months. To satisfy criterion 2, subjects had to report objective bulimic episodes twice per week for the past 3 months. According to Fairburn and Wilson,[21] objective bulimic episodes (commonly called "binging") are defined by the consumption of a large amount of food (eg, 1000 calories or more) in a very short time, with the subject experiencing a loss of control over eating. Subjective bulimic episodes are those in which subjects "feel" they ate too much, but in reality the amount of food was "normal," and loss of control also was exhibited. Finally, to satisfy criterion 3, subjects had to score 4 to 6 on "importance of shape or weight" (ie, how important any change in shape or weight would be in influencing how they felt about themselves as people).

NOS-anorexic subjects had to meet any 4 of the 5 anorexia criteria, but lower scores were required (Table 1). For criteria 1 and 2, subjects had to score 3 on questions dealing with fear of weight gain and feelings of fatness. For criterion 3, dieting behavior, a score of 1 was required. To satisfy criterion 4,

amenorrhea, subjects had to score 1 through 8 (meaning they had at least 1 menstrual period, but no more than 8 during the last 12 months). To satisfy criterion 5, body composition, subjects had to have 14% to 17% body fat or a BMI of 18 to 19.9.

NOS-bulimic subjects had to meet all 3 criteria for bulimia but to a lesser degree of severity or frequency. To satisfy criterion 1 for NOS-bulimia, purging could occur less frequently than twice per week for the last 3 months. To satisfy criterion 2, objective or subjective bulimic episodes could occur less often than twice per week for the last 3 months. To satisfy criterion 3, subjects had to score 3 on the importance of shape or weight. Some subjects were classified as NOS-both, meaning they met the NOS criteria for both anorexia and bulimia.

The American College of Sports Medicine[1] has defined disordered eating as "a wide spectrum of harmful and often ineffective eating behaviors used in attempts to lose weight or achieve a lean appearance. The spectrum of behaviors ranges in severity from restricting food intake to binging and purging, to the *DSM-IV*–defined disorders of anorexia nervosa and bulimia nervosa." Criteria used in classification of subjects as disordered eating in this study were based on the *DSM-III-R*[18] and *DSM-IV*[19] diagnostic criteria for an eating disorder. Disordered eating subjects had to meet 2 of the 4 major criteria. First, subjects had to display signs of amenorrhea, meaning they had to score 9 of 12 on item 36 of the EDE (ie, they had to have had 9 menstrual periods in the last 12 months). Second, body composition scores had to show an abnormally low percentage of body fat (14% to 17%) or a BMI of 18.0 to 19.9. Third, subjects had to use purging methods (ie, self-induced vomiting, laxative or diuretic misuse) on a regular basis (ie, less than once per week for 3 months). Finally, bulimic episodes (subjective or objective) had to occur on a regular basis (ie, less than once per week for 3 months). Additionally, to be classified as disordered eating, subjects had to respond with 3 on at least 2 of the following minor criteria: importance of shape or weight, fear of weight gain, feelings of fatness, and maintaining a low weight.

Data Analyses

Item Identification. We statistically reduced a large item pool to increase the probability of identifying highly discriminant items.[25,26] The original item pool was composed of items from 3 tests: the AMDQ (119 items), EDI-2 (91 items), and BULIT-R (36 items). Because the EDI-2 uses only 3 of the 11 subscales (bulimia, drive for thinness, body dissatisfaction) for diagnoses of anorexia or bulimia, we only included items from these subscales (23 total) in the data analyses. Logistic regression, along with item analysis techniques, were used to determine the overall sensitivity and specificity of each screening test and the value of each individual item and to compare tests (and individual items from different tests) in terms of their ability to identify athletes with ED/DE. Logistic regression was selected because confounding variables can be controlled statistically and because it is appropriate for analyses of items on different scales of measurement (eg, 4- or 5-level Likert questions and responses on a continuous scale, such as body weight and age). Subjects were classified as ED/DE or OK (not ED/DE) based on EDE interview data. A nominal response variable based on the EDE (0 = no ED/DE, 1 = ED/DE) was used to compare results of the screening tests.

The tests or the best combination of items also were evaluated for their ability to identify each of the 4 types of ED/DE (anorexia, bulimia, NOS, disordered eating), resulting in a multinomial response. Analyses were undertaken using generalized logits to predict group membership. All test items were analyzed for internal consistency with the Cronbach coefficient α.

Item Selection. Items that were selected for analyses had to meet all of the following 5 statistical criteria. The first was mean separation: the mean score for ED/DE subjects on an item had to be significantly ($P < .01$) greater than the mean score for OK (not ED/DE) subjects. This value was used to reduce the number of items (although many other items had significant mean separation, with a P value between 1% and 5%). The second criterion was logistic regression: the item had to be a significant predictor of ED/DE versus OK subjects in a logistic regression model with α set at 1%. The third was correlation with total: the correlation of each retained item with the total score had to be at least 0.4. The fourth criterion was Cronbach α: each item's individual Cronbach α had to be at least 0.85. The fifth was sensitivity and specificity: the sensitivity and specificity had to be "satisfactory," meaning a sensitivity of 80% or greater and specificity of approximately 75 to 80% (criterion 5 above is essentially [1] and [2] of the epidemiologic analyses presented below).

Epidemiologic Analyses. We conducted epidemiologic analyses to evaluate the effectiveness of each test. These calculations included 9 interrelated analyses: (1) sensitivity, (2) specificity, (3) percentage of false-positives, (4) percentage of false-negatives, (5) positive predictive value, (6) negative predictive value, (7) yield, (8) accuracy, and (9) validity.[27–29] Sensitivity is the ability of the test to correctly classify those with the disorder (ie, ED/DE subjects). Specificity is the ability of the test to correctly classify those without the disorder (ie, OK subjects). False-positives are the percentage of subjects without the disorder who test positive. False-negatives are the percentage of subjects with the disorder who test negative. Positive predictive value is the probability that a person who tests positive does have the disorder. Negative predictive value is the probability that a person who tests negative does not have the disorder. Yield is the number of true positives correctly identified (ie, the proportion of true positives divided by the total number of subjects screened). Accuracy is the degree of agreement between the screening test and the gold standard (ie, the EDE) for identifying true-positives and true-negatives. Validity is the ability of a test to give a true measure: how well it measures what it is supposed to measure. Further indications of the validity of the 3 tests include sensitivity, specificity, positive predictive value, and negative predictive value.

RESULTS

Epidemiologic Analyses

The response rate was 85.5%. Results for each of the 9 epidemiologic values described above for the AMDQ, EDI-2, and BULIT-R and a combination of 26 items from all 3 tests called the AEBSC (AMDQ, EDI-2, and BULIT-R subsets combined) are presented in Table 2. Three different versions of the AMDQ and the BULIT-R also are reported in Table 2, for a total of 8 tests. The BULIT-R1 is the 28-item version using the scoring guidelines (cutoffs of 84 and 112) of Thelen et al.[11]

Table 2. Epidemiologic Evaluation (%) of Eating Disorder Assessment Tools

	AMDQ*				BULIT†			
	1	2	3	EDI-2‡	-R1†	-R2	-R3	AEBSC§
Sensitivity	80.00	80.00	82.00	64.00	26.92	69.23	69.23	70.59
Specificity	77.17	75.27	79.57	74.23	98.94	78.72	77.66	73.68
False-Positives	22.83	24.73	20.43	25.77	1.06	21.28	22.34	26.32
False-Negatives	20.00	20.00	18.00	36.00	73.08	30.77	30.77	29.41
Positive Predictive Value	65.57	63.49	68.33	56.14	93.33	64.29	63.16	59.02
Negative Predictive Value	87.65	87.50	89.16	80.00	70.99	82.22	82.02	82.35
Yield	28.17	27.97	28.67	21.77	9.59	24.66	24.66	24.66
Accuracy	78.17	76.92	80.42	70.75	73.29	75.34	74.66	72.60
Validity	57.17	55.27	61.57	38.23	25.86	47.95	46.89	44.27

*AMDQ, Athletic Milieu Direct Questionnaire. Three different versions of the AMDQ are presented. Each version of the AMDQ has a different number of items, with some items common to all 3 tests. The AMDQ 1 has 35 items; AMDQ 2 has 19 items; and AMDQ 3 has 9 items.
†BULIT-R, Bulimia Test-Revised. The BULIT-R1 is the 28-item version using the scoring guidelines (cutoffs of 84 and 112) of Thelen et al.[11] The BULIT-R2 is the 36-item version with a cutoff of 60, and the BULIT-R3 is the 28-item version with cutoff of 60. The cutoffs for the BULIT-R2 and R3 were chosen by a discriminant analysis to improve sensitivity and specificity.
‡EDI-2, Eating Disorder Inventory, 2nd edition.
§AEBSC = AMDQ, EDI-2, and BULIT-R subsets combined. The AEBSC has 26 items.

The BULIT-R2 is the 36-item version with a cutoff of 60, and the BULIT-R3 is the 28-item version with cutoff of 60. The cutoffs for the BULIT-R2 and R3 were chosen by a discriminant analysis to maximize sensitivity and specificity.

Sensitivity (criterion 5) was highest for the 3 AMDQ subsets (80% to 82%), lower for the EDI-2 (64%), and lowest for the BULIT-R1 (27%). Only the 3 AMDQ tests met criterion 5 for sensitivity (> 80%). In contrast, the AEBSC and BULIT-R2 and R3 have values that are moderately high but below the criterion (70%), while the EDI-2 and BULIT-R1 were far below the criterion (64% and 27%, respectively). More specifically, when AMDQ subset 2 was used, 80% (n = 40) of ED/DE subjects were correctly classified, and 10 ED/DE subjects were misclassified as OK (false-negatives) (Table 3). Of the 10 ED/DE subjects who were misclassified, 1 was bulimic, 1 was NOS, and 8 exhibited disordered eating. In contrast, when the BULIT-R1 was used, only 27% (n = 14) of ED/DE subjects were correctly classified, and 38 ED/DE subjects were misclassified as OK. Of the 38 misclassified subjects, 1 was anorexic, 7 were bulimic, 4 were NOS, and 26 exhibited disordered eating.

Specificity values (criterion 5) for all 8 tests were high (74% to 99%) and 5 tests met criterion 5 for specificity (approximately 75% to 80%). These tests included the 3 AMDQ tests, the BULIT-R2, and R3. The values for the EDI-2 and AEBSC (74%) were just under the criterion. The BULIT-R1 (99%) exceeded the criterion; however, it was accompanied by a high false-negative value (73%) and a low false-positive value (1%) (Table 1).

The false-positive value was highest for the EDI-2, AEBSC, and AMDQ subset 2 (25% to 26%) and lowest for the BULIT-R1 (1%). The other tests varied between 20% (AMDQ subset 3) and 23% (AMDQ subset 1). False-positive values for all 8 tests were acceptable; however, a false-positive value of 1% for the BULIT-R1 is not likely to be accurate and is probably due to the extremely high specificity (99%) of this test.

False-negative values were highest for the BULIT-R1 (73%), with moderately high values for the EDI-2 (36%), BULIT-R2 (31%), BULIT-R3 (31%), and AEBSC (29%). The lowest false-negative values were for the 3 AMDQ subsets (18% to 20%). Because of the seriousness of eating disorders,

Table 3. Sensitivity of Screening Tests by ED/DE Classification

	Frequency (%)					
	Anorexia Nervosa	Bulimia Nervosa	Not Otherwise Specified	Eating Disordered	Disordered Eating	Sensitivity
AMDQ 1	1 (100)	12 (92)	3 (75)	16 (89)	24* (75)	40 (80)
AMDQ 2	1 (100)	12 (92)	3 (75)	16 (89)	24* (75)	40 (80)
AMDQ 3	1 (100)	12 (92)	3 (75)	15 (83)	25* (78)	41 (82)
EDI-2	1 (100)	10 (77)	3 (75)	14 (78)	18* (56)	32 (64)
BULIT-R1	0 (0)	6 (46)	0 (0)	6 (33)	8 (23)	14 (27‡)
BULIT-R2	1 (100)	12 (92)	3 (75)	16 (89)	20 (59)	36 (69‡)
BULIT-R3	1 (100)	12 (92)	3 (75)	16 (89)	20 (59)	36 (69‡)
AEBSC	1 (100)	10 (77)	3 (75)	14 (78)	22† (67)	36 (71§)
Total	1	13	4	18	34	50

Note: AMDQ, Athletic Milieu Direct Questionnaire; EDI-2, Eating Disorder Inventory, 2nd edition; BULIT-R, Bulimia Test-Revised; AEBSC, AMDQ, EDI-2, and BULIT-R subsets combined resulted in 3 different versions of the AMDQ. Each version of the AMDQ has a different number of items, with some items common to all 3 tests. The AMDQ 1 has 35 items; AMDQ 2 has 19 items; and AMDQ 3 has 9 items. The BULIT-R1 is the 28-item version using Thelen et al (BULIT-R author) scoring guidelines (cutoffs of 84 and 112). The BULIT-R2 is the 36-item version with a cutoff of 60, and the BULIT-R3 is the 28-item version with cutoff of 60. The cutoffs for the BULIT-R2 and R3 were chosen by a discriminant analysis to improve sensitivity and specificity. The AEBSC is the AMDQ, EDI-2, and BULIT-R subsets combined (26 items total).
*Frequency missing = 2; †Frequency missing = 1; ‡Total number of items = 52; §Total number of items = 51.

a lower false-negative value is desirable in a screening test. The 3 AMDQ tests had the lowest false-negative values (18% to 20%). The BULIT-R3 and AEBSC values (31% and 29%, respectively) are questionable in terms of acceptability, and the EDI-2 value (36%) and BULIT-R1 value (73%) are unacceptable.

Positive predictive value was highest for the BULIT-R1 (93%) and lowest for the EDI-2 (56%). Positive predictive values for the other tests, including the AMDQ, were comparable (63% to 68%). Positive predictive values for all tests were acceptable, except for the EDI-2. However, the high positive predictive value of the BULIT-R1 is misleading and not necessarily desirable because this test also had the lowest negative predictive value (71%), which is of greater concern when screening for eating disorders.

Negative predictive value was highest for the 3 AMDQ subsets (88% to 89%) and lowest for the BULIT-R1 (71%). The EDI-2, BULIT-R2, BULIT-R3, and AEBSC tests all had similar values (80% to 82%). When screening for eating disorders, higher negative predictive values are desirable. Therefore, negative predictive values for the 3 AMDQ tests were preferable to the values for the other 5 tests.

Yield was highest for the 3 AMDQ subsets (28% to 29%) and lowest for the BULIT-R1 (10%). The EDI-2, BULIT-R2, BULIT-R3, and AEBSC had comparable values (22% to 25%). Yields for the 3 AMDQ tests were preferable to yields for the other 5 tests. The highest yield possible for ED/DE subjects (ie, the maximum ED/DE yield) is 34.90% (52/149; 52 ED/DE or true-positive subjects were identified by the EDE of 149 total subjects). Percentage of total maximum yield for the AMDQ tests was 80.14 to 82.15 (AMDQ yield/maximum ED/DE yield; 27.97/34.90 = 80.14; 28.67/34.90 = 82.15), meaning the AMDQ correctly identified approximately 4 of every 5 persons who were truly ED/DE.

Accuracy in identification of true-positives and true-negatives was highest for the 3 AMDQ subsets (77% to 80%) and lowest for the EDI-2, BULIT-R1, and AEBSC (71% to 73%). Accuracy for the 3 AMDQ tests was preferable to that of the other 5 tests.

Validity was highest for the 3 AMDQ subsets (55% to 62%) and lowest for the BULIT-R1 (26%) and EDI-2 (38%). The AEBSC, BULIT-R2, and BULIT-R3 had comparable values (44% to 48%). Validity for the 3 AMDQ tests was preferable to that of the other 5 tests.

AMDQ Subset Analyses

Initial Analyses. A total of 51 items were significant ($P = .01$) on both the mean separation t tests and the logistic regression analyses (criteria 1 and 2). Another 23 items were significant ($P = .05$) on both mean separation and logistic regression. These 23 items were not considered in the latter 3 analyses.

Using the 51 AMDQ items that were significant for criteria 1 and 2, several equally good subsets of AMDQ items potentially meet all 5 criteria. Two examples of such subsets and the corresponding analyses associated with each subset follow. Also included is a third subset, AMDQ subset 3, which did exceptionally well on criterion 5 but does not meet criteria 3 and 4, largely due to its small item set (only 9 items).

AMDQ Subset 1 (35 Items). Correlations with total and Cronbach α (criteria 3 and 4) were both used to demonstrate internal consistency of test items. Most of these items (32/35) met criterion 3, having a correlation with the total of 0.40 or above; 3 items were slightly below. Cronbach α values were all above 0.85 (criterion 4), and the average Cronbach α for raw and standardized variables was 0.9401 and 0.9441, respectively. Raw variables represent actual values of subjects' responses on items. Standardized variables represent values adjusted to a uniform scale of measurement. Based on criteria 3 and 4, 32 of 35 items in AMDQ subset 1 met the criteria for inclusion in the item pool.

AMDQ Subset 2 (19 Items). Most of these items (18/19) met criterion 3, having a correlation with the total of 0.40 or above; 1 item was slightly below. Cronbach α values were all above 0.85 (criterion 4), and the average Cronbach α for raw and standardized variables was 0.9043 and 0.9161, respectively. Based on criteria 3 and 4, 18 of 19 items in AMDQ subset 2 met the criteria for inclusion in the item pool.

AMDQ Item Subset 3 (9 Items). One third of these items did not meet criterion 3, having a correlation with the total of 0.40 or above. None of the Cronbach α values met criterion 4 of 0.85 or above. The average Cronbach α for raw and standardized variables was 0.7587 and 0.7706, respectively. Based on criteria 3 and 4, none of the items in AMDQ subset 3 met the criteria for inclusion in the item pool. This subset, however, had the highest sensitivity and specificity (82% and 79.6%, respectively) of any subset. Therefore, fewer subjects were misclassified with this subset. Cronbach α and item correlations with the total, however, were low. The high sensitivity and specificity of this small item set may reflect an artificially high performance of these items for the current subjects that is unlikely to generalize to other athlete populations because of the lower than acceptable correlations with the total and Cronbach α. These items would be better used as part of a larger item pool, such as the AMDQ subset 1 or 2.

EDI-2 Analyses

Initial Analyses. A total of 15 of 23 diagnostic subscale items were significant ($P = .01$) on both mean separation and logistic regression (criteria 1 and 2). The remaining 8 items were dropped from the item pool, although they were included in the analyses for criteria 3 through 5. Items not meeting criteria 1 and 2 were predominantly from the bulimia subscale (6/7), although 2 items were from the drive-for-thinness subscale. In other words, only 1 of 7 items from the bulimia subscale met criteria 1 and 2 for inclusion in the item pool.

The results for criteria 3 and 4 for the EDI-2 diagnostic subscales are presented below. The results for all 3 subscales are combined and then presented separately.

Three Diagnostic Subscales Combined (23 Items). When all diagnostic subscale items were analyzed as a group, 21 of the 23 items met criterion 3, having a correlation with the total of 0.40 or above (using standardized variables); 1 on the drive-for-thinness subscale and 1 on the bulimia subscale did not meet the criterion. Cronbach α values were all above 0.85 (criterion 4), and the average Cronbach α for raw and standardized variables was 0.9282 and 0.9246, respectively. When all 3 diagnostic subscales were combined for analysis, 21 of 23 EDI-2 items met criteria 3 and 4 for inclusion in the item pool.

Bulimia Subscale (7 Items). All of these items met criterion 3, having a correlation with the total of 0.40 or above. Cronbach α values were all below 0.85 (criterion 4) using raw variables, and only 2 items met the criterion using standardized variables. The average Cronbach α for raw and standardized variables was

0.8252 and 0.8483, respectively. Based on criteria 3 and 4, only 1 of 7 items from the bulimia subscale of the EDI-2 met the criteria for inclusion in the item pool.

Body Dissatisfaction Subscale (9 Items). All of these items met criterion 3, having a correlation with the total of 0.40 or above, and all items had Cronbach α values above 0.85 (criterion 4). The average Cronbach α for raw and standardized variables was 0.9253 and 0.9258, respectively. Based on criteria 3 and 4, all 9 items on the body dissatisfaction diagnostic subscale of the EDI-2 met the criteria for inclusion in the item pool.

Drive-for-Thinness Subscale (7 Items). Most of these items (6/7) met criterion 3, having a correlation with the total of 0.40. The majority of Cronbach α values (5/7) were all above 0.85 (criterion 4). The average Cronbach α for raw and standardized variables was 0.8804 and 0.8767, respectively. Based on criteria 3 and 4, 5 of 7 items on the drive-for-thinness diagnostic subscale of the EDI-2 met the criteria for inclusion in the item pool.

BULIT-R

Initial Analyses. A total of 29 of 36 items were significant ($P = .01$) on both mean separation and logistic regression (criteria 1 and 2). The remaining 7 items were dropped from the item pool, although they were included in the analyses for criteria 3 through 5.

As Table 1 indicates, the primary difference between the 3 versions of the BULIT-R tests is the scoring method. Therefore, we analyzed all 36 BULIT-R items as a group for comparison with criteria 3 and 4.

BULIT-R (36 Items). Five of the 36 items did not meet criterion 3, having a correlation with the total of 0.40 or above, and were dropped from the item pool. Cronbach α values were all above 0.85 (criterion 4), and the average Cronbach α for raw and standardized variables was 0.9514 and 0.9501, respectively. Based on criteria 3 and 4, 31 of 36 BULIT-R items met the criteria for inclusion in the item pool.

AEBSC Subset (26 Items)

This subset included a total of 26 items: 9 from the AMDQ, 9 from the EDI-2, and 8 from the BULIT-R. All of these items were significant ($P = .01$) on both criteria 1 and 2 (mean separation and logistic regression). Additionally, all items met criterion 3, having a correlation with the total of 0.40 or above and all Cronbach α values were above 0.85 (criterion 4). The average Cronbach α for raw and standardized variables was 0.9548 and 0.9580, respectively. Based on criteria 3 and 4, all of the items in the AEBSC subset met the criteria for inclusion in the item pool.

DISCUSSION

The purpose of our study was to develop a screening test specifically for female college athletes with ED/DE. Six major conclusions can be drawn from the results of this study. First, the results supported the hypothesis that the AMDQ more accurately screened eating disorders and disordered eating among female college athletes than the EDI-2 or BULIT-R and established that the AMDQ subsets were more discriminating than any combination of items from all 3 tests. This finding confirms that a screening test such as the AMDQ, which is specifically intended for the athletic population, is required. The AMDQ subsets produced superior results for sensitivity (81%), false-negatives (19%) positive predictive value (66%), negative predictive value (88%), yield (28%), accuracy (79%), and validity (58%) than commercial tests not normed for athletes (Table 2), while maintaining acceptable values for specificity (77%) and false-positives (23%).

Second, the AMDQ subsets were the only tests that met criterion 5 for sensitivity. The AMDQ subsets correctly classified 80% to 82% of ED/DE subjects (sensitivity) and 75 to 80% of OK subjects (specificity) (Tables 2 and 3). In contrast, the 2 valid commercial eating disorder tests correctly classified only 64% and 70% of ED/DE subjects and 74% and 78% of OK subjects (EDI-2[9] and BULIT-R,[11] respectively), while the combined test (AEBSC) correctly classified 71% and 74%, respectively. It should be noted, however, that the BULIT-R[10,11] was designed to screen only for bulimia, and the EDI-2 and BULIT-R were not designed to screen for NOS or disordered eating. The use of the AMDQ with a population of 1000 ED/DE female college athletes who are truly positive for an ED/DE would theoretically result in correct classification of 4 of every 5 (80%, n = 800) athletes, which is superior to the sensitivity achieved using the EDI-2 (64%, n = 640) or BULIT-R (69%, n = 690). Additionally, if a second screening were conducted using different versions of the AMDQ, the number of subjects correctly classified as ED/DE would improve (ie, sensitivity and specificity would increase).

Third, a large percentage of ED/DE subjects were identified in this study, which suggests that ED/DE are serious problems for female college athletes. A total of 35% of the sample was at risk (disordered eating) or had a definite problem (eating disordered) as determined by the EDE diagnostic interview. A higher prevalence of bulimia (25% of the 52 ED/DE subjects, n = 13, and 8.72% of the total sample) than anorexia (2% of ED/DE subjects, n = 1) was noted. Also, a large number of subjects met the criteria for disordered eating (n = 34, 65.38% of ED/DE subjects and 22.82% of the total sample) and 4 (8% of ED/DE subjects) met NOS criteria. These findings are consistent with prior prevalence studies of college students,[30–32] but our ED/DE percentage (35%) was higher than that found in 1 study of college athletes,[33] which focused only on anorexia and bulimia and noted an eating disorder prevalence of 25% among 695 college athletes. Many of the athletes who participated in our study appear to be at risk for future development of an eating disorder, and the eating behaviors and attitudes that were most prevalent were consistent with bulimia rather than anorexia. This is an important consideration for all support personnel closely affiliated with athletes because bulimia tends to be more difficult to detect, especially for someone not trained in the recognition and treatment of eating disorders. This study also provides the first operational definition of disordered eating, which is important in order to advance the field conceptually and empirically.

Fourth, it is noteworthy that all 3 AMDQ subsets met all 5 inclusion criteria. In contrast, the 2 validated published eating disorder tests each had several items that fell short on 1 or more inclusion criteria. For example, of the 3 diagnostic subscales of the EDI-2 (23 items), 8 items did not meet all 5 inclusion criteria (2 from the drive-for-thinness subscale, 6 from the bulimia subscale). Similarly, of 36 BULIT-R items, 7 did not meet all 5 inclusion criteria. All AEBSC items met criteria 1 through 4 but fell slightly short on criterion 5 (sensitivity and specificity).

Fifth, although we reported results for only 3 AMDQ subsets, many additional possible combinations of items could potentially produce similar results. In addition to the 51 AMDQ items that met all 5 statistical criteria, another 23 would be potentially acceptable for retention in the item pool after rewording the questions or changing the response options, or both. An additional advantage to the AMDQ over the current tests is its use in screening not only for anorexia and bulimia but also screening for NOS and disordered eating (which current tests do not consider). Further analysis of EDI-2 and BULIT-R items that did not discriminate well (ie, did not meet all 5 statistical criteria) may be beneficial if a common theme can be determined to explain why they did not perform well as discriminators and if they can be modified after psychometric testing.

Finally, the research literature suggests that the prevalence of ED/DE is highest in sports emphasizing low body weight and leanness.[1] This study also supports this conclusion, but ED/DE subjects were observed in every sport sampled except basketball and softball. Of the 13 bulimic athletes, 3 were involved in cheerleading, 3 in a dance company, 1 in modern dance, 1 in golf, 4 in swimming, and 1 in track. The anorexic athlete was in the dance company, and the 4 NOS subjects were involved (1 each) in cheerleading, gymnastics, swimming, and track. Disordered eating subjects were found in every sport except basketball and softball.

Three potential limitations existed in the study. First, the internal validity of the study might be questioned if subjects were not truthful in their responses to test or interview questions. The AMDQ underwent substantial pilot testing before we conducted our study to evaluate whether each item would produce a response bias. Response bias was an issue for only 9 of the 119 AMDQ items; all 9 items were revised based on the athletes' comments. Additionally, measures were taken to encourage truthful responding. Coaches were excluded from the study, and all subjects were guaranteed confidentiality and were assured that the data would be shared with no one (specifically coaches, athletic trainers, teammates, or parents), except for the research team. Observationally, subjects were open and candid in responses to interview questions, and many provided valuable, unsolicited comments regarding concerns about their eating habits and specific behaviors. Second, content validity might be questioned if the AMDQ did not adequately reflect the categories of ED/DE. AMDQ items, however, were developed based on *DSM-III-R*[18] and *DSM-IV*[19] diagnostic criteria for eating disorders and the research literature regarding the athletic milieu.[14–17,22,23,33–38] Additionally, 3 experts specializing in the area of eating disorders and athletes evaluated content validity of the AMDQ during the pilot test, and content validity was within psychometric standards. Third, results of this study may not be generalizable to all other female college athletes because we included female college athletes from only 1 major university. A national study using female college athletes from many representative universities in combination with the EDE interview is recommended to further verify use of the AMDQ subsets with other universities nationwide.

Future research with the AMDQ subsets that met statistical criteria in this study is recommended for 2 primary reasons. First, because this study and others[12,13] have shown that published eating-disorder tests (ie, EDI-2, BULIT-R, EAT) are not accurate or suitable for comprehensive screening of athletes, the magnitude of the problem of ED/DE in female college athletes nationwide needs to be verified. Prior prevalence studies indicate that athletes are 2 to 3 times more likely than nonathletes (eg, general population and college students) to manifest characteristics of eating disorders.[2] These data also suggest that approximately 3.0% of athletes meet the medical criteria for anorexia and 21.5% meet the criteria for bulimia.[2] Based on a figure of 352 000 athletes in American colleges and universities,[39,40] this translates to an estimated projection of approximately 10 560 athletes nationally who exhibit symptoms of anorexia and 75 680 who exhibit symptoms of bulimia. These estimates are conservative because a less stringent category, NOS, was not considered, nor was disordered eating.

Second, early detection of eating problems in athletes using an athlete-specific test, such as the AMDQ, is a high priority because of medical problems that accompany eating disorders. A recent survey (W. Wooten, unpublished communication, 1990) indicated that eating disorders are the third leading cause of morbidity among college students, preceded by depression and substance abuse. Potential adverse medical complications from eating disorders include amenorrhea, bradycardia, dental erosion, dehydration, electrolyte abnormalities, hypotension, hypothermia, and swollen salivary glands.[7] Another significant problem is unrecoverable bone loss (osteoporosis) associated with being underweight and the cessation of menses.[41–44] Anorexic athletes have the bone density of senescent women 3 to 4 times older, and there are no current methods to restore bone loss.[41–44] Other potential problems of special concern to athletes include diminished muscle power and endurance due to reduced protein synthesis and inadequate glycogen and fluid stores. Dehydration, Benardot et al[45] contended, also is commonly found in anorexic and bulimic individuals due to restricted food intake or the use of self-induced vomiting, laxatives, or diuretics, or a combination of these practices. Attempting to train in a chronically dehydrated state will not only decrease performance but may lead to acute complications, such as heat exhaustion or heat stroke. The personal effects of these medical signs and symptoms of eating disorders may be underestimated by athletes.

Unsolicited comments from our subjects indicated that athletes felt pressure from coaches to lose weight, often with little or no guidance regarding how to do so (or how to lose weight in a healthy, safe way). Interviews suggested that subjects also were misinformed about such topics as weight management, the role of food and body weight in sport performance, and basic nutrition. The athletes were eager to learn more about nutrition and, as such, are an excellent target population for nutrition education and for nutrition studies focusing on performance, total caloric intake, and energy expenditure. Athletes in 1 study[45] believed that diet makes little difference, disordered-eating practices are harmless, and losing weight, regardless of method, enhances performance. Athletes need education about the impact of inadequate caloric intake and disordered-eating practices on athletic performance and health (eg, problems resulting from depletion of muscle glycogen, dehydration, loss of muscle mass, hypoglycemia, electrolyte disturbances, anemia, osteoporosis, and amenorrhea).[7,41–45]

Seven specific recommendations for athletic trainers from the research literature[14–17,46,47] include the following. First, objective goals should be set with the athlete to determine an optimal range for individual body fat, as opposed to setting weight goals based on appearance, standards, or tables that do

not account for individual differences (eg, height/weight tables), or body weight, which does not account for muscle mass. Second, the use of rapid weight-loss methods should be discouraged, and weight-loss programs, when indicated, should be initiated well before the season begins, so that gradual loss of body fat (as opposed to the loss of muscle mass and body water that occurs with crash diets) can occur. Third, nutritional guidance should be provided. Athletic trainers and coaches should not tell athletes to lose weight without also providing sound nutritional guidance on how to do so. Referral of athletes to a registered dietitian for such guidance is recommended. Fourth, dietary intervention should focus on providing adequate calories to support basic needs and the demands of the sport. Nutrient-dense foods, such as complex carbohydrates, should provide most of the calories (55% to 70%), and protein sources should be high in biologic value (ie, animal as opposed to vegetable protein). Athletes in general, but especially those trying to reduce weight, must be encouraged to regularly replace fluids (2 cups of fluid for every 0.45 kg [1 lb] lost during exercise). Female athletes need to be educated about healthy food choices that provide adequate intake of iron and calcium. Fifth, weigh-ins and measurement of body composition should be private to reduce the stress, anxiety, and embarrassment of public assessment. Athletic trainers, coaches, and parents must be aware that their comments and opinions regarding body weight can strongly influence, even trigger, the development of an eating problem in some athletes. Sixth, athletic trainers need to be familiar with ED/DE symptoms and should talk to any athlete who exhibits a problem. Athletes should not be punished or dismissed from the team because of eating problems or the existence of an eating disorder, but professional counseling should be recommended. Finally, the AMDQ subsets (available from the authors) that met statistical criteria in this study could be used to screen for ED/DE, so that the emphasis is on prevention. By identifying the potential problems at earlier, less severe stages, (ie, NOS or disordered eating), we may be able to prevent the progression to an eating disorder (anorexia or bulimia). In fact, the AMDQ could be used as part of a total process to identify potential athletes with ED/DE. The initial step, of course, would be to administer the AMDQ. The next step would be either to retest (with the same test) those athletes who were positive for an ED/DE or to use an alternate form of the screening test for evaluation. The last step would be to have a trained professional verify those who tested positive. Athletic trainers are a pivotal and integral part of the screening, referral, and rehabilitation process.

In summary, an important step has been taken to develop a screening test for at-risk female athletes at the collegiate level. Clearly, a test is needed because ED/DEs are pandemic worldwide. Early detection (eg, identification of athletes at the disordered eating or NOS stage) is a salient priority so that appropriate prevention initiatives can be introduced, athletes can be returned to healthy competition, and ED/DE prevalence can be reduced.

ACKNOWLEDGMENTS

We thank the National Athletic Trainers' Association for support and endorsement of this project. This manuscript is the result of research funded by 2 National Athletic Trainers' Association Research and Education Foundation research grants (project number 0395-B007) awarded to David R. Black and Larry J. Leverenz for the periods July 1, 1994 through June 30, 1995 and July 1, 1995 through June 30, 1996. Both grants were for a study titled "Multicenter study to evaluate a screening test for athletes with disordered eating." Deborah (Chalmers) Nagel, PhD, RD, EdS, was one of the research assistants for both grants. We thank Brenda Stevens, School of Pharmacy; the late Dr. Robert K. Chalmers, School of Pharmacy; Dr. Stephen Badylak, Head Team Physician; and Denny Miller, Head Athletic Trainer, all from Purdue University, for their help and support of this project. We also thank Laurie Larkin, James Larkin, and Lisa (Lauer) Chrysler for their help with data collection and subject recruitment.

REFERENCES

1. Otis CL, Drinkwater B, Johnson M, Loucks A, Wilmore J. American College of Sports Medicine position stand: the female athlete triad. *Med Sci Sports Exerc.* 1997;29:i-ix.
2. Burckes-Miller ME, Black DR. Male and female college athletes: prevalence of anorexia nervosa and bulimia nervosa. *Athl Train, J Natl Athl Train Assoc.* 1988;23:137–140.
3. Black DR, Burckes-Miller ME. Male and female college athletes: use of anorexia nervosa and bulimia nervosa weight loss methods. *Res Q Exerc Sport.* 1988;59:252–256.
4. Burckes-Miller ME, Black DR. Behaviors and attitudes associated with eating disorders: perceptions of college athletes about food and weight. *Health Educ Res Theory Pract.* 1988;3:203–208.
5. Rosen LW, McKeag DB, Hough DO, Curley V. Pathogenic weight control behavior in female athletes. *Physician Sportsmed.* 1986;14(1):79–86.
6. Rosen LW, Hough DO. Pathogenic weight-control behaviors of female college gymnasts. *Physician Sportsmed.* 1988;16(9):141–144.
7. Comerci GD. Medical complications of anorexia nervosa and bulimia nervosa. *Med Clin North Am.* 1990;74:1293–1310.
8. Garner DM [catalog]. Lutz, FL: Psychological Assessment Resources, Inc; 1994.
9. Keyser DJ, Sweetland RC, eds. *Test Critiques.* Vol. VI. Kansas City, MO: Wesport Publishers; 1987;177–181.
10. Brelsford TN, Hummel RM, Barrios BA. The Bulimia Test-Revised: a psychometric investigation. *Psychol Assess.* 1992;4:399–401.
11. Thelen MH, Farmer J, Wonderlich S, Smith M. A revision of the Bulimia Test: the BULIT-R. *Psychol Assess.* 1991;3:119–124.
12. Wilmore JH. Eating and weight disorders in the female athlete. *Int J Sport Nutr.* 1991;1:104–117.
13. O'Connor PJ, Lewis RD, Kirchner EM. Eating disorder symptoms in female college gymnasts. *Med Sci Sports Exerc.* 1995;27:550–555.
14. Black DR, ed. *Eating Disorders Among Athletes: Theory, Issues, and Research.* Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance; 1991.
15. Brownell KD, Rodin J, Wilmore JH, eds. *Eating, Body Weight and Performance in Athletes: Disorders of Modern Society.* Philadelphia, PA: Lea & Febiger; 1992.
16. Holliman SC, ed. *Eating Disorders and Athletes: A Handbook for Coaches.* Dubuque, IA: Kendall/Hunt; 1991.
17. Thompson RA, Sherman RT. *Helping Athletes with Eating Disorders.* Bloomington, IN: Human Kinetics; 1993.
18. *Diagnostic and Statistical Manual of Mental Disorders.* 3rd ed. Washington, DC: American Psychiatric Association; 1987.
19. *Diagnostic and Statistical Manual of Mental Disorders.* 4th ed. Washington, DC: American Psychiatric Association; 1994.
20. Gunning R. *The Technique of Clear Writing.* New York, NY: McGraw-Hill; 1968.
21. Fairburn CG, Wilson GT, eds. *Binge Eating: Nature, Assessment and Treatment.* New York, NY: Guilford Press; 1993.
22. Sundgot-Borgen J. Risk and trigger factors for the development of eating disorders in female elite athletes. *Med Sci Sports Exerc.* 1994;26:414–419.
23. Sundgot-Borgen J. Eating disorders in female athletes. *Sports Med.* 1994;17:176–188.
24. Jackson AS, Pollock ML. Steps toward the development of generalized

equations for predicting body composition of adults. *Can J Appl Sport Sci.* 1982;7:189–196.
25. Cronbach LJ. *Essentials of Psychological Testing.* 3rd ed. New York, NY: Harper and Row; 1970.
26. Nunnally JC. *Psychometric Theory.* 2nd ed. New York, NY: McGraw-Hill; 1978.
27. Browner WS, Newman TB, Cummings SR. In: Hulley SB, Cummings SR, eds. *Designing Clinical Research: An Epidemiological Approach.* Baltimore, MD: Williams & Wilkins; 1988:87–97.
28. Friis RH, Sellers TA. *Epidemiology for Public Health Practice.* Gaithersburg, MD: Aspen; 1996.
29. Last J, ed. *A Dictionary of Epidemiology.* 2nd ed. New York, NY: Oxford University Press; 1988.
30. Halmi KA, Falk JR, Schwartz E. Binge-eating and vomiting: a survey of a college population. *Psychol Med.* 1981;11:697–706.
31. Katzman M, Wolchik S, Braver S. The prevalence of frequent binge eating and bulimia in a non-clinical sample. *Intl J Eat Disord.* 1984;3:53–61.
32. Pyle RL, Mitchell JE, Eckert ED, Halvorson PA, Neuman PA, Goff GM. The incidence of bulimia in freshman college students. *Intl J Eat Disord.* 1985;2:73–84.
33. Conoley JC, Kramer JJ, eds. *The Tenth Mental Measurements Yearbook.* Lincoln, NE: University of Nebraska Press; 1989;274–276.
34. Fairburn CG, Beglin SJ. Studies of the epidemiology of bulimia nervosa. *Am J Psychiatry.* 1990;147:401–408.
35. Lakin JA, Steen SN, Oppliger RA. Eating behaviors, weight loss methods, and nutrition practices among high school wrestlers. *J Community Health Nurs.* 1990;7:223–234.
36. Lindeman AK. Self-esteem: its application to eating disorders and athletes. *Int J Sport Nutr.* 1994;4:237–252.
37. Sundgot-Borgen J. Eating disorders, energy intake, training volume, and menstrual function in high-level modern rhythmic gymnasts. *Int J Sport Nutr.* 1996;6:100–109.
38. Slavin JL. Eating disorders in athletes. *J Phys Educ Recr Dance.* 1987;58:33–36.
39. US Department of Commerce. *Statistical Abstract of the United States.* 110th ed. Washington, DC: US Government Printing Office; 1990.
40. US House Representatives Committee on Education and Liaison Subcommittee on Post Secondary Education. *Hearing on the Role of Athletics and College Life.* Washington, DC: US Government Printing Office; May 18 and 24, 1989.
41. Drinkwater BL, Nilson K, Chesnut CH 3rd, Bremner WJ, Shainholtz S, Southworth MB. Bone mineral content of amenorrheic and eumenorrheic athletes. *N Engl J Med.* 1984;277–281.
42. Johnson CL. Managing problems of weight: anorexia nervosa/bulimia. Presented at: 65th Annual American Dietetic Association Convention; October 1982; San Antonio, TX.
43. Nattiv A, Agostini R, Drinkwater B, Yeager KK. The female athlete triad: the inter-relatedness of disordered eating, amenorrhea, and osteoporosis. *Clin Sports Med.* 1994;13:405–418.
44. Yeager KK, Agostini R, Nattiv A, Drinkwater B. The female athlete triad: disordered eating, amenorrhea, osteoporosis. *Med Sci Sports Exerc.* 1993;775–777.
45. Benardot D, Engelbert-Fenton K, Freeman K, Hartsough C, Nelson-Steen S. Eating disorders in athletes: the dietitian's perspective. *Sports Sci Exch.* 1994;5:1–4.
46. Berning JR, Steen SN. *Sports Nutrition for the 90s: The Health Professional's Handbook.* Gaithersburg, MD: Aspen; 1991.
48. Clark N. *Sports Nutrition Guidebook: Eating to Fuel Your Active Lifestyle.* Champaign, IL: Leisure Press; 1990.