Published in final edited form as:

Psychol Sport Exerc. 2012 November ; 13(6): 779–788. doi:10.1016/j.psychsport.2012.06.002.

Psychometric Evaluation of the Timeline Followback for Exercise among College Students

Gregory A. Panza^{a,*}, Jeremiah Weinstock^{b,c}, Garrett I. Ash^a, and Linda S. Pescatello^a

Gregory A. Panza: GPan222@gmail.com; Jeremiah Weinstock: Jweinsto@slu.edu; Garrett I. Ash: Garrett.Ash@huskymail.uconn.edu; Linda S. Pescatello: Linda.Pescatello@uconn.edu

^aCenter for Health Intervention and Prevention, Department of Kinesiology, University of Connecticut, 2095 N. Hillside Road, U-1110, Storrs, CT 06269 -1110, USA

^bDepartment of Psychology, Saint Louis University, 3511 Laclede Avenue, Saint Louis, MO 63103-2010, USA

^cCalhoun Cardiology Center - Behavioral Health, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030-2202, USA

Abstract

Objectives—Two separate studies assessed the psychometric properties of a retrospective behavioral measure adapted for exercise called the Timeline Followback for Exercise (TLFB-E). Study one examined criterion, convergent, and predictive validity. Study two examined test-retest reliability.

Methods—Study one participants ($N = 66$) were college students 20.0 ± 1.4 yr. Validity of frequency, intensity, time, and type (FITT) of exercise as assessed on the TLFB-E was examined using Pearson r correlations with accelerometers, weekly exercise contracts between participants and researchers, question four of the College Alumni Questionnaire, and a health-related physical fitness battery. Study two participants were a different sample ($N = 40$) of college students 18.63 ± 1.0 yr. Pearson r correlations determined reliability of the TLFB-E for exercise frequency, intensity, and time between two interviews separated by one month. Kappa statistic determined reliability of the TLFB-E for type of exercise.

Results—The TLFB-E displayed evidence of criterion validity when compared to accelerometers ($r = .35$ to $.39$) and evidence of convergent validity when compared to weekly exercise contracts ($r = .65$ to $.80$) and question four of the College Alumni Questionnaire ($r = .06$ to $.75$). The TLFB-E displayed evidence of modest to adequate test-retest reliability ($r = .79$ to $.97$) for exercise frequency, intensity, and time and moderate Kappa ($k = .49$) for exercise type.

Conclusions—The TLFB-E produces evidence of reliable and valid scores among college students and improves upon other self-report, retrospective questionnaires by enabling daily collection of exercise FITT over a specified time period.

© 2012 Elsevier Ltd. All rights reserved.

*Correspondence concerning this article should be addressed to Gregory Panza, University of Connecticut, 2095 N. Hillside Road, U-1110, Storrs, CT 06269 -1110, USA. GPan222@gmail.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

physical activity assessment; validity; reliability; college students; health behavior

The US Department of Health and Human Services (Health and Human Services [HHS], 2008), American College of Sports Medicine (ACSM; Garber et al., 2011), and the American Heart Association (AHA; Haskell et al., 2007) have released exercise guidelines for all Americans because of the many health benefits that result from a physically active lifestyle. However, only 46.7% of college students meet the minimum guidelines of participating in moderate intensity exercise for $30 \text{ min} \cdot \text{d}^{-1}$ on $5 \text{ d} \cdot \text{wk}^{-1}$, or vigorous intensity exercise for $20 \text{ min} \cdot \text{d}^{-1}$ on $3 \text{ d} \cdot \text{wk}^{-1}$, or a combination of the two (American College Health Association [ACHA], 2010). Following the college years, exercise participation progressively declines with approximately 22% of adults 25 – 64yr, 15% of those 65 – 74yr, and 6% of those 75yr participating in the recommended amount of exercise (Center for Disease Control [CDC], 2010). The development of reliable and accurate assessment tools for exercise among college students is important in efforts to develop behavioral strategies that address the decline in exercise participation that occurs with aging.

Self-report physical activity (PA) questionnaires that assess exercise are widely used to quantify exercise participation (Craig et al., 2003; Strath, Bassett, & Swartz, 2004); however, many suffer from limitations such as over reporting (Rzewnicki, Auweele, & Bourdeaudhui, 2003) and often use quantity-frequency methods to collect exercise information. Quantity-frequency methods require individuals to report an “average” of pattern and volume (e.g., “I exercised about two days a week in the past two months”) rather than a specific pattern and volume (e.g., “I exercised on Tuesday and Thursday this week”). Variations in health behaviors, such as exercise, commonly occur over time because of injury, changes in motivation, and other factors that affect exercise participation and are not adequately captured by quantity-frequency methods (Sobell & Sobell, 1996; van Poppel, Chinapaw, Mokkink, van Mechelen, & Terwee, 2010). Due to the limitations of quantity-frequency methods, self-report PA questionnaires that assess more specific exercise patterns are needed.

The Timeline Followback (TLFB) is a retrospective self-report tool used in clinical and research settings and is the standard self-report metric for assessing substance use outcomes in clinical trials for alcohol and illicit drug use (Donovan et al., 2012). The TLFB uses a calendar method to retrospectively assess a target behavior daily over a specified time for up to one year through an interview style approach. Many studies have compared quantity-frequency methods to the TLFB for drinking, and have concluded that the TLFB for drinking is superior to quantity-frequency methods in its ability to collect useful data (Lemmens, Tan, & Knibbe, 1992; O'Hare, Bennett & Leduc, 1991; Saunders & Conigrave, 1990; Werch, 1989; 1990) and accurately categorize drinking levels (i.e., heavy drinker, moderate drinker; Flegal, 1990). Concerns exist around the accuracy of quantity-frequency methods for drinking and other health behaviors (Flegal, 1990; Schroder, Carey, & Venable, 2003). The TLFB interview method has also been psychometrically supported to assess a variety of other behaviors including spousal abuse (Fals-Stewart, Birchler, & Kelley, 2003), gambling (Hodgins & Makarchuk, 2003; Weinstock, Whelan, & Meyers, 2004), sexual behaviors (Weinhardt et al., 1998), smoking (Brown et al., 1998), and panic attacks (Nelson & Clum, 2002). However, the TLFB has not been utilized for the assessment of exercise.

We adapted the TLFB to assess planned forms of exercise (TLFB-E). The TLFB-E has several potential advantages over self-reported PA questionnaires that assess exercise

including the ability to: (1) collect daily exercise behavior over a specified time period by obtaining the *frequency*, *intensity*, *time*, and *type* or FITT components of an exercise prescription; (2) provide documentation of these exercise patterns; (3) allow for analysis of exercise behavior data longitudinally; and (4) provide tailored individual feedback about these exercise patterns. Providing individually tailored feedback is an empirically supported behavior change intervention (Allen & Litten, 1993). Overall, these features of the TLFB-E allow for collection of more specific and useful information for both clinical and research applications than self-report questionnaires that use quantity-frequency methods, and ultimately a more precise depiction of exercise engagement. Furthermore, there are no monetary costs or fees to use the TLFB-E. It is in the public domain. The exercise adaptation of the TLFB has not previously been empirically validated as an assessment tool.

Thus, the purposes of this investigation were to conduct two separate studies to assess the psychometric properties of the TLFB-E among college students. Study one tested criterion, predictive, and convergent validity of the TLFB-E. Study two assessed test-retest reliability of the TLFB-E between two interviews separated by one month. In study one we hypothesized the TLFB-E would display evidence of criterion, convergent, and predictive validity through correlations with other measures of exercise. In study two we hypothesized the TLFB-E would display modest test-retest reliability for *frequency*, *intensity*, and *time*, and moderate Kappa statistic for *type* of exercise self-reported on the TLFB-E at interview one and two.

Study 1: Validity

Method

Data for this study were derived from the National Institute of Health funded project entitled, *Motivational Interventions for Exercise in Hazardous Drinking College Students* (R21-AA017717). The study investigated the utility of exercise as an intervention for sedentary hazardous drinking college students.

Participants—Participants ($N = 66$, $n = 37$ women, $n = 29$ men) were English speaking, currently enrolled in college, 20.0 ± 1.4 yr, and normal weight [body mass index = 24.5 ± 3.3 kg/m²]. Participant classification by ethnic category was 91.6% Caucasian, 4.2% African American, and 4.2% Asian, and was consistent with the local university demographics. Criteria for eligibility included: (a) sedentary, defined as <16 bouts of exercise in the past two months; (b) hazardous drinking as assessed by the Alcohol Use Disorder Identification Test (Saunders, Aasland, Amundsen, & Grant, 2006); (c) reporting at least four heavy drinking episodes in the past two months (Women four drinks, Men five drinks); (d) enrolled in > six course credits; and (e) between 18–26yr. Participants were excluded if they were currently receiving treatment for alcohol use or desired such treatment, had an acute psychiatric problem that may require immediate treatment, or reported any contraindications for exercise on the Service Utilization Form (McLellan, Alterman, Cacciola, Metzger, & O'Brien, 1992) and/or Physical Activity Readiness-Questionnaire (Thompson, Gordon, & Pescatello, 2009). All participants signed an informed consent approved by the local university Institutional Review Board.

Study Overview—Participants were enrolled in an exercise intervention for two months and followed for an additional four months (i.e., six months total) with assessments completed at baseline, two months (post-treatment), and six months (follow up). At all three assessments, participants completed the TLFB-E via a recall interview, completed question four of the College Alumni Questionnaire (Kriska & Casperson, 1997), and a health-related physical fitness assessment battery. Participants wore an accelerometer for four days at baseline and the two month assessment. In addition, as part of an exercise intervention,

weekly exercise contracts between participants and researchers were completed from baseline until the two month assessment.

Subjective Physical Activity/Exercise Measures

Demographic Questionnaire: Participants were asked to complete a demographic questionnaire at baseline only. Information obtained included: age, gender, ethnicity, marital status, grade point average (GPA), and year in school.

Timeline Followback for Exercise: The TLFB-E was completed via paper and pencil through interviews conducted by trained research assistants and covered the past two months of exercise behavior. Training included reviewing the *TLFB User's Guide* (Sobell & Sobell, 1996) and administering pilot interviews under the supervision of a clinical psychologist (JW) experienced in using the TLFB.

The TLFB-E represented a traditional monthly calendar and assessed the FITT components of exercise. *Frequency* of exercise was the number of bouts recorded. *Intensity* or the level of physical exertion was assessed with two methods: (1) the Rating of Perceived Exertion Borg Scale (RPE-Scale; Borg, 1998), and (2) metabolic equivalents (METs) for each exercise bout obtained from the compendium of PA (Ainsworth et al., 2000). *Time* was expressed as minutes per bout. Exercise *type* was reported as the modality and categorized as aerobic, resistance, flexibility, or a combination of modalities.

Additional sections on the TLFB-E included: “special day” and “notes.” The “special day” section was used as a memory aid to enhance recall by recording events that were unique to participants such as birthdays, vacations, hospitalizations, and other. Such events served as anchor points for recall, and therefore aided in remembering exercise behavior. The “notes” section was utilized for recording any important information acquired by the research assistant during the interview or to clarify any data recorded if clarification was needed. On average, the TLFB-E took approximately 20 minutes to assess two months of exercise behavior.

Question four of the College Alumni Questionnaire: The College Alumni Questionnaire has demonstrated good psychometric properties as a measure of PA behavior among college students (Ainsworth, Leon, Jacobs, & Paffenbarger, 1993; Strath et al., 2004). Only data from question four of the College Alumni Questionnaire was used because this question collects information on exercise while the remaining questions gather information on PA. Research assistants administered question four of the College Alumni Questionnaire to ensure its understanding. Exercise variables from question four of the College Alumni Questionnaire were: Total bouts of exercise (*frequency*), average MET hours, total Kcal (*intensity*), total minutes (*time*), and total aerobic bouts, resistance bouts, and flexibility bouts (*type*).

Objective Physical Activity/Exercise Measures

Actical® Accelerometer: An omnidirectional Actical® accelerometer (Mini Mitter Co Inc, Bend, OR), an objective measure of PA, was attached to the participant's hip continuously for four days including two week and two weekend days at baseline and two months. Exercise variables collected from the accelerometers were: Total aerobic exercise bouts (*frequency*), total minutes of aerobic exercise (*time*), and estimated energy expended in Kcal (*intensity*). Exercise logs were completed concurrent with the four days the accelerometers were worn. Aerobic exercise data were calculated as follows: Moderate intensity rating of 3 METs for 20 minutes was equal to one aerobic exercise bout, the sum of Kcal/min/kg x body weight (kg) equaled total Kcal expended for that exercise bout, and the sum of total

minutes of moderate to vigorous intensity exercise (3 METs) expressed as total time over the four days.

Exercise contract: Participants met with the study personnel over eight consecutive weeks beginning at baseline to review the prior week's contract and create new exercise contracts for the upcoming week. Each exercise contract outlined specific exercise activities to be completed (e.g., run 3.0 miles, attend spin class, and swim laps for 20 minutes). Participants were required to provide objective verification of the exercise completed. Examples of objective verification were a fitness instructor's note verifying exercise class attendance, pedometers, and short videos of the participant beginning and completing the exercise activity (i.e., "cell phone videos"). Participants were asked to select three exercise activities and one alternate exercise activity to complete each week. Exercise variables from the exercise contracts were: Total bouts (*frequency*), total aerobic bouts, total resistance bouts, total flexibility bouts, total bouts of aerobic and resistance exercise (*type*), total minutes of exercise (*time*), average rating of perceived exertion (Borg, 1998), average MET hours of all bouts, and total Kcal expended (*intensity*).

Health-Related Fitness Assessments—All health-related fitness assessments for a given subject were administered by the same research assistant. All research assistants were trained by the exercise physiologist study investigator (LP). Fitness assessments were administered in the following order: Resting heart rate (RHR), resting blood pressure (BP), body mass index (BMI), waist circumference (WC), push-up test, handgrip dynamometer, sit-and-reach, and YMCA submaximal bicycle ergometer test (Thompson et al., 2009).

Resting Heart Rate: RHR was used as a measure of cardiorespiratory fitness (Thompson et al., 2009). RHR was obtained prior to all other fitness assessments using a Polar T31-Coded Heart Rate Monitor (Polar Electro Oy, Kempele, Finland) and Polar Heart Rate Watch model F6 ceo537 (Polar Electro Oy, Kempele, Finland). Participants were seated comfortably for a minimum of 15 minutes before RHR in beats per minute was recorded.

Resting Blood Pressure: BP was used as a measure of cardiovascular health (Thompson et al., 2009). Participants were seated quietly for at least 10 minutes in a chair with their back supported, feet on the floor, legs uncrossed, bladder empty, and upper arm supported at heart level (Pickering et al., 2005). Participants were asked to refrain from exercise, smoking cigarettes or ingesting caffeine the day of the measurement. BP was measured in the left arm using an Omron HEM711 automatic deluxe BP monitor (Omron Healthcare, Inc., Bannockburn, IL, 60015) three times with one minute intervals between measurements. If the readings were within 5 mmHg, the readings were averaged and recorded as resting systolic and diastolic blood pressure. If there was a difference of > 5 mmHg between readings, the measurements were repeated until three readings were within 5 mmHg.

Body Mass Index: BMI was used as an indicator of overall adiposity (Thompson et al., 2009). Height and weight were measured using a calibrated Detecto® Scale (Webb City, MO 64870) and used to calculate BMI (kg/m^2 ; Thompson et al., 2009).

Waist Circumference: WC was used as a measure of abdominal adiposity and overall cardiometabolic health (Thompson et al., 2009). WC was measured below the rib cage, 1 inch (2.54 cm) above the umbilicus or at the smallest circumference to the nearest 0.2 inches (0.5 cm). Multiple measures were taken until two measures were within ¼ inch (0.64 cm; Thompson et al., 2009).

Push-Up Test: The push-up test was used to assess arm and shoulder girdle muscle strength and endurance (Mozumdar, Liguori, & Baumgartner, 2010). Men assumed a standard push-up position while women used a modified position with knees on the mat. Participants performed as many consecutive push-ups as possible without resting until s/he either could not continue or could not maintain the appropriate form for two consecutive repetitions (Thompson et al., 2009). Number of push-ups until failure was recorded.

Handgrip Dynamometer: The handgrip test measured overall muscular strength (Thompson et al., 2009; Hamilton, McDonald, & Chenier, 1992). Handgrip strength was assessed using a Jamar® Hydraulic Handgrip Dynamometer model 5030J1 (Sammons Preston Rolyan, Bolingbrook, IL). Two trials for each hand were conducted. Data from the dominant hand was analyzed and recorded in kg.

Sit and Reach: The sit and reach test was used as a measure of flexibility, primarily of the lower back and hip-joint (Chung & Yuen, 1999). Three trials were completed, using the farthest reach of the three trials as the number recorded to the nearest 0.10 cm.

YMCA Submaximal Ergometer Test (YSET): Cardiorespiratory physical fitness was measured using the YSET multistage cycle ergometer protocol (Thompson et al., 2009). HR and workrates were used to predict cardiorespiratory maximal capacity using the YMCA plotting technique (Thompson et al., 2009). Maximum oxygen uptake ($\text{VO}_{2\text{max}}$) was estimated and expressed in $\text{L}\cdot\text{min}^{-1}$.

Statistical Analyses—Descriptive statistics for participants were analyzed using one way Analysis of Variance to determine if there were differences between genders. Correlations were calculated using Pearson product-moment correlation coefficients with $p < .05$ established as the level of significance for the TLFB-E compared to accelerometer (criterion validity, hypothesis one), exercise contract (convergent validity, hypothesis two), question four of the College Alumni Questionnaire (convergent validity, hypothesis three), and health-related fitness assessments (predictive validity, hypothesis four). Paired t -tests examined the presence of under and over reporting on the TLFB-E.

Several PA questionnaires that assess exercise have been validated using samples with a wide spectrum of exercise levels (Ainsworth et al., 1993; Craig et al., 2003; Rzewnicki et al., 2003; Strath et al., 2004) including a study consisting of only college students (Dishman & Steinhardt, 1988). As part of the larger study's inclusion/exclusion criteria all participants were sedentary at baseline. To ensure a range of exercise engagement, two month TLFB-E, question four of the College Alumni Questionnaire, and health-related fitness data were randomly selected from one of three time points: baseline, two month, or six month assessments using the 2007 Microsoft Excel randomization tool (Microsoft Co., Redmond, WA). Accelerometer data were only collected at baseline and two months, therefore accelerometer data were randomly selected from these two time points. Absolute sample sizes at each time point were unequal due to fewer assessments completed at six months compared to baseline and two month assessments. However, random selection of data for each participant did ensure equal relative sample sizes at each time point determined by dividing the number of participants selected at a given time point by the total possible number of participants at a given time point multiplied by 100 to calculate the percent of participants sampled at a given time point. All statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) version 14.0 (SPSS Inc., Chicago, IL).

Results

Participant Characteristics: The overall sample was 20.0 ± 1.4 yr, normal weight, and had optimal BP. All physical fitness tests showed evidence of poor to below average physical fitness for individuals of their age except the push-up test in which participants scored good to very good (See Table 1; Thompson et al., 2009). Men had significantly higher systolic BP ($p < .001$), BMI ($p = .004$), and WC ($p < .001$), and scored significantly higher on the push-up ($p = .029$), and handgrip ($p < .001$) fitness tests than women. Men had pre-hypertension and were overweight while women had optimal BP and were of normal weight.

As shown in Table 2, a range of mean exercise levels was found for randomly selected exercise data at differing time points for all bouts, aerobic bouts, resistance bouts, total time, average RPE, and total Kcal. Overall, the total mean of the sample fell below the ACSM guidelines for all variables (Garber et al., 2011). Although the total sample mean fell below the ACSM guidelines (Garber et al., 2011), when looking at the individual data, over half (56%) of the validity sample ($N = 66$) was exercising according to ACSM guidelines (Garber et al., 2011), exceeding the national average of 46.7% of college students (ACHA, 2010).

Criterion Validity, Hypothesis One: Table 3 displays validity coefficients between the TLFB-E and accelerometer. Correlations were significant for all variables ($r = .35$ to $.39$, $p < .01$), displaying evidence of criterion validity supporting hypothesis one.

Convergent Validity, Hypotheses Two & Three: Validity coefficients (Table 3) between the TLFB-E and eight week exercise contract were significant for all variables ($r = .47$ to $.80$, $p < .001$), displaying evidence of convergent validity of the TLFB-E supporting hypothesis two. Validity coefficients (Table 3) between the TLFB-E and question four of the College Alumni Questionnaire had significant correlations for all variables assessed ($r = .49$ to $.75$, $p < .01$) except average MET hours per bout ($r = .06$, $p > .05$), displaying evidence of convergent validity of the TLFB-E supporting hypothesis three.

Predictive Validity, Hypothesis Four: As shown in Table 4, systolic BP negatively correlated with total bouts ($p = .044$) and total aerobic bouts ($p = .001$) reported on the TLFB-E. Diastolic BP negatively correlated with average MET hours per bout ($p = .043$). WC was positively correlated with bouts of exercise on the TLFB-E that included aerobic and resistance ($p = .023$). Handgrip ($p = .009$) had a positive correlation with resistance bouts reported on the TLFB-E. Sit and reach results positively correlated with total bouts of exercise ($p = .034$) and total aerobic bouts ($p = .037$) reported on the TLFB-E. Pearson correlations among RHR, BMI, estimated VO_2max , and the push-up test and the corresponding variables assessed on the TLFB-E were not statistically significant ($p > .05$). Based on these results, hypothesis four was only partially supported by correlations among the TLFB-E and health-related fitness assessments.

Under and over reporting of the TLFB-E: Discrepancy scores indicated slight under reporting of total bouts ($p = .003$) and slight over reporting of Kcal expended over four days ($p = .038$) on the TLFB-E compared to accelerometer (See Table 5). Discrepancy scores for exercise contracts and TLFB-E indicated over reporting of total bouts ($p < .001$), total aerobic bouts ($p < .001$), total time ($p < .001$), and displayed lower total Kcal expended ($p < .001$) on the TLFB-E over two months. Compared to question four of the College Alumni Questionnaire, participants under reported resistance bouts of exercise on the TLFB-E ($p = .019$) and reported greater average MET hours per bout ($p < .01$).

Study 2: Reliability

Method

Participants—A separate sample of participants was recruited from an undergraduate subject pool at the same state university and students received class research credit for completing the reliability study. Prior to participation, all participants signed an informed consent approved by the university Institutional Review Board. Participants ($N = 40$, $n = 28$ women, $n = 12$ men) were English speaking college students 18.6 ± 1.0 yr. Participant breakdown by ethnic category was 72.5% Caucasian, 20.0% Asian, 2.5% African American, 2.5% Hispanic, and 2.5% Other. Participants were excluded if they were not a college student, <18 yr, and/or have previously filled out the TLFB-E.

Study Overview—Participants met with study personnel two times. The first visit consisted of completion of a demographics questionnaire and the TLFB-E. Information obtained on the demographics questionnaire included: age, gender, ethnicity, marital status, GPA, and year in school. The TLFB-E collected information regarding exercise habits for the past two months. Visit two occurred one month later. Participants completed the TLFB-E covering the same two months as in visit one. We hypothesized the TLFB-E would display evidence of modest test-retest reliability (hypothesis five) and a moderate Kappa statistic (hypothesis six).

Statistical Analyses—Pearson r correlations assessed test-retest reliability of the TLFB-E for the following variables: Total bouts (*frequency*), average RPE (*intensity*), and total minutes (*time*) from interviews one and two. Test-retest reliability criteria standards from Nunnally and Bernstein (1994) were used and included poor ($.69$), modest ($.70$), and adequate ($.80$). Test-retest reliability for the categorical variable *type* was calculated using Kappa statistic. Type included: Aerobic, resistance, and flexibility bouts. Reliability analysis using the Kappa statistic was performed to determine consistency among *type* of exercise reported by participants between interview one and interview two conducted one month later (Hsu & Field, 2003). Kappa statistic criteria for *type* were poor ($< .00$), slight ($.00 - .20$), fair ($.21 - .40$), moderate ($.41 - .60$), substantial ($.61 - .80$), and almost perfect ($.81 - 1.00$) (Landis & Koch, 1977). Statistical analyses were performed using SPSS version 14.0 (SPSS Inc., Chicago, IL) with $p < .05$ established as the level of significance. Statistical analysis for Kappa was performed using calculations based on equations presented in Statistical Methods for Rates and Proportions (Fleiss, 1981).

Results

Test-Retest Reliability, Hypotheses 5 and 6: At interview one, participants recorded an average of 22.0 total bouts ($SD = 12.1$, range = 6.0 – 50.0), 1,379.9 minutes ($SD = 1,425.3$, range = 140.0 – 8940.0) of exercise, RPE of 13.7 ($SD = 1.9$, range = 10.3 – 18.3), and expended an average of 1211.0 Kcal ($SD = 1722.8$, range = 192.9 – 11268.7). At the retest interview, participants recorded an average of 20.0 total bouts ($SD = 12.4$, range = 4.0 – 50.0), 1,308.6 minutes ($SD = 1,445.8$, range = 80.0 – 8880.0), RPE of 13.6 ($SD = 1.8$, range = 10.2 – 18.1), and expended an average of 1188.1 Kcal ($SD = 1784.2$, range = 73.5 – 11543.0). Exercise reported on the TLFB-E at interview one significantly correlated with exercise reported at retest interview of the TLFB-E one month later for total bouts ($r = .93$, $p < .001$), total time ($r = .97$, $p < .001$), average RPE ($r = .79$, $p < .001$), and average Kcal expended ($r = .98$, $p < .001$) for all bouts. Thus, the TLFB-E demonstrated evidence of modest to adequate test-retest reliability (Nunnally & Bernstein, 1994) supporting hypothesis five. Kappa = .49 ($p < .05$) for *type* of exercise reported indicated moderate agreement between the two interviews (Landis & Koch, 1977) supporting hypothesis six.

Discussion

The primary purpose of this study was to test validity (study one) and reliability (study two) of the TLFB-E. We sought to test validity by correlating the FITT components of exercise collected on the TLFB-E with the FITT components of exercise collected using objective and subjective measures of exercise. We sought to test reliability by using a test-retest method between two interviews separated by one month. Results suggest that the TLFB-E produces valid test scores when assessing self-reported exercise behavior in a sample of college students who have decreased or are seeking to decrease their sedentary behaviors. Results also suggest the TLFB-E demonstrates reproducibility in a separate sample of college students with no particular exclusion/inclusion criteria for participation in the research study.

The magnitude of the correlations we found between the TLFB-E and accelerometer ($r = 0.35$ to 0.39 , $p < .01$) were slightly lower than those reported by studies assessing criterion validity via objective measures of the TLFB for smoking (Brown et al., 1998), cocaine, and heroine (Ehrman & Robbins, 1994; $r = .51$ to $.97$, $p < .05$). These slightly lower correlations may be partially explained by the typical validity study design of the TLFB for these other behaviors. TLFB validation studies for smoking (Brown et al., 1998), cocaine, and heroine (Ehrman & Robbins, 1994) only correlated the presence or absence of an event with an objective measure by recording “yes” they did the behavior or “no” they did not (*frequency*), and did not correlate the *intensity* of the behavior. One method by which the TLFB-E gathers information on *intensity* is through Kcal expenditure calculated by METs of the *type* of exercise reported. Therefore, accuracy of *intensity* is also dependent on accuracy of *type* of exercise reported, thus making it more difficult to yield higher correlations with objective measures when compared to other studies where *intensity* was not taken into account. However, when looking at other self-report measures of exercise, Sallis and Saelens (2000) found mean r values of criterion validity to be .30. Therefore, the TLFB-E with a mean r of .37 has slightly higher criterion validity scores than these other self-report questionnaires when compared to objective measures of exercise.

Correlations comparing the TLFB-E to weekly exercise contracts and question four of the College Alumni Questionnaire were similar in magnitude to those found in convergent validity studies of the TLFB for gambling (Weinstock, Whelan, & Meyers, 2004) and panic attacks (Nelson & Clum, 2002). Predictive validity correlations of the TLFB-E were slightly lower compared to those Dennis, Funk, Godley, Godley, and Waldron, (2004) found for predictive validity of the TLFB for substance use. The literature suggests health benefits resulting from participating in a regular exercise program do not begin to appear until 12 weeks and more likely at 16 weeks (Thompson et al., 2009). At the two month follow-up, study participants did not yet begin to manifest improvements in the health-related fitness assessments, and they represented 40% of the sample (Table 2). Therefore, the slightly lower predictive validity scores of the TLFB-E may be attributed to the two month health-related fitness assessments. In addition, participants were sedentary at baseline shown by exercise reported on the TLFB-E, however, daily PA levels (e.g., walking to class, carrying groceries, etc.) may have been higher than we were able to quantify on the TLFB-E which assesses exercise but not physical activities of daily living, thus contributing to the low predictive validity scores of the TLFB-E when correlated with health-related fitness outcomes.

Discrepancies were noticed for exercise *frequency*, *intensity* (Kcal), and *type* when the TLFB-E was compared to weekly exercise contracts and question four of the College Alumni Questionnaire. When compared to weekly exercise contracts, an objective measure, the TLFB-E showed significant over reporting for *frequency*, *intensity* (Kcal) and *type* of exercise. However, when compared to question four of the College Alumni Questionnaire, a

subjective self-report measure, significant under reporting was shown on the TLFB-E for *frequency*, *intensity* (Kcal) and *type* of exercise. Over reporting of the TLFB-E compared to weekly exercise contracts may be explained by contracted exercise bouts that we were not able to verify on the exercise contracts as well as participant involvement in non-contracted exercise. This unverified exercise was not counted as exercise on the weekly contracts, thus leading to over reporting on the TLFB-E compared to weekly exercise contracts.

The discrepancy of under reporting on the TLFB-E compared to the College Alumni Questionnaire maybe due to question four of the College Alumni Questionnaire only using quantity-frequency methods to collect exercise performed. Question four of the College Alumni Questionnaire asks participants to list all exercise participated in over the past two months by type of exercise, total times participated in the exercise, and average time per bout of each exercise. By collecting information in this manner, bouts of exercise that include more than one type of exercise are more likely to not be captured by question four of the College Alumni Questionnaire than the TLFB-E. For example, a participant may have reported participating in three bouts of resistance training exercise and three bouts of aerobic exercise equaling six total bouts as recorded on question four of the College Alumni Questionnaire. However, this individual may have combined resistance training and aerobic exercise together in the same exercise sessions resulting in only a total of three exercise bouts recorded on the TLFB-E. Consequently, question four of the College Alumni Questionnaire is displaying six bouts of exercise completed while this individual completed three bouts as shown on the TLFB-E. This example illustrates how the TLFB-E may have underreported exercise performed compared to question four of the College Alumni Questionnaire.

In study two, we sought to test reliability by analyzing FITT components of exercise reported over two months at two separate interviews separated by one month. The TLFB-E demonstrated evidence of modest to adequate test-retest reliability supporting hypothesis five (Nunnally & Bernstein, 1994). Correlations for test-retest reliability of the TLFB-E ($r = .79$ to $.97$, $p < .01$) were similar in magnitude to correlations for reliability of the TLFB for other complex health-related behaviors ($r = .55$ to $.99$, $p < .05$; Brown, Burgess, Sales, Evans, & Miller, 1998; Ehrman & Robbins, 1994; Hodgins & Makarchuck, 2003; Weinhardt et al., 1998). Kappa statistic indicated moderate agreement ($k = .49$; Landis & Koch, 1977) between the two interviews for *type* of exercise and was significantly higher than reported for the TLFB for heroine ($k = .06$) and cocaine ($k = .05$; Ehrman & Robbins, 1994). Therefore, we can conclude that the TLFB-E has the ability to record *type* of exercise to a much greater degree than by chance supporting hypothesis six.

Our validity and reliability test score results support the TLFB-E as a potentially preferred measure compared to current psychometrically supported retrospective self-report PA questionnaires that assess exercise (See Table 6). Van Poppel and colleagues (2010) conducted a meta-analysis on PA questionnaires and concluded the International Physical Activity Questionnaire (IPAQ) is the most widely used and validated self-report PA questionnaire available, and therefore may be considered the “gold standard” (Van Poppel et al., 2010). Table 6 displays results from studies that tested IPAQ validity scores, and when compared to the TLFB-E, validity scores of the IPAQ were lower for most of the studies. Although the TLFB-E may take longer to administer (approximately 20 minutes per two months) than many other self-report PA measures that assess exercise, it has the capability of providing a greater depth of information about exercise compared to many existent self-report PA questionnaires that assess exercise. Most current PA questionnaires that assess exercise lack the ability to provide detailed information about the FITT of exercise over an extended period of time. For example, Godin’s Questionnaire asks, “during a typical seven day period” (Jacobs, Ainsworth, Hartman, & Leon, 1993) while the IPAQ asks, “over the

last seven days” (Hagstromer, Oja, & Sjostrom, 2006). Additionally, several other self-report PA questionnaires that assess exercise use a quantity-frequency method, collecting information on average exercise completed rather than the specific FITT components completed on a day to day basis which may lead to over or under reporting and an overall inaccurate depiction of exercise participation (Rzewnicki et al., 2003). Another limitation noted for many self-report PA measures that assess exercise is activities of less than 10 minutes and activities with a level of physical exertion lower than brisk walking are difficult to capture (Tudor-Locke & Myers, 2001). Overall, the TLFB-E is an easily administered, comprehensible method of collecting the FITT information of past exercise over two months.

Limitations were present in study one. First, the event marker feature available for accelerometers was not used; therefore accelerometer data ≥ 3 METs (moderate to vigorous intensity) for ≥ 20 minutes were coded as a bout of aerobic exercise. Future studies using the accelerometer as an exercise measure should consider using event markers that participants can set to “tag” when exercise is being done which will allow for a more concise and accurate analysis by simply analyzing the “tagged” exercise data. Second, question four of the College Alumni Questionnaire utilizes the quantity-frequency method and did not adequately collect information on *type* of exercise completed. As discussed previously, this may explain the slight under reporting of the TLFB-E when compared to question four of the College Alumni Questionnaire. Third, the College Alumni Questionnaire when used in its entirety is shown to have valid and reliable test scores; however this may not be the case when only question four is used as done in our study. Fourth, as a result of the larger study design, divergent validity was not evaluated. Divergent validity can be examined by comparing the TLFB-E to an assessment that measures different constructs and testing if these two measures display a low correlation. Future studies that are evaluating convergent validity of the TLFB-E should also test divergent validity in order to determine the exact cause of the relationships shown.

Lastly, specificity of the sample population and small sample size offered limitations for the current study. The current study only investigates the psychometric properties of the TLFB-E among a small sample of college students, with the validity study more narrowly targeting hazardous drinking college students. The lifestyle and exercise habits of college students tend to be different than those of other populations (Behrens & Dinger, 2003). Future studies should assess the psychometric properties of the TLFB-E in a larger, more diverse sample with an ultimate goal of enhancing its generalizability. Study two participants are more representative as they did not have to be sedentary at baseline nor were they heavy drinkers. These differences present another limitation. Validity and reliability study samples ideally should have similar sample characteristics and meet the same inclusion/exclusion criteria.

Strengths of the current study include an interdisciplinary research team of experts in the areas of exercise physiology and clinical psychology. In addition, random selection of assessment points to analyze in this study allowed for data consisting of a wide spectrum of exercise levels. Furthermore, several types of validity were used to test the psychometric properties of the TLFB-E. These included criterion that tested the abilities of the TLFB-E to collect the FITT components of exercise, convergent that tested that the FITT components collected on the TLFB-E were in fact related to the FITT components collected on the weekly exercise contracts and question four of the College Alumni Questionnaire, and predictive that tested the ability of the TLFB-E to forecast health-related fitness outcomes.

The TLFB-E’s ability to collect precise data on the FITT components of past exercise patterns allow for documentation, longitudinal analysis, and tailored feedback for individual exercise behaviors. Most existing self-report PA questionnaires that assess exercise (Strath

et al., 2004; Kriska & Caspersen, 1997) use quantity-frequency methods that require individuals to report an “average” of pattern and volume rather than a specific pattern and volume, therefore variations in exercise that happen over time are not adequately assessed (Sobell & Sobell, 1996; van Poppel et al., 2010). Like the psychometrically supported TLFBs for other behaviors, we have shown that the TLFB-E displays reliable and valid test scores as an exercise recall measure and improves upon quantity-frequency methods by enabling collection of the FITT components of an individual’s daily exercise over a specified period of time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

- Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, Leon AS. Compendium of Physical Activities: An update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*. 2000; 32 (Supplement):S498–S516. [PubMed: 10993420]
- Ainsworth BE, Leon AS, Richardson MT, Jacobs DR, Paffenbarger RS Jr. Accuracy of the college alumnus physical activity questionnaire. *Journal of Clinical Epidemiology*. 1993; 46(12):1403–1411. [PubMed: 8263567]
- Allen JP, Litten RZ. Psychometric and laboratory measures to assist in the treatment of alcoholism. *Clinical Psychology Review*. 1993; 13:223–240.
- American College Health Association. American college health association-national college health assessment II, reference group executive summary, fall 2010. 2010. Retrieved August 1, 2011 from http://www.acha-ncha.org/docs/ACHA-NCHA_Reference_Group_ExecutiveSummary_Fall2009.pdf
- Behrens T, Dinger M. A preliminary investigation of college students’ physical activity patterns. *American Journal of Health Studies*. 2003; 18(2/3):169–172.
- Borg, G. Borg’s perceived exertion and pain scales. Champaign, IL: Human Kinetics Publishers; 1998.
- Brown RA, Burgess ES, Sales SD, Whiteley JA, Evans DM, Miller IW. Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors*. 1998; 12(2):101–112.
- Brown WJ, Burton NW, Marshall AL, Miller YD. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. *The Australian and New Zealand Journal of Public Health*. 2008; 32(6):535–541.
- Center for Disease Control. Early release of selected estimates based on data from the 2010 national health interview survey. 2010. Retrieved August 1, 2011, from http://www.acha-ncha.org/docs/ACHA-NCHA-II_ReferenceGroup_ExecutiveSummary_Fall2010.pdf
- Chung PK, Yuen CK. Criterion-related validity of sit-and-reach tests in university men in hong kong. *Perceptual and Motor Skills*. 1999; 88(1):304–316. [PubMed: 10214658]
- Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Sallis JF. International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*. 2003; 35(8):1381. [PubMed: 12900694]
- Dennis ML, Funk R, Godley SH, Godley MD, Waldron H. Cross-validation of the alcohol and cannabis use measures in the Global Appraisal of Individual Needs (GAIN) and Timeline Followback (TLFB; Form 90) among adolescents in substance abuse treatment. *Addiction*. 2004; 99(2):120–128. [PubMed: 15488110]
- Dinger MK, Behrens TK, Han JL. Validity and reliability of the International Physical Activity Questionnaire in college students. *American Journal of Health Education*. 2006; 37(6):337–343.
- Dishman RK, Steindhardt M. Reliability and concurrent validity for a 7-d recall of physical activity in college students. *Medicine & Science in Sports & Exercise*. 1988; 20:14–25. [PubMed: 3343912]
- Donovan DM, Bigelow GE, Brigham GS, Carroll KM, Cohen AJ, Gardin JG, Lindblad R. Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection

- and measurement of drug use end-points in clinical trials. *Addiction*. 2012; 107(4):694–708. [PubMed: 21781202]
- Ehrman RN, Robbins SJ. Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology*. 1994; 62(4):843–850. [PubMed: 7962889]
- Fals-Stewart W, Birchler GR, Kelley ML. The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence*. 2003; 18(3):131–142.
- Flegal KM. Agreement between two dietary methods in the measurement of alcohol consumption. *Journal of Addiction*. 1990; 85:71–79.
- Fleiss, JL. *Statistical methods for rates and proportions*. New York, NY: John Wiley & Sons; 1981.
- Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, Swain DP. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine & Science in Sports & Exercise*. 2011; 43(7):1334–1359. [PubMed: 21694556]
- Hagstromer M, Oja P, Sjostrom M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutrition*. 2006; 9(6):755–762. [PubMed: 16925881]
- Hamilton GF, McDonald C, Chenier TC. Measurement of grip strength: Validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of Orthopedic and Sports Physical Therapy*. 1992; 16(5):215–219.
- Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Bauman A. Physical activity and public health. Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Medicine & Science in Sports & Exercise*. 2007; 39(8):1423–1434. [PubMed: 17762377]
- Health and Human Services. Introducing the 2008 physical activity guidelines for Americans. 2008. Retrieved August 1, 2011, from <http://www.health.gov/paguidelines/guidelines/chapter1.aspx>
- Hodgins DC, Makarchuk K. Trusting problem gamblers: Reliability and validity of self-reported gambling behavior. *Psychology of Addictive Behaviors*. 2003; 17(3):244–248. [PubMed: 14498819]
- Hsu LM, Field R. Interrater agreement measures: Comments on kappa n, cohen's kappa, scott's π , and aickin's α . *Understanding Statistics*. 2003; 2(3):205–219.
- Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Medicine & Science in Sports & Exercise*. 1993; 25(1):81–91. [PubMed: 8423759]
- Kriska A, Caspersen C. A collection of physical activity questionnaires for health-related research. *Medicine & Science in Sports & Exercise*. 1997; 29(6):S203–S205.
- Kurtze N, Rongul V, Hustvedt BE, Flaners WD. Reliability and validity of self-reported physical activity in the Nord-Trøndelag Health Study: HUNT 1. *Scandinavian Journal of Public Health*. 2008; 36(1):52–61. [PubMed: 18426785]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159. [PubMed: 843571]
- Lemmens P, Tan ES, Knibbe RA. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *Journal of Studies on Alcohol*. 1992; 53(5):476. [PubMed: 1405641]
- MacFarlane DJ, Lee CCY, Ho EYK, Chan KL, Chan DTS. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). *Journal of Science & Medicine in Sport*. 2007; 10(1):45–51. [PubMed: 16807105]
- McLellan AT, Alterman AI, Cacciola J, Metzger D, O'Brien CP. A new measure of substance abuse treatment initial studies of the treatment services review. *Journal of Nervous and Mental Disease*. 1992; 180(2):101–110. [PubMed: 1737971]
- Mini Mitter Company, Inc. Actical: Physical activity monitoring system—instruction manual software version 2.0; Oregon, USA.

- Mozumdar A, Liguori G, Baumgartner TA. Additional revised push-up test norms for college students. *Measurement in Physical Education and Exercise Science*. 2010; 14(1):61–66.
- Nelson WA, Clum GA. Assessment of panic frequency: Reliability and validity of a time-line follow-back method. *Journal of Psychopathology and Behavioral Assessment*. 2002; 24(1):47–54.
- Norman A, Bellocco R, Bergstrom A, Wolk A. Validity and reproducibility of self-reported total physical activity: differences by relative weight. *International Journal of Obesity and Related Metabolic Disorders*. 2001; 25(5):682–688. [PubMed: 11360151]
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. New York, NY: McCraw-Hill; 1994.
- O'Hare T, Bennett P, Leduc D. Reliability of self-reports of alcohol use by community clients. *Hospital & Community Psychiatry*. 1991; 42(4):406–408. [PubMed: 2050354]
- Pickering TG, Hall JE, Appel LJ, Falkner BE, Graves J, Hill MN, Roccella EJ. Recommendations for blood pressure measurement in humans and experimental animals: Part 1: Blood pressure measurement in humans: A statement for professionals from the subcommittee of professional and public education of the American Heart Association council on high blood pressure research. *Circulation*. 2005; 111(5):697–716. [PubMed: 15699287]
- Pols MA, Peeters PH, Ocke MC, Bueno-de-Mesquita HB, Slimani N, Kemper HCG, Collett HJA. Relative validity and repeatability of a new questionnaire on physical activity. *Preventive Medicine*. 1997; 26(1):37–43. [PubMed: 9010896]
- Rzewnicki R, Auweele YV, Bourdeaudhuij ID. Addressing overreporting on the international physical activity questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition*. 2003; 6(3):299–305. [PubMed: 12740079]
- Sallis JF, Saelens BE. Assessment of physical activity by self-report: Status, limitations, and future directions. *Research Quarterly for Exercise and Sport*. 2000; 71(2):S1–14. [PubMed: 10925819]
- Saunders JB, Aasland OG, Amundsen A, Grant M. Alcohol consumption and related problems among primary health care patients: WHO collaborative project on early detection of persons with harmful alcohol consumption—I. *Addiction*. 2006; 88(3):349–362. [PubMed: 8461852]
- Saunders JB, Conigrave KM. Early identification of alcohol problems. *Canadian Medical Association Journal*. 1990; 143(10):1060–1069. [PubMed: 2224674]
- Schroder, Kerstin EE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: II. Accuracy of self-reports. *Annals of Behavioral Medicine*. 2003; 26(2):104–123. [PubMed: 14534028]
- Sidney S, Jacobs DR Jr, Haskell WL, Armstrong MA, Dimicco A, Oberman A, Van Hom L. Comparison of two methods of assessing physical activity in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *American Journal of Epidemiology*. 1991; 133(12):1231–1245. [PubMed: 2063831]
- Sobell, LC.; Sobell, MB. *Timeline followback: A calendar method for assessing alcohol and drug use. User's guide*. Addiction Research Foundation. Toronto, Canada: Center for Addiction and Mental Health; 1996.
- Strath SJ, Bassett DR, Swartz AM. Comparison of the college alumnus questionnaire physical activity index with objective monitoring. *Annals of Epidemiology*. 2004; 14(6):409–415. [PubMed: 15246329]
- Thompson, WR.; Gordon, NF.; Pescatello, LS. *ACSM's guidelines for exercise testing and prescription*. 8. Philadelphia, PA: Lippincott, Williams, and Wilkins; 2009.
- Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Medicine*. 2001; 31:91–100. [PubMed: 11227981]
- Vandelandotte C, de Bourdeaudhuij I, Philippaerts R, Sjostrom M, Sallis JF. Reliability and validity of a computerized and Dutch version of the International Physical Activity Questionnaire (IPAQ). *Journal of Physical Activity & Health*. 2005; 2(1):63.
- van Poppel MNM, Chinapaw MJM, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: A systematic review of measurement properties. *Sports Medicine*. 2010; 40(7):565–600. [PubMed: 20545381]
- Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine*. 1998; 20(1):25–30. [PubMed: 9755348]

- Weinstock J, Whelan JP, Meyers AW. Behavioral assessment of gambling: An application of the timeline followback method. *Psychological Assessment*. 2004; 16(1):72–79. [PubMed: 15023094]
- Werch CE. Quantity-Frequency and diary measures of alcohol consumption for elderly drinkers. *International Journal of the Addictions*. 1989; 24:859–865. [PubMed: 2621007]
- Werch CE. Two procedures to reduce response bias in reports of alcohol consumption. *Journal of Studies on Alcohol*. 1990; 51(4):327–330. [PubMed: 2359305]

Highlights

- We examine psychometric properties of an exercise behavioral measure.
- Validity is shown through correlations with other measures of physical activity.
- Reliability is shown using a test-retest method.
- The exercise behavioral measure may improve upon current physical activity measures.

Table 1

Participant Characteristics for the Total Sample and by Gender Presented in Mean (Standard Deviation)

Variable	Total (n = 66)	Men (n = 29)	Women (n = 37)
Age (yr)	20.0 (1.4)	20.1 (1.6)	19.9 (1.3)
Resting Heart Rate (beats per minute)	72.5 (9.6)	72.1 (8.8)	72.8 (10.3)
Systolic Blood Pressure (mmHg)	116.3 (10.2)	122.9 (9.0) ***	111.1 (7.9)
Diastolic Blood Pressure (mmHg)	67.2 (7.2)	66.1 (8.4)	68.0 (6.2)
Body Mass Index (kg/m ²)	24.3 (3.3)	25.6 (3.0) **	23.3 (3.3)
Waist Circumference (cm)	78.0 (10.8)	84.0 (7.2) ***	73.2 (11.0)
Push – Up (repetitions)	26.0 (15.3)	32.0 (16.9) *	20.3 (11.3)
Handgrip (kg)	28.8 (10.3)	37.0 (9.7) ***	22.9 (5.7)
Sit & Reach (cm)	32.2 (10.2)	29.6 (9.1)	34.2 (10.6)
YMCA Submax Ergometer Test (L·min ⁻¹)	36.8 (6.5)	37.8 (6.8)	36.0 (6.1)

Note. Asterisks denote significant differences between genders.

*
p < .05.

**
p < .01.

p < .001.

Table 2
Exercise Reported and Health-Related PA Assessment Outcomes by Total Sample and Individual Time Points

Measure	Variable	Total <i>M</i> (<i>SD</i>) <i>n</i> = 66	Baseline <i>M</i> (<i>SD</i>) <i>n</i> = 29	Two Month <i>M</i> (<i>SD</i>) <i>n</i> = 26	Six Month <i>M</i> (<i>SD</i>) <i>n</i> = 11
TLFB-E (2 months)	All Bouts	17.8 (11.9) ***	8.2 (4.2)	23.7 (7.2)	29.0 (15.6)
	Aerobic Bouts	13.2 (11.0) ***	6.0 (4.2)	15.9 (8.4)	25.5 (15.1)
	Resistance Bouts	3.3 (4.6) ***	0.8 (1.9)	6.1 (5.4)	3.4 (3.8)
	Flexibility Bouts	0.3 (1.5)	0.0 (0.0)	0.7 (2.3)	0.0 (0.0)
	Aerobic & Resistance Bouts	1.0 (2.5)	1.3 (2.5)	1.0 (2.9)	0.1 (.30)
	Total Time (min)	1,090.0 (1,121.0) ***	414 (256.5)	1319.5 (773.9)	2329.7 (1822.0)
	Average RPE	13.8 (1.43) ***	13.0 (1.3)	14.3 (1.4)	14.4 (0.7)
	Average MET hrs/bout	6.3 (5.1)	5.98 (5.4)	5.3 (2.5)	9.3 (7.6)
	Total Kcal	8504.3 (9285.1) ***	3198.1 (2112.7)	9969.6 (6738.3)	9929.9 (8077.2)
	All Bouts	20.5 (17.3) ***	9.9 (12.3)	28.1 (13.8)	30.4 (21.3)
Q4 College Alumni Questionnaire	Aerobic Bouts	14.6 (13.0) ***	7.7 (8.6)	18.2 (10.5)	24.5 (18.5)
	Resistance Bouts	5.6 (8.7) **	2.3 (6.0)	9.6 (10.2)	4.8 (7.3)
	Flexibility Bouts	0.3 (1.6)	0.0 (0.0)	0.4 (2.0)	1.0 (2.5)
	Total Time (min)	878.6 (435.0) ***	598.4 (804.0)	1693.8 (1344.3)	1902.3 (1436.5)
	Average MET hrs/bout	5.3 (4.2)	4.6 (5.0)	5.9 (3.8)	6.0 (2.3)
	Total Kcal	8504.3 (9285.1) **	4541.2 (5895.7)	10972.8 (8744.7)	13118.1 (13677.6)
	Resting Heart Rate (bpm)	72.5 (9.6)	74.5 (9.8)	70.5 (10.3)	71.8 (6.5)
	Systolic BP (mmHg)	116.3 (10.2)	117.6 (10.0)	116.0 (10.0)	113.5 (11.4)
	Diastolic BP (mmHg)	67.2 (7.2)	66.4 (8.6)	67.6 (6.3)	68.4 (5.6)
	Body Mass Index (kg/m ²)	24.3 (3.3)	24.0 (3.3)	24.4 (2.5)	24.6 (4.9)
Health Related Fitness Tests	Waist Circumference (cm)	78.0 (10.8)	77.1 (13.2)	80.0 (6.6)	75.7 (12.1)
	Push-Ups (reps)	26.0 (15.3)	23.9 (10.4)	27.5 (20.3)	30.5 (2.1)
	Handgrip (kg)	28.8 (10.3)	25.9 (9.2)	30.4 (11.0)	33.2 (10.0)
	Sit & Reach (cm)	32.2 (10.2)	29.6 (10.9)	33.3 (8.2)	36.7 (11.5)
	YSET (L·min ⁻¹)	36.8 (6.5)	36.6 (7.2)	37.5 (5.7)	35.7 (6.6)

Measure	Variable	Total <i>M</i> (<i>SD</i>) <i>n</i> = 66	Baseline <i>M</i> (<i>SD</i>) <i>n</i> = 29	Two Month <i>M</i> (<i>SD</i>) <i>n</i> = 26	Six Month <i>M</i> (<i>SD</i>) <i>n</i> = 11
		<i>n</i> = 32			
		<i>n</i> = 34			
TLFB-E (4 days)	All Bouts	0.8 (1.0) *	0.5 (0.8)	1.1 (1.1)	—
	Total Time (min)	50.2 (68.0) *	31.7 (50.6)	66.5 (77.4)	—
	Total Kcal	389.8 (520.3) *	243.2 (416.8)	519.0 (572.1)	—
Accelerometer (4 days)	All Bouts	1.3 (1.1)	1.0 (1.1)	1.5 (1.0)	—
	Total Time (min)	57.5 (56.1)	53.1 (63.7)	61.3 (49.1)	—
	Total Kcal	262.3 (261.6)	252.1 (298.9)	271.3 (227.9)	—

Note. Accelerometer and four day TLFB-E *n* sizes are different than other measures because accelerometer data was not collected at the 6 month assessment. YSET = YMCA submax cycle ergometer test; TLFB-E = Timeline Followback for Exercise; RPE = rating of perceived exertion (6–20); Kcal = Kilocalories; Q4 = Question 4. Asterisks represent differences among all time points.

* $p < .05$.

** $p < .01$.

*** $p < .001$

Table 3
Criterion and Convergent Validity Coefficients of the Timeline Followback for Exercise

Variable	Accelerometer (four days)		Exercise Contract		Question four of the CAQ	
	r	p	r	p	r	p
All Bouts	.35	.004	.78	<.001	.74	<.001
Aerobic Bouts	—	-	.71	<.001	.75	<.001
Resistance Bouts	—	-	.79	<.001	.49	<.001
Flexibility Bouts	—	-	.80	<.001	.69	<.001
Aerobic & Resistance Bouts	—	-	.71	<.001	—	-
Total Time (min)	.37	.003	.66	<.001	.72	<.001
Average RPE	—	-	.65	<.001	—	-
Average MET hrs/bout	—	-	.47	<.001	.06	.621
Total Kcal	.39	.001	.67	<.001	.61	<.001

Note. RPE = rating of perceived exertion (6–20 scale); CAQ = College Alumni Questionnaire; Kcal = Kilocalories; MET = metabolic equivalent.

Table 4
Predictive Validity Coefficients of the Timeline Followback for Exercise for Health-Related Fitness Measures

Variable	Resting Heart Rate (bpm)	Systolic Blood Pressure (mmHg)	Diastolic Blood Pressure (mmHg)	Body Mass Index (kg/m ²)	Waist Circ. (cm)	Push – Ups (reps)	Handgrip (kg)	Sit & Reach (cm)	YSET (L·min ⁻¹)
	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
All Bouts	-.12	-.25*	.11	.04	.06	.04	.13	.26*	-.01
Aerobic Bouts	-.23	-.41**	.05	-.03	-.10	.16	-.03	.26*	.02
Resistance Bouts	.19	.18	.05	.07	.23	.24	.32**	.09	-.05
Flexibility Bouts	.07	.21	.09	.06	.09	.06	.05	-.02	.01
Aerobic & Resistance Bouts	.05	.18	.14	.13	.28*	.27	.08	-.04	-.03
Total Time (min)	.05	-.18	-.02	.04	.09	.13	.17	.13	.03
Average RPE	-.10	-.08	.02	.03	.09	-.07	.01	.07	-.03
Average MET hrs/bout	.125	.100	-.25*	.12	.14	.06	.21	-.03	.12
Total Kcal	.04	.13	.07	.01	.12	.02	.102	.08	.05

Note. Negative value = negative relationship; positive value = positive relationship. YSET = YMCA submaximal ergometer test; RPE = rating of perceived exertion (6–20 scale), bpm = beats per minute; MET = metabolic equivalent; Kcal = Kilocalories.

* $p < .05$.

** $p < .01$

Table 5

Over and Under Reporting of the TLFB-E Compared to Accelerometers, Exercise Contracts, and Question Four of the College Alumni Questionnaire

Variable	Accelerometer Discrepancy Score			Exercise Contract Discrepancy Score			Question four of the CAQ Discrepancy Score		
	M (SD)	P		M (SD)	P		M (SD)	P	
All Bouts	− 0.5 (1.2)	.003		4.6 (4.9)	<.001		−2.7 (11.7)		.062
Aerobic Bouts	—	-		3.6 (5.9)	<.001		−1.4 (8.8)		.188
Resistance Bouts	—	-		0.7 (3.8)	.142		−2.3 (7.6)		.019
Flexibility Bouts	—	-		0.2 (1.1)	.077		−0.1 (1.2)		.761
Aerobic & Resistance Bouts	—	-		0.2 (3.7)	.692		—		-
Total Time (min)	−7.3 (70.3)	.408		421.3 (564.2)	<.001		157.2 (912.7)		.166
Average RPE	—	-		0.3 (2.1)	.231		—		-
Average MET hrs/bout	—	-		0.20 (2.3)	.487		1.03 (3.0)		.006
Total Kcal	127.5 (482.2)	.038		2160.3 (4684.1)	<.001		−1516.7 (7377.7)		.100

Note. Negative value = under reporting of the TLFB-E; positive value = over reporting of the TLFB-E. TLFB-E = Timeline Followback for Exercise; CAQ = College Alumni Questionnaire; RPE = rating of perceived exertion (6–20 scale); Kcal = Kilocalories; MET = metabolic equivalent.

Table 6
Validity of the TLFB-E Compared to Selected Examples of Physical Activity Questionnaires that Assess Exercise

Questionnaire	Comparison Measure (Equivalency measure used with TLFB-E correlations)	Study Population (n; mean age or range)	Author(s)	Results	Questionnaire with $> r$
MAAS	Accelerometer	44♂; 55y	Brown, Burton, Marshall, & Miller, 2008	Total $r = .52$	MAAS
Baecke	Accelerometer	78; 20–59y	Jacobs, Ainsworth, Hartman, & Leon, 1993	Total $r = .19$	TLFB-E
EPIC original Q	Activity Diary (Exercise Contract/Log)	59♂; 41 y; 52♀; 49y	Pols et al., 1997	Total $r = .55$	TLFB-E
Godin Q	Accelerometer	64–73; 37y	Jacobs, Ainsworth, Hartman, & Leon, 1993	Total $r = .32$	TLFB-E
HUNT 1	Accelerometer	108♂; 32y	Kurtze, Rongul, Hustvedt, & Flanders, 2008	Total $r = .07$	TLFB-E
College Alumni Q	Accelerometer	64–73; 37y	Jacobs, Ainsworth, Hartman, & Leon, 1993	TEE $r = .30$	TLFB-E
	Accelerometer	28♂; 50♀; 38y	Ainsworth, Leon, Richardson, Jacobs, & Paffenbarger, 1993	Leisure EE $r = .19$	TLFB-E
IPAQ	Accelerometer	22♂; 24♀; 41y	Hagstromer, Oja, & Sjostrom, 2006	Total $r = .55$	IPAQ
	Accelerometer	32♂; 91♀; 21y	Dinger, Behrens, & Han, 2006	Total $r = .23$	TLFB-E
	Accelerometer	108♂; 32 y	Kurtze, Rongul, Hustvedt, & Flanders, 2008	Total $r = .26$	TLFB-E
	Accelerometer	30♂; 26 y; 19♀; 34y	MacFarlane, Lee, Ho, Chan, & Chan, 2007	Total $r = .09$	TLFB-E
Computerized	BMI	22♂; 24♀; 41y	Hagstromer, Oja, & Sjostrom, 2006	Total $r = .25$	IPAQ
IPAQ	PA Diary (Exercise Contract/ Log)	23♂; 30♀; 31y	Vandelaanote, de Bourdeaudhuij, Philippaert, Sjostrom, & Sallis, 2005	$r = .38$ (Total PA)	IPAQ
Norman Q	7 d Activity Diary (Exercise Contract/Log)	111♂; 63y	Norman, Bellico, Bergstrom, & Wolk, 2001	$r = .39$ (min)	TLFB-E
Modified Stanford SDR	Accelerometer	4956; 18–30y	Sidney et al., 1991	Total PA $r = .56$	TLFB-E
				Total $r = .33$	TLFB-E

Note. MAAS = Modified Active Australia Survey; EPIC = European Prospective investigation into Cancer and Nutrition original Questionnaire; HUNT 1 = The Nord-Trondelag Health Study 1; IPAQ = International Physical Activity Questionnaire; SDR = seven day recall; d = day; ♀ = females; ♂ = males; y = years of age; TEE = Total Energy Expenditure; EE = Energy expenditure; Q = Questionnaire; $> r$ = greater correlation (questionnaire vs. TLFB-E).