12-18-2011

Psychometric Evaluation of the Timeline Followback for Exercise among Young Adults

Gregory A. Panza
University of Connecticut - Storrs, Gregory.Panza@uconn.edu

Psychometric Evaluation of the Timeline Followback for Exercise

among Young Adults

Gregory A. Panza

B.S. Central Connecticut State University, 2009

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

At the

University of Connecticut

2011

APPROVAL PAGE

Master of Arts Thesis

Psychometric Evaluation of the Timeline Followback for Exercise

among Young Adults

Presented by

Gregory Panza, B.S.

Major Advisor ____
Linda S. Pescatello, Ph.D, FACSM

Associate Advisor ____
Jeremiah Weinstock, Ph.D

Associate Advisor ____
Lawrence E. Armstrong, Ph.D, FACSM

University of Connecticut

2011

Acknowledgements

Special thanks to my family, fellow students, and advisory committee for your guidance and support over the past two years.

Table of Contents

Chapter 1 - Introduction

Background and Significance

Benefits of Physical Activity

The US Department of Health and Human Services (Health and Human Services [HHS], 2008), American College of Sports Medicine (Garber et al., 2011) and the American Heart Association (Haskell et al., 2007) have released new physical activity (PA) guidelines for all Americans because of the many health benefits that result from a physically active lifestyle. Scientific evidence confirms that regular participation in physical activity (PA) has several health benefits including: Reduction in risk of pre-mature all-cause mortality, prevention and elimination of overweight and obesity by reducing body fat, reduction in risk of developing chronic diseases, reduction or elimination of presence of risk factors for chronic disease, increased aerobic capacity, increased muscular strength and endurance, increased functional capacity, decreased risk of injury, and improved cognitive function. (Macera, Hootman, & Sniezek, 2003). Habitual PA and exercise also improve symptoms of anxiety, mood, and depression (Paluska & Schwenk, 2000; Saxena, Van Ommeren, Tang, & Armstrong, 2005).

Despite the many health benefits of habitual PA, adherence rates are relatively low in the United States. These adherence rates vary greatly among age groups.

Exercise Behavior among College Students

Only 46.7% of college students compared to approximately 61% of high school students meet the minimum guidelines of participating in moderate intensity PA for 30 min•d^{-1} on $\geq$5 d•wk^{-1}, or vigorous intensity PA for 20 min•d^{-1} on $\geq$3 d•wk^{-1} (American College Health Association [ACHA], 2010). These participation rates indicate that exercise levels decline from high school to college (Center for Disease and Control [CDC], 2010; American College Health Association [ACHA], 2010). Bray and Born (2004) found half of the students who had been physically active in high school failed to meet the same exercise standards during their first semester at university. Additionally, a meta-analysis examining college student exercise behaviors concluded that about 40-50% of college students do not meet the ACSM recommendations for adequate amounts of PA and exercise (Keating, Guan, Pinero, & Bridges, 2005). Subsequent to the college years, PA participation displays a progressive declined trend with approximately 22% of adults 25 – 64yr, 15% of those 65 – 74yr, and 6% of those ≥75yr participating in the recommended amount of PA (CDC, 2010). Thus, the declined PA trend indicates importance of intervening before exercise levels continue to decrease after college years.

Kruger, Ham, & Kohl, (2007) examined the colloquial term “weekend warrior” (somebody who compresses their weekly activity into long durations on 1 or 2 days). They discovered the prevalence of this pattern is highest among those aged 45–64yr and is lowest among those aged 18–24yr, suggesting that the college aged population tends to be more active on weekdays than on weekends. Therefore, different PA assessment strategies may be necessary for the college age group than other adult populations (i.e., aged 25-64 and 75yr and older). However, Keating and colleagues (2005) conducted a meta-analysis on college students’ PA behaviors and noted that PA assessments currently used in studies investigating college students’ PA and exercise behaviors lack the ability to collect information on habitual PA and exercise patterns. Thus, Keating et al. (2005) suggested there is a need to standardize exercise measures to help researchers’ better understand students’ habitual PA behaviors.

Physical Activity / Exercise Assessments

The assessment of exercise behavior is an essential part of exercise adherence research. PA assessments serve as a key component in research and clinical settings as they are often used to detect PA trends over time, measure PA levels and health status, determine health benefits, and evaluate the effects of interventions (van Poppel, Mokkina, van Mechelen, & Terwee, 2010). However, there are limited cost-efficient, reliable and valid measures to examine exercise (Prince et al., 2008). Self-report measures such as diaries/logs, surveys, and interviews are frequently used due to their low cost, practicality with a large population, general acceptance, and minimal burden to the participant (Dishman, Washburn, & Schoeller, 2006). Self-report measures have the ability to provide information over a long period of time while other measures such as accelerometers and maximal aerobic capacity tests target shorter periods of exercise measurement (Ehrman & Robbins, 1994).

However, self-reports often display issues of recall and response bias along with their lack of ability to capture absolute levels of exercise (Prince et al., 2008). In addition, self-report assessments often use quantity-frequency methods to collect PA information (Sobell, L.C. & Sobell, 1996). 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 PA 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, L.C. & Sobell, 1996; van Poppel, Chinapaw, Mokkink, van Mechelen, & Terwee, 2010).

The College Alumni Questionnaire, also known as the Paffenbarger PA questionnaire, is a commonly used self-report questionnaire that uses the quantity-frequency method to collect information on past PA behavior (Ainsworth, Leon, Richardson, Jacobs, & Paffenbarger, 1993). However, studies have shown that the College Alumni Questionnaire under reports walking and stair climbing when compared to PA records (Ainsworth et al., 1993) and pedometers (Bassett, Cureton, & Ainsworth, 2000), and is limited in its ability to capture light to moderate PA (Strath, Bassett, & Swartz, 2004).

The International Physical Activity Questionnaire (IPAQ) is another well known and widely used self-report PA questionnaire that applies the quantity-frequency method and presents limitations. The IPAQ has been shown to over report in a population sample of randomly selected Belgian adults whose PA response patterns ranged from very low to very high (Rzewnicki, Auweele, & Bourdeaudhuij, 2003). The IPAQ has also been considered to have only "acceptable" measurement properties when compared to accelerometers (Craig et al., 2003). Fogelholm et al. (2006) compared PA reported on the IPAQ to results of several physical fitness tests. Results showed that sedentary individuals reported participation in very high levels of PA on the IPAQ. Therefore, they concluded there is a further need to solve the over reporting problem by apparently sedentary individuals (Fogelholm et al., 2006).

In contrast, direct measures (e.g., accelerometers, doubly labeled water) of exercise are objective and display a more accurate means of assessing exercise in adults (Prince et al., 2008). However, these strategies require trained professionals and are significantly more expensive than self-report assessments. Having a low-cost, valid, and reliable measure to evaluate PA and exercise behavior would be ideal for researchers and exercise professionals. Table 1 displays advantages and disadvantages of various types of self-report or subjective assessments of PA/exercise. Table 2 displays advantages and disadvantages of various objective assessments of PA/exercise.

Table 1. *Advantages and Disadvantages of Self-Report PA/Exercise Assessments.*

Type of Assessment	**Advantages**	**Disadvantages**
Diaries / Logs	• Practical • Low in cost • Daily information obtained • Specific information obtained • Easy to administer • Provides exercise behavior over a period of time • Provide the possibility to categorize respondents into activity categories	• Requires daily subject cooperation • May present response bias • Inability to capture absolute level of exercise • Typically used over short periods of time
Questionnaires / Surveys	• Practical • Low in Cost • Easy to administer • Low subject burden • Takes minimal time • Provide the possibility to categorize respondents into activity categories	• May present recall and response bias • Inability to capture absolute level of exercise
Recall Interviews	• Practical • Low in Cost • Specific information obtained • Greater reliability and validity than quantity / frequency measures • Low subject burden • Provide memory aids • Provide exercise behavior over a period of time • Interviewer assistance • Provide the possibility to categorize respondents into activity categories	• May present recall and response bias • Inability to capture absolute level of exercise • Interviewer must be trained in appropriate administration • May potentially take longer to complete compared to questionnaires / surveys depending on amount of PA recalled.

Note: EE = Energy Expenditure, HR = Heart Rate, PA = Physical Activity

References: Battley, 1995; Dishman, Washburn, & Schoeller, 2006; Heyward, 2006; Prince et al., 2008; Rush, Valencia, & Plank, 2008; Schoeller & Racette, 1990.

Table 2. *Advantages and Disadvantages of Objective PA/exercise*

Type of Assessment	**Advantages**	**Disadvantages**
Accelerometer	• Provide detailed description of activity patterns • User specific epoch intervals • Excellent data storage capacity • Extensively validated • Applicable to children and adults • Applicable in relatively large studies	• Wearing device for days at a time may be inconvenient • EE from complex movements are not reflected by acceleration of the body (i.e. bicycling, upper body work, walking up/down, carrying goods, etc.) • Data needs careful interpretation • Costly (several hundred dollars) • Require individual calibration

Doubly Labeled Water	• Measures metabolic rate over time (days to weeks) • Accuracy (1-2%) Precision (4-8%) • Use of stable isotopes • Easy use in free-living objects	• Isotopes and instrument may be costly • Trained personnel needed to administer • May be a burden to the subject • Measure of Co2 production rather than O2 consumption • Does not measure intensity or frequency of exercise
Indirect Calorimetry	• Can provide the actual energy cost of a particular activity • Heat loss as well as ratios of fat, carbohydrates catabolized can be calculated • When combined with duration, total energy expended during any appropriate time period may be calculated • Corrections do not have to be made for heat loss by convection, conduction, evaporation, and radiation. • Excellent accuracy and precision (less than 1% error)	• Trained personnel needed to administer • May be a burden to the subject (Enclosed chamber) • Restrict activities and change activity patterns because they are not free-living environments • Considerable investment of time by both the subject and investigator
Direct Calorimetry	• Can make direct measurements of total heat loss or rate of heat loss • Can measure heat loss in various conditions • Energy budgets can be formulated based on data • Excellent accuracy and precision	• Calorimeters are expensive and complex • Trained personnel needed to administer • May be a burden to the subject • Restrict activities and change activity patterns because they are not free-living environments • Considerable investment of time by both the subject and investigator
Peak Oxygen Consumption (V_{O2} Max)	• Used as an estimate of physical activity • More objective than a questionnaire • Beneficial in research looking at conditioning of an individual	• About 30% of variation may be explainable by genetic variation • Measurement is uncomfortable for the subject • Equipment is costly • Trained personnel needed to administer • Does not reflect physical fitness level in terms of muscular strength • Does not measure total daily expenditure
Pedometers	• Low in cost • Practical • Low subject burden • Applicable in large studies	• Poor reliability and validity • Do not necessarily detect activities performed in a static position • Only provide amount of steps taken
Heart Rate Monitoring	• The FLEX HR method is extensively validated in adults, children, athletes, and obese • Applicable in relatively large studies • Provide data on EE and pattern of PA	• Require individual calibration • Other factors other than PA may alter true HR • HR monitor and software may cost several hundred dollars

Note: EE = Energy Expenditure, HR = Heart Rate, PA = Physical Activity

References: Battley, 1995; Dishman, Washburn, & Schoeller, 2006; Heyward, 2006; Prince et al., 2008; Rush, Valencia, & Plank, 2008; Schoeller & Racette, 1990.

Timeline Followback Calendar Method

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., in press). Originally developed in the early 1970s, the TLFB was an alternative to procedures being utilized at the time to classify individuals as "drinkers" or "abstinent." 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. By analyzing behavior daily, the TLFB has the ability to generate a variety of variables by its ability to gather information on pattern, variability, and level that the individual is partaking in the behavior (Sobell, L.C. & Sobell, 1996). The TLFB is superior to quantity-frequency methods in terms of reliability and validity (Sobell, L.C. & Sobell, 1996). The TLFB is 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). Over the years the TLFB has gained international acceptance having been referenced in several publications with studies in the United States (Weinstock et al., 2004), Canada (Hodgins & Makarchuk, 2003), Finland (Aalto, Tuunanen, Sillanaukee, & Seppa, 2006), Sweden (Carlbring, Jonsson, Josephson, Forsberg, 2009), Germany (Collins, Eck, Torchalla, Schroter, Batra, 2009) among others.

Timeline Followback for Exercise

The TLFB has not been utilized for the assessment of PA. We propose to adapt the TLFB to assess PA behavior that includes more structured and planned forms of exercise (TLFB-E). The TLFB-E has several potential advantages over other self-reported PA questionnaires including the ability to: (1) collect daily PA 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 PA behavior data longitudinally; and (4) provide tailored individual feedback about these exercise patterns. 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 PA/exercise engagement.

Purpose of Study

The purposes of this study are to conduct two separate studies to assess the psychometric properties of the TLFB-E among college students. Study one is a validity study examining criterion, predictive, and convergent validity of the TLFB-E. Study two assesses test-retest reliability of the TLFB-E between two interviews separated by one month. Data for study one was derived from the National Institute of Health funded project entitled, *Motivational Interventions for Exercise in Hazardous Drinking College Students* (MILE) (R21-AA017717). MILE investigated the utility of exercise as an intervention for sedentary hazardous drinking college students. Test-retest reliability data was collected from a separate sample of college students.

Specific Aims and Hypotheses

Study 1 - Validity

Specific Aim 1: The first aim was to assess criterion validity of the TLFB-E by examining the relationship between exercise reported on the TLFB-E and data obtained from accelerometers on matching days and time periods of 96 hours (e.g., 2 weekdays / 2 weekend days).

Hypothesis 1: We hypothesized that bouts of exercise recorded on the TLFB-E would correlate with bouts of exercise shown by accelerometers over the four day period. Variables compared included: All bouts (*frequency)*, total Kcal expended (*intensity)*, and total minutes (*time*) of exercise bouts.

Specific Aim 2: The second aim of this study was to measure convergent validity by examining the relationship between the TLFB-E and weekly exercise contracts administered over the same 8 week period.

Hypothesis 2: We hypothesized that there would be a positive correlation between exercise *frequency* (number of bouts), *intensity* (average MET hours, Kcal expended and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, combined aerobic and resistance) of exercise recorded on the TLFB-E and weekly exercise contracts.

Specific Aim 3: The third aim of this study was to measure convergent validity by evaluating the association between the TLFB-E and question four of the College Alumni Questionnaire also known as the Paffenbarger Physical Activity Questionnaire (Bassett Jr. & Ainsworth, 2000).

Hypothesis 3: We hypothesized that there would be a positive correlation between exercise recorded on the TLFB-E and responses on the College Alumni Questionnaire. Variables assessed were: *Frequency* (number of bouts), *intensity* (average MET hours, Kcal expended and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, combined aerobic and resistance) of exercise.

Specific Aim 4: The fourth aim of this study was to measure predictive validity by observing the relationship between exercise recorded on the TLFB-E and results from health-related physical fitness assessments including the YMCA submaximal bicycle ergometer test (YSET) (Thompson, Gordon, & Pescatello, 2009; Poldermans et al., 1993), handgrip dynamometer

(Hamilton, McDonald, & Chenier, 1992; Rantanen et al., 1999), push-up test, sit-and-reach test, resting heart rate (RHR), resting blood pressure (BP), waist circumference, and body mass index (BMI; Thompson et al., 2009).

Hypothesis 4: We hypothesized that there would be either a positive or negative relationship depending on the assessment, between exercise (i.e., FITT) recorded on the TLFB-E and levels of health and fitness measured by health-related PA assessments. Variables correlated from the TLFB-E included: *Frequency* (number of bouts), *intensity* (average MET hours, Kcal expended, and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, or combined bout of aerobic and resistance) of exercise.

Study 2 - Reliability

Specific Aim 5: The fifth aim of this study was to determine test-retest reliability of the TLFB-E.

Hypothesis 5: We hypothesized that the TLFB-E would show modest test-retest reliability ($r \geq .70$) for *frequency* (number of bouts), *intensity* (Kcal expended and rating of perceived exertion), and *time* (total minutes).

Specific Aim 6: The last aim of this study was to determine kappa statistic for *type* of exercise reported at interview one compared to *type* of exercise reported at interview two.

Hypothesis 6: We hypothesized that the TLFB-E would display a moderate kappa statistic (.41 – .60; Landis & Koch, 1977) for *type* of exercise reported at interview one and interview two.

Significance of Study

This study will provide insight on a possible PA assessment tool used to examine retrospective habitual PA/exercise behavior among college students. The TLFB method has shown to be a valid and reliable assessment in substance use and other behaviors (Hodgins & Makarchuk, 2003; Weinstock et al., 2004; Weinhardt et al., 1998; Brown et al., 1998; Nelson &

Clum, 2002). However, its validity and test-retest reliability as an assessment of PA/exercise is unknown. Psychometric evaluation of an assessment tool such as the TLFB-E may provide exercise professionals with a cost-effective instrument that improves upon commonly used quantity-frequency self-report measures by collecting information on past patterns of exercise on a daily basis. By obtaining this knowledge, exercise professionals will enhance their ability to prescribe appropriate interventions and exercise programs for college students based on their patterns and activities. Providing more appropriate programs and interventions for college students may contribute to a decrease in sedentary behavior which may ultimately result in a more active and healthier adulthood.

References

Aalto, M., Tuunanen, M., Sillanaukee, P., & Seppä, K. (2006). Effectiveness of structured questionnaires for screening heavy drinking in middle-aged women. *Alcoholism: Clinical and Experimental Research, 30*(11), 1884-1888.

American College Health Association. (2010). *American college health association-national college health assessment II, reference group executive summary, fall 2010.* Retrieved August 1, 2011 from http://www.achancha.org/docs/ACHA-NCHA_ Reference_ Group_ ExecutiveSummary_Fall2009.pdf.

Ainsworth B.E., Haskell W.L., Herrmann S.D., Meckes N., Bassett Jr. D.R., Tudor-Locke C., Greer J.L., Vezina J, Whitt-Glover M.C., Leon A.S. (2011). Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise*, 43(8):1575-1581.

Bassett Jr., D. R., Cureton, A.L. & Ainsworth, B.E. (2000). Measurement of daily walking distance-questionnaire versus pedometer. *Medicine & Science in Sports & Exercise, 32*(5), 1018-1023.

Battley, E. H. (1995). The advantages and disadvantages of direct and indirect calorimetry. *Thermochimica Acta, 250*(2), 337-352.

Behrens, T., & Dinger, M. (2003). A preliminary investigation of college students' physical activity patterns. *American Journal of Health Studies, 18*(2/3), 169-172.

Brown, R. A., Burgess, E. S., Sales, S. D., Whiteley, J. A., Evans, D. M., & Miller, I. W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors, 12*(2), 101-112.

Carey, K. B. (1997). Reliability and validity of the time-line follow-back interview among psychiatric outpatients: A preliminary report. *Psychology of Addictive Behaviors, 11*, 26-33.

Carlbring, P., Jonsson, J., Josephson, H., & Forsberg, L. (2009). Motivational interviewing versus cognitive behavioral group therapy in the treatment of problem and pathological gambling: A randomized controlled trial. *Cognitive Behaviour Therapy,* 39, 92-103.

Collins, S. E., Eck, S., Torchalla, I., Schröter, M., & Batra, A. (2009). Validity of the timeline followback among treatment-seeking smokers in germany. *Drug and Alcohol Dependence,* 105 (2), 164-167.

Dinger, M. (1999). Physical activity and dietary intake among college students. *American Journal of Health Studies, 15*(3), 139–149.

Dishman, R. K., Washburn, R., & Schoeller, D. (2006). Measurement of physical activity. *Active Living, Cognitive Functioning, and Aging, 65*, 74-91.

Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62*(4), 843-850.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2003). The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence, 18*(3), 131-142.

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the computer science and applications, inc. accelerometer. *Medicine & Science in Sports & Exercise, 30*(5), 777.

Garber, C. E., Bryan Blissmer, Michael R. Deschenes, Barry A. Franklin, Michael J. Lamonte, I-Min, . . . David P. Swain. (2011). 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. 43*(7), 1334-1359.

Hamilton, G. F., McDonald, C., & Chenier, T. C. (1992). Measurement of grip strength: Validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of Orthopedic and Sports Physical Therapy, 16*(5), 215-219.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S.N., Franklin, B. A., Macera, C. A., Heath, G. W., Thompson, P. D., Bauman, A. (2007). 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, 39*(8), 1423–1434.

Heil, D. P. (2006). Predicting activity energy expenditure using the actical activity monitor. *Research Quarterly for Exercise & Sport, 77*(1), 64-80.

Heyward, V. H., (2006). *Advanced fitness assessment and exercise prescription.* Human Kinetics Publishers.

Hodgins, D. C., & Makarchuk, K. (2003). Trusting problem gamblers: Reliability and validity of self-reported gambling behavior. *Psychology of Addictive Behaviors, 17*(3), 244-248.

Keating, X. D., Guan, J., Pinero, J. C., & Bridges, D. M. (2005). A meta-analysis of college students' physical activity behaviors. *Journal of American College Health, 54*(2), 116-126.

Kruger, J., Ham, S. A., & KOHL III, H. W. (2007). Characteristics of a "weekend warrior": Results from two national surveys. *Medicine & Science in Sports & Exercise, 39*(5), 796.

Kwan, M. Y. W., Bray, S. R., & Ginis, K. A. M. (2009). Predicting physical activity of first-year university students: An application of the theory of planned behavior. *Journal of American College Health, 58*(1), 8.

Macera, C. A., Hootman, J. M., & Sniezek, J. E. (2003). Major public health benefits of physical activity. *Arthritis Care & Research, 49*(1), 122-128.

Matthews, C. E., Ainsworth, B. E., Thompson, R. W., & Bassett Jr., D. R. (2002). Sources of variance in daily physical activity levels as measured by an accelerometer. *Medicine & Science in Sports & Exercise, 34*(8), 1376.

Nelson, W. A., & Clum, G. A. (2002). Assessment of panic frequency: Reliability and validity of a time-line follow-back method. *Journal of Psychopathology and Behavioral Assessment, 24*(1), 47-54.

Newman, W. P., Freedman, D., Voors, A., Gard, P., Srinivasan, S., Cresanta, J., . . . Berenson, G. (1986). Relation of serum lipoprotein levels and systolic blood pressure to early atherosclerosis. the bogalusa heart study. *New England Journal of Medicine, 314*(3), 138.

Paluska, S., & Schwenk, T. (2000). Physical activity and mental health: Current concepts. *Sports Medicine, 29*(3), 167-180.

Poldermans, D., Fioretti, P., Forster, T., Thomson, I., Boersma, E., El-Said, E., . . . van Urk, H. (1993). Dobutamine stress echocardiography for assessment of perioperative cardiac risk in patients undergoing major vascular surgery. *Circulation, 87*(5), 1506.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *The International Journal of Behavioral Nutrition and Physical Activity, 5*, 56. doi:10.1186/1479-5868-5-56

Rantanen, T., Guralnik, J. M., Foley, D., Masaki, K., Leveille, S., Curb, J. D., & White, L. (1999). Midlife hand grip strength as a predictor of old age disability. *The Journal of the American Medical Association, 281*(6), 558.

Rush, E. C., Valencia, M. E., & Plank, L. D. (2008). Validation of a 7-day physical activity diary against doubly-labeled water. *Annals of Human Biology, 35*(4), 416-421.

Saxena, S., Van Ommeren, M., Tang, K., & Armstrong, T. (2005). Mental health benefits of physical activity. *Journal of Mental Health, 14*(5), 445-451.

Schoeller, D. A., & Racette, S. B. (1990). A review of field techniques for the assessment of energy expenditure. *Journal of Nutrition, 120*(11), 1492.

Sobell, L. C., & Sobell, M. B. (1996). 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.

Strong, W., Deckelbaum, R., Gidding, S., Kavey, R., Washington, R., Wilmore, J., & Perry, C. (1992). Integrated cardiovascular health promotion in childhood. A statement for health professionals from the subcommittee on atherosclerosis and hypertension in childhood of the council on cardiovascular disease in the young, American Heart Association. *Circulation, 85*(4), 1638.

Thompson, W. R., Gordon, N. F., & Pescatello, L. S. (2009). *ACSM's guidelines for exercise testing and prescription* (8th ed.). Philadelphia, PA: Lippincott, Williams, and Wilkins.

van Poppel, M. N. M., Chinapaw, M. J. M., Mokkink, L. B., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for adults: A systematic review of measurement properties. *Sports Medicine, 40*(7), 565-600.

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine, 20*(1), 25-30.

Weinstock, J., Whelan, J. P., & Meyers, A. W. (2004). Behavioral assessment of gambling: An application of the timeline followback method. *Psychological Assessment, 16*(1), 72-79.

Preface

Chapter 2 provides information about the psychometric evaluation of the Timeline Followback for Exercise in the format to be submitted for publication. Additional details and explanation about the methods used in the study are printed in chapter 3 and a fuller discussion of the results is presented in chapter 4.

Chapter 2

Psychometric Evaluation of the Timeline Followback for Exercise among Young Adults

Gregory A. Panza[1], Jeremiah Weinstock[2], Garrett I. Ash[1], and Linda S. Pescatello[1]
[1]University of Connecticut
[2]Saint Louis University and University of Connecticut Health Center

Author Note

Gregory A. Panza, Department of Kinesiology, University of Connecticut; Garrett I. Ash, University of Connecticut; Linda S. Pescatello, Department of Kinesiology, University of Connecticut; Jeremiah Weinstock, Department of Psychology, Saint Louis University and Calhoun Cardiology Center - Behavioral Health, University of Connecticut Health Center.

We acknowledge the financial support for this research from the National Institutes of Health: R21-AA017717.

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

Email: Gregory.Panza@uconn.edu

ABSTRACT

Objectives: Two separate studies assessed 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 ± 1.4yr. Validity of frequency, intensity, time, and type of exercise as assessed on the TLFB-E was examined using Pearson r correlations with accelerometers, weekly exercise contracts between participants and researchers, College Alumni Questionnaire, and a health-related physical fitness battery. Study two participants were a separate sample (N = 40) of college students 18.63 ± 1.0yr. 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 criterion validity when compared to accelerometers (rs = .35 to .39) and convergent validity when compared to weekly exercise contracts (rs = .65 to .80) and College Alumni Questionnaire (rs = .06 to .75). The TLFB-E displayed modest to adequate test-retest reliability (rs = .79 to .97) for exercise frequency, intensity, and time and moderate Kappa (k = .49) for exercise type. ***Conclusions:*** The TLFB-E is a reliable and valid measure of physical activity and improves upon quantity-frequency methods by enabling collection of the exercise components of an individual's daily physical activity over a specified time period.

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 (Garber et al., 2011), and the American Heart Association (Haskell et al., 2007) have released new physical activity (PA) 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 PA for 30 $min \bullet d^{-1}$ on ≥ 5 $d \bullet wk^{-1}$, or vigorous intensity PA for 20 $min \bullet d^{-1}$ on ≥ 3 $d \bullet wk^{-1}$ (American College Health Association [ACHA], 2010). Following the college years, PA participation progressively declines with approximately 22% of adults 25 – 64yr, 15% of those 65 – 74yr, and 6% of those $\geq$75yr participating in the recommended amount of PA (Center for Disease and Control [CDC], 2010). The development of reliable and accurate assessment tools for exercise among young adults is important in efforts to develop behavioral strategies to advert the decline in PA participation that occurs with aging.

Self-report PA questionnaires are widely used to assess PA (Strath, Bassett, & Swartz, 2004; Craig et al., 2003); however, many suffer from limitations such as over reporting (Rzewnicki, Auweele, & Bourdeaudhui, 2003) and often use quantity-frequency methods to collect PA information (Sobell, L.C. & Sobell, 1996). 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 PA 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, L.C. & 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., in press). 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. The TLFB is superior to quantity-frequency methods in terms of reliability and validity (Sobell, L.C. & Sobell, 1996). The TLFB is 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 PA.

We propose to adapt the TLFB to assess PA behavior that includes more structured and planned forms of exercise (TLFB-E). The TLFB-E has several potential advantages over other self-reported PA questionnaires including the ability to: (1) collect daily PA 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 PA behavior data longitudinally; and (4) provide tailored individual feedback about these exercise patterns. 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

PA/exercise engagement. The PA/exercise adaptation of the TLFB has not previously been empirically validated as an assessment tool.

Thus, the purposes of this study are to conduct two separate studies to assess the psychometric properties of the TLFB-E among college students. Study one is a validity study examining criterion, predictive, and convergent validity of the TLFB-E. Study two assesses test-retest reliability of the TLFB-E between two interviews separated by one month. In study one we hypothesized that the TLFB-E will display criterion, convergent, and predictive validity through correlations with other measures of exercise. In study two we hypothesized that 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 was derived from the National Institute of Health funded project entitled, *Motivational Interventions for Exercise in Hazardous Drinking College Students* (MILE) (R21-AA017717). MILE 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±1.4yr, and normal weight [body mass index = 24.5±3.3 kg/m²]. Participant classification by racial 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 covering the previous two months, 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 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 research assistants. Participants were asked to complete TLFB-

E calendars covering the past two months. Research assistants read the *TLFB User's Guide* (Sobell, L.C. & Sobell, 1996). Prior to conducting participant interviews for data collecting purposes, research assistants were trained and administered 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 two ways with: (1) the Rating of Perceived Exertion Borg Scale (RPE-Scale) (Borg, 1998) and (2) metabolic equivalents (METs) for each exercise bout reported calculated using the compendium of PA (Ainsworth et al., 2011). *Time* was expressed as min 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.

College Alumni Questionnaire. Question four of the College Alumni Questionnaire gathers information on exercise participation in the previous two months and is a valid and reliable self-report measure of PA behavior among college students (Strath et al., 2004; Ainsworth, Leon, Jacobs, & Paffenbarger, 1993). Research assistants administered the College Alumni Questionnaire to ensure understanding of the questions asked. Exercise variables from question four of the College Alumni Questionnaire were: Total bouts of exercise (*frequency*), average

MET hours, total Kcal (*intensity*), total min (*time*), total aerobic bouts, resistance bouts, and flexibility bouts (*type*).

Objective Physical Activity / Exercise Measures

Actical® Accelerometer. An omnidirectional Actical® accelerometer (Mini Mitter, Bend OR, USA), 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 month. PA variables collected from the accelerometers were: Total aerobic exercise bouts (*frequency*), total min 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 ≥ 20min 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 min of moderate to vigorous (≥3 METs) exercise expressed as total time over the four days.

Exercise contract. Participants met with study personnel weekly during the baseline to two month assessment period to review prior week's 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 min). 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 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 min before RHR in beats per min was recorded.

Resting Blood Pressure. BP was used as a measure of cardiovascular health (Thompson et al., 2009). Subjects were seated quietly for at least 10 min 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). Subjects 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 in (2.54 cm) above the umbilicus or at the smallest circumference to the nearest 0.2 in (0.5 cm). Multiple measures were taken until two measures were within ¼ in (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). Estimated maximal aerobic capacity was expressed in mL·kg·min.

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 product-moment correlation coefficients with $p < .05$ established as the level of significance for the TLFB-E compared to accelerometer (criterion validity, hypothesis 1), exercise contract (convergent validity, hypothesis 2), question four of the College Alumni Questionnaire (convergent validity, hypothesis 3), and health-related fitness assessments (predictive validity, hypothesis 4). Paired t-tests examined the presence of under and over reporting on the TLFB-E.

Several PA questionnaires have been validated using samples with a wide spectrum of PA 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). To ensure a range of exercise engagement, two month TLFB-E, 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). As part of the larger study's inclusion/exclusion criteria all participants were sedentary at baseline. Accelerometer data were only collected at baseline and two month, therefore accelerometer data were selected from these two time points. There were fewer assessments completed at six months than baseline and two month assessments. To ensure similar relative sample sizes at each assessment time frame, a time period selected with no available data was re-randomized between the two time points that data were collected. 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 displayed a 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 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 normal weight.

As shown in Table 2, a wide spectrum of mean PA 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 sample fell below ACSM guidelines for all variables (Garber et al., 2011). However, PA levels increased from baseline to two month and continued to increase from two month to the six month time point.

Criterion Validity, Hypothesis 1. Table 3 displays validity coefficients between the TLFB-E and accelerometer. Correlations were significant for all variables (rs = .35 to .39, $ps < .01$), displaying criterion validity supporting hypothesis one.

Convergent Validity, Hypotheses 2 & 3. Validity coefficients (Table 3) between the TLFB-E and eight week exercise contract were significant for all variables (rs = .47 to .80, $ps < .001$), displaying 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 displayed a significant correlation for all variables assessed (*r*s = .49 to .75, *ps* < .01) except average MET hours per bout ($r = .06$, $p > .05$), displaying convergent validity of the TLFB-E supporting hypothesis three.

Predictive Validity, Hypothesis 4. As shown in Table 4, systolic BP displayed a negative relationship with total bouts ($p = .044$) and total aerobic bouts ($p = .001$) reported on the TLFB-E. Diastolic BP displayed a negative relationship with average MET hours per bout ($p = .043$). WC displayed a positive relationship with bouts of exercise on the TLFB-E that included aerobic and resistance ($p = .023$). Handgrip ($p = .009$) had a positive relationship with resistance bouts reported on the TLFB-E. Sit and reach results displayed a positive relationship with total bouts of exercise ($p = .034$) and total aerobic bouts ($p = .037$) reported on the TLFB-E. Pearson correlations among RHR, BMI, estimated aerobic capacity, and push-up test data and the corresponding variables assessed on the TLFB-E were not statistically significant (*p*s > .05). Based on results, hypothesis four was modestly 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 the College Alumni Questionnaire, participants under reported resistance bouts of exercise on the TLFB-E ($p = .019$) and displayed higher 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 local state university and received class research credit for completing the 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±1.0yr. Participant breakdown by racial 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, <18yr, 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 PA/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 that the TLFB-E would display modest test-retest reliability (hypothesis 5) and moderate Kappa statistic (hypothesis 6).

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 min (*time)* from interview 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 min (*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 min (*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 modest to adequate test-retest reliability (Nunnally and Bernstein, 1994) supporting hypothesis 5. Kappa = .49 ($p < .05$) for *type* of exercise reported and indicated a moderate classification agreement rate 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 on objective and subjective measures of PA/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 is a valid and reliable instrument for assessing self-reported exercise behavior among college students.

The magnitude of the correlations we found between the TLFB-E and accelerometer (rs = 0.35 to 0.39, ps < .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; rs = .51 to .97, ps < .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 of 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 *intensity* of the behavior. One way 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 PA, Sallis & 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 than these other self-report questionnaires when compared to objective measures of PA.

Correlations comparing the TLFB-E to weekly exercise contracts and 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 and colleagues (2004) found for predictive validity of the TLFB for substance use. Slightly lower predictive validity correlations may have been caused by two month health-related fitness assessment data that did not display health-related benefits or changes of exercise. The scientific literature suggests mental and physical health benefits of exercise do not begin to manifest until at least 12 weeks and typically more likely after 16 weeks (Thompson et al., 2009). Two month data represented approximately 40% of the sample.

Discrepancies were noticed for exercise *frequency, intensity* (Kcal), and *type* when the TLFB-E was compared to weekly exercise contracts and 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 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 exercise bouts that were not able to be verified on the exercise contracts due to lack of "proof of participation." Therefore, this unverified exercise was not counted as exercise on the weekly contracts. Additionally, exercise bouts may have been completed outside of the exercise contracts that were not accounted for at weekly exercise contracting sessions. However, participants may have reported on the TLFB-E all exercise completed outside of verified and contracted exercise in addition to exercise that was contracted, leading to over reporting on the TLFB-E compared to weekly exercise contracts. The discrepancy of under reporting on the TLFB-E may be due to the College Alumni Questionnaire using the quantity-frequency method to collect exercise.

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 modest to adequate test-retest reliability supporting hypothesis five (Nunnally and Bernstein, 1994). Correlations for test-retest reliability of the TLFB-E (*r*s = .79 to .97, *p*s < .01) were similar in magnitude to correlations for reliability of the TLFB for other complex health-related behaviors (*r*s = .55 to .99, *p*s < .05; Brown, Burgess, Sales, Evans, & Miller, 1998; Ehrman & Robbins, 1994; Hodgins & Makarchuck, 2003; Weinhardt et al., 1998). Kappa statistic indicated a moderate classification 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 just by chance supporting hypothesis six.

Significant validity and reliability results support the TLFB-E as a potentially preferred measure compared to current psychometrically supported retrospective self-report PA measures. Although the TLFB-E may take longer to administer (approximately 20 min per 2 months) than many other self-report PA measures, it has the capability of providing a greater depth of information about exercise/PA behavior compared to many existent self-report PA questionnaires. Most current PA questionnaires lack the ability to provide detailed information about the FITT of exercise over an extended period of time. Additionally, several other self-report PA questionnaires 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 is activities of less than 10 minutes and activities with a level of 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 of past exercise over two months.

Limitations were present in the current study. First, the event marker feature available for accelerometers was not used; therefore accelerometer data ≥ 3 METs (moderate to vigorous intensity) for ≥ 20 min 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. Bouts of exercise that included more than one modality may have been reported as multiple bouts of exercise. This may explain the slight under reporting of the TLFB-E when compared to the College Alumni Questionnaire. Third, the current study only investigates the psychometric properties of the TLFB-E among the college student population. 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 other populations to enhance its generalizability. Lastly, a majority of the data from the sixty day TLFB-E, College Alumni Questionnaire, and health-related fitness battery although randomly selected involved college students that were sedentary or not meeting the PA levels recommended by the ACSM (Garber et al., 2011).

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 analysis using a wide spectrum of PA 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 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 PA behaviors. Most existing self-report questionnaires (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 behavior that happen over time are not adequately assessed (Sobell, L.C. & Sobell, 1996; van Poppel et al., 2010). Like the psychometrically-supported TLFBs for other behaviors, we have shown that the TLFB-E is a reliable and valid PA measure and improves upon quantity-frequency methods by enabling collection of the FITT components of an individual's daily PA over a specified period of time.

References

Actical: Physical activity monitoring system—instruction manual software version 2.0. Oregon, USA: Mini Mitter Company, Inc.

Ainsworth B.E., Haskell W.L., Herrmann S.D., Meckes N., Bassett Jr. D.R., Tudor-Locke C., Greer J.L., Vezina J, Whitt-Glover M.C., Leon A.S. (2011). Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise*, 43(8):1575-1581.

Ainsworth, B. E., Leon, A. S., Richardson, M. T., Jacobs, D. R., & Paffenbarger, R. S.,Jr. (1993). Accuracy of the college alumnus physical activity questionnaire. *Journal of Clinical Epidemiology, 46*(12), 1403-1411.

American College Health Association. (2010). *American college health association-national college health assessment II, reference group executive summary, fall 2010.* Retrieved August 1, 2011 from http://www.achancha.org/docs/ACHA-NCHA_ Reference_ Group_ ExecutiveSummary_Fall2009.pdf.

Behrens, T., & Dinger, M. (2003). A preliminary investigation of college students' physical activity patterns. *American Journal of Health Studies, 18*(2/3), 169-172.

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Champaign, IL: Human Kinetics Publishers.

Brown, R. A., Burgess, E. S., Sales, S. D., Whiteley, J. A., Evans, D. M., & Miller, I. W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors, 12*(2), 101-112.

Center for Disease Control (2010). *Early release of selected estimates based on data from the 2010 national health interview survey.* Retrieved August 1, 2011, from http://www.acha-ncha.org/docs/ACHA-NCHA-II_ReferenceGroup_ExecutiveSummary_Fall2010.pdf.

Chung, P. K., & Yuen, C. K. (1999). Criterion-related validity of sit-and-reach tests in university men in hong kong. *Perceptual and Motor Skills, 88*(1), 304-316.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E.,.. Sallis, J. F. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise, 35*(8), 1381.

Dennis, M.L., Funk, R., Godley, S.H., Godley, M.D., Waldron, H. (2004). 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, 99(2), 120-128.

Dishman, R. K. & Steindhardt, M. (1988). Reliability and concurrent validity for a 7-d recall of physical activity in college students. *Medicine & Science in Sports & Exercise, 20*, 14-25.

Donovan, D. M., Bigelow, G. E., Brigham, G. S., Carroll, K. M., Cohen, A. J., Gardin, J. G., . . . Lindblad, R. (2011). Primary outcome indices in illicit drug dependence treatment research: Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end‐points in clinical trials. *Addiction,* Article first published online: 22 July 2011. doi: 10.1111/j.1360-0443.2011.03473.x.

Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62*(4), 843-850.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2003). The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence, 18*(3), 131-142.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. *New York: John Wiley & Sons.*

Garber, C. E., Bryan Blissmer, Michael R. Deschenes, Barry A. Franklin, Michael J. Lamonte, I-Min, . . . David P. Swain. (2011). 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. 43*(7), 1334-1359.

Hamilton, G. F., McDonald, C., & Chenier, T. C. (1992). Measurement of grip strength: Validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of Orthopedic and Sports Physical Therapy, 16*(5), 215-219.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S.N., Franklin, B. A., Macera, C. A., Heath, G. W., Thompson, P. D., Bauman, A. (2007). 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, 39*(8), 1423–1434.

Health and Human Services (2008). *Introducing the 2008 physical activity guidelines for Americans.* Retrieved August 1, 2011, from http://www.health.gov/paguidelines/guidelines /chapter1.aspx.

Hodgins, D. C., & Makarchuk, K. (2003). Trusting problem gamblers: Reliability and validity of self-reported gambling behavior. *Psychology of Addictive Behaviors, 17*(3), 244-248.

Hsu, L. M., & Field, R. (2003). Interrater agreement measures: Comments on kappa n, cohen's kappa, scott's π, and aickin's α. *Understanding Statistics, 2*(3), 205-219.

Kriska, A., & Caspersen, C. (1997). A collection of physical activity questionnaires for health-related research. *Medicine & Science in Sports & Exercise, 29*(6), S203-S205.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159.

McLellan, A. T., Alterman, A. I., Cacciola, J., Metzger, D., & O'Brien, C. P. (1992). A new measure of substance abuse treatment initial studies of the treatment services review. *Journal of Nervous and Mental Disease, 180*(2), 101-110.

Mozumdar, A., Liguori, G., & Baumgartner, T. A. (2010). Additional revised push-up test norms for college students. *Measurement in Physical Education and Exercise Science, 14*(1), 61-66.

Nelson, W. A., & Clum, G. A. (2002). Assessment of panic frequency: Reliability and validity of a time-line follow-back method. *Journal of Psychopathology and Behavioral Assessment, 24*(1), 47-54.

Nunnally, J. C., & Bernstein, I. H. Psychometric Theory (1994). New York, NY: McCraw-Hill.

Pickering, T. G., Hall, J. E., Appel, L. J., Falkner, B. E., Graves, J., Hill, M. N., . . . Roccella, E. J. (2005). 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, 111*(5), 697-716.

Rzewnicki, R., Auweele, Y. V., & Bourdeaudhuij, I. D. (2003). Addressing overreporting on the international physical activity questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition, 6*(3), 299-305.

Sallis, J. F., Saelens, B. E. (2000). Assessment of physical activity by self-report: Status, limitations, and future directions. *Research Quarterly for Exercise and Sport.* 71(2), S1-14.

Saunders, J. B., Aasland, O. G., Amundsen, A., & Grant, M. (2006). Alcohol consumption and related problems among primary health care patients: WHO collaborative project on early detection of persons with harmful alcohol consumption—I. *Addiction, 88*(3), 349-362.

Sobell, L. C., & Sobell, M. B. (1996). 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.

Strath, S. J., Bassett, D. R., & Swartz, A. M. (2004). Comparison of the college alumnus questionnaire physical activity index with objective monitoring. *Annals of Epidemiology, 14*(6), 409-415.

Thompson, W. R., Gordon, N. F., & Pescatello, L. S. (2009). *ACSM's guidelines for exercise testing and prescription* (8^{th} ed.). Philadelphia, PA: Lippincott, Williams, and Wilkins.

Tudor-Locke C. E., Myers A. M. (2001). Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Medicine, 31*, 91-100.

van Poppel, M. N. M., Chinapaw, M. J. M., Mokkink, L. B., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for adults: A systematic review of measurement properties. *Sports Medicine, 40*(7), 565-600.

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine, 20*(1), 25-30.

Weinstock, J., Whelan, J. P., & Meyers, A. W. (2004). Behavioral assessment of gambling: An application of the timeline followback method. *Psychological Assessment, 16*(1), 72-79.

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 (mL·kg·min)	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 *M* (*SD*) *n* = 66	Baseline *M* (*SD*) *n* = 29	Two Month *M* (*SD*) *n* = 26	Six Month *M* (*SD*) *n* = 11	All Time Points *p*	Baseline vs Two Month *p*	Two Month vs Six Month *p*
TLFB-E (2 months)	All Bouts	17.8 (11.9)	8.2 (4.2)	23.7 (7.2)	29.0 (15.6)	<.001	<.001	.053
	Aerobic Bouts	13.2 (11.0)	6.0 (4.2)	15.9 (8.4)	25.5 (15.1)	<.001	<.001	.007
	Resistance Bouts	3.3 (4.6)	0.8 (1.9)	6.1 (5.4)	3.4 (3.8)	<.001	<.001	.262
	Flexibility Bouts	0.3 (1.5)	0.0 (0.0)	0.7 (2.3)	0.0 (0.0)	.168	.107	.356
	Aerobic & Resistance Bouts	1.0 (2.5)	1.3 (2.5)	1.0 (2.9)	0.1 (.30)	.380	.709	.324
	Total Time (min)	1,090.0 (1,121.0)	414 (256.5)	1319.5 (773.9)	2329.7 (1822.0)	<.001	<.001	.009
	Average RPE	13.8 (1.43)	13.0 (1.3)	14.3 (1.4)	14.4 (0.7)	.001	.001	.597
	Average MET hrs/bout	6.3 (5.1)	5.98 (5.4)	5.3 (2.5)	9.3 (7.6)	.076	.531	.019
	Total Kcal	8504.3 (9285.1)	3198.1 (2112.7)	9969.6 (6738.3)	9929.9 (8077.2)	<.001	<.001	.988
College Alumni Questionnaire	All Bouts	20.5 (17.3)	9.9 (12.3)	28.1 (13.8)	30.4 (21.3)	<.001	<.001	.363
	Aerobic Bouts	14.6 (13.0)	7.7 (8.6)	18.2 (10.5)	24.5 (18.5)	<.001	<.001	.079
	Resistance Bouts	5.6 (8.7)	2.3 (6.0)	9.6 (10.2)	4.8 (7.3)	.006	.002	.277
	Flexibility Bouts	0.3 (1.6)	0.0 (0.0)	0.4 (2.0)	1.0 (2.5)	.205	.295	.359
	Total Time (min)	878.6 (435.0)	598.4 (804.0)	1693.8 (1344.3)	1902.3 (1436.5)	.001	<.001	.419
	Average MET hrs/bout	5.3 (4.2)	4.6 (5.0)	5.9 (3.8)	6.0 (2.3)	.415	.261	.960
	Total Kcal	8504.3 (9285.1)	4541.2 (5895.7)	10972.8 (8744.7)	13118.1 (13677.6)	.006	.002	.570
Health Related Fitness Tests	Resting Heart Rate (bpm)	72.5 (9.6)	74.5 (9.8)	70.5 (10.3)	71.8 (6.5)	.284	.138	.751
	Systolic BP (mmHg)	116.3 (10.2)	117.6 (10.0)	116.0 (10.0)	113.5 (11.4)	.538	.567	.285
	Diastolic BP (mmHg)	67.2 (7.2)	66.4 (8.6)	67.6 (6.3)	68.4 (5.6)	.702	.572	.999
	Body Mass Index (kg/m^2)	24.3 (3.3)	24.0 (3.3)	24.4 (2.5)	24.6 (4.9)	.844	.603	.371
	Waist Circumference (cm)	78.0 (10.8)	77.1 (13.2)	80.0 (6.6)	75.7 (12.1)	.456	.313	.524
	Push-Ups (reps)	26.0 (15.3)	23.9 (10.4)	27.5 (20.3)	30.5 (2.1)	.762	.556	.840
	Handgrip (kg)	28.8 (10.3)	25.9 (9.2)	30.4 (11.0)	33.2 (10.0)	.092	.104	.492
	Sit & Reach (cm)	32.2 (10.2)	29.6 (10.9)	33.3 (8.2)	36.7 (11.5)	.069	.244	.321
	YSET (mL·kg·min)	36.8 (6.5)	36.6 (7.2)	37.5 (5.7)	35.7 (6.6)	.716	.575	.492
			n = 32	*n* = 34				
TLFB-E (4 days)	All Bouts	0.8 (1.0)	0.5 (0.8)	1.1 (1.1)	–	–	.021	–
	Total Time (min)	50.2 (68.0)	31.7 (50.6)	66.5 (77.4)	–	–	.040	–
	Total Kcal	389.8 (520.3)	243.2 (416.8)	519.0 (572.1)	–	–	.033	–
Accelerometer (4 days)	All Bouts	1.3 (1.1)	1.0 (1.1)	1.5 (1.0)	–	–	.083	–
	Total Time (min)	57.5 (56.1)	53.1 (63.7)	61.3 (49.1)	–	–	.565	–
	Total Kcal	262.3 (261.6)	252.1 (298.9)	271.3 (227.9)	–	–	.771	–

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. *P* values represent differences among time points. YSET = YMCA submax cycle ergometer test; TLFB-E = Timeline Followback for Exercise; RPE = rating of perceived exertion (6-20); Kcal = Kilocalories.

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 (mL·kg·min)
	r	*r*	*r*	*r*	*r*	*r*	*r*	*r*	*r*
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

	Accelerometer Discrepancy Score		Exercise Contract Discrepancy Score		Question four of the CAQ Discrepancy Score	
Variable	*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.

Chapter 3 - Methods

Study 1 – Validity

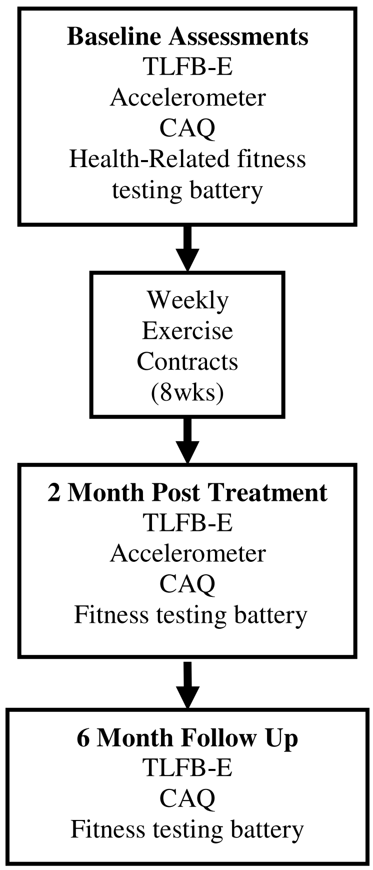
Participants

Participants (*N* = 66, *n* = 37 women, *n* = 29 men) were English speaking, currently enrolled in college, 20.0±1.4yr, and normal weight [body mass index = 24.5±3.3 kg/m^2]. Participant classification by racial 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 Procedure

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; See Figure 1). At all three assessments, participants completed the TLFB-E covering the previous two months, 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 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.

Figure 1. *Validity of the Timeline Followback for Exercise: Study Overview*



Note. TLFB-E = Timeline Followback for Exercise; PA = Physical Activity; CAQ = College Alumni Questionnaire

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 research assistants. Participants were asked to complete TLFB-E calendars covering the past two months. Research assistants read the *TLFB User's Guide* (Sobell, L.C. & Sobell, 1996). Prior to conducting participant interviews for data collecting purposes, research assistants were trained and administered pilot interviews under the supervision of a clinical psychologist (JW) experienced in using the TLFB.

The TLFB-E represented a traditional monthly calendar (See Appendix A) and assessed the frequency, intensity, time, and type (FITT) components of exercise (See Table 1). Frequency refers to *how often* the exercise bouts take place (i.e., 3d·wk). Intensity refers to *how hard* the exercise is (i.e., light, moderate, vigorous). Time refers to *how many* and *how long* each exercise bout is (i.e., 30 mins·d, 5 d·wk). Type refers to the modality or *kind* of activity completed (i.e., walking, resistance training, cycling).

Table 1. *Exercise Variables (FITT) Produced by the TLFB-E*

FITT	Unit(s) of Measure	How Variable is Produced
Frequency	Bouts	• Total Bouts
Intensity	MET hours, Kcal, & RPE	• Compendium of physical activities (Ainsworth et al., 2000) • Borg Scale (6-20; Borg, 1998)
Time	Minutes	• Total Time
Type	Modality	• e.g., Aerobic, Resistance, Flexibility, Aerobic & Resistance combination

MET = Metabolic equivalent; Kcal = Kilocalories; RPE = Rating of Perceived Exertion; TLFB=E = Timeline Followback for Exercise; MET = Metabolic Equivalents.

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.

College Alumni Questionnaire. Question four of the College Alumni Questionnaire gathers information on exercise participation in the previous two months and is a valid and reliable self-report measure of PA behavior among college students (Strath et al., 2004; Ainsworth, Leon, Jacobs, & Paffenbarger, 1993; See Appendix B). Research assistants administered the College Alumni Questionnaire to ensure understanding of the questions asked. Exercise variables from question four of the College Alumni Questionnaire were: Total bouts of exercise (*frequency*), average MET hours, total Kcal (*intensity*), total min (*time*), total aerobic bouts, resistance bouts, and flexibility bouts (*type*).

Objective Physical Activity / Exercise Measures

Actical® Accelerometer. An omnidirectional Actical® accelerometer (Mini Mitter, Bend OR, USA), 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 month. PA variables collected from the accelerometers were: Total aerobic exercise bouts (*frequency*), total min 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 extraction calculations are displayed in Table 2.

Table 2. *Methods for Extraction of Exercise Components of an Aerobic Exercise Bout from Accelerometer Data*

Variable	Unit of Measure	Criteria / Extraction Method
Frequency	Bouts	Consistent moderate intensity rating of ≥ 3 METs for ≥ 20min. = 1 aerobic exercise bout
Intensity	Kcal	Sum of Kcal/min./kg. x body weight (kg) = Total Kcal expended for exercise bout
Time	Minutes	Sum of total consistent minutes of moderate to vigorous exercise over 4d
Type	N/A	N/A

Note. Criteria adapted from *Actical:* Physical activity monitoring system—instruction manual software version 2.1 (Mitter, 2003). METs = metabolic equivalents; Kcal = Kilocalorie.

Exercise Contract. Participants met with study personnel weekly during the baseline to two month assessment period to review prior week's 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 min; See Appendix C). 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 of all bouts, average MET hours, and total Kcal expended (*intensity*; Borg, 1998).

Health-Related Fitness Assessments

All 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 min before RHR in beats per min was recorded.

Resting Blood Pressure. BP was used as a measure of cardiovascular health (Thompson et al., 2009). Subjects were seated quietly for at least 10 min 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). Subjects were asked to refrain from exercise, smoking cigarettes or ingesting caffeine at least 24 hours prior to the time of measurement. The bladder inside the cuff encircled 80% of the arm circumference with a minimal amount of Velcro showing. The cuff was placed snug around the upper arm at heart level, and the center of the bladder was placed directly above the brachial artery.

The lower edge of the cuff was approximately 1 in (2.50cm) above the antecubital fossa (bend of the elbow; Pickering et al., 2005). The researcher ensured the subject did not push sleeved clothing up on the arm in order to avoid a tourniquet effect. 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 (Thompson et al., 2009). Participants were asked to remove footwear and all items from their pockets as well as any additional heavy clothing (e.g., sweatshirts, jackets, etc.) and asked to stand on the scale. Weight was recorded in lbs and later converted into kg. Height was measured by having the participant stand with their back facing the scale and the height rod resting at a 90 degree angle on top of the participant's head. BMI was determined by the following calculation: BMI (kg/m^2) = Weight (kg) / Height $(m)^2$ (weight in kilograms divided by height in meters squared; Thompson et al., 2009). BMI can be used to classify disease risk relative to normal weight using a Classification of Disease Risk scale (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 in (2.54 cm) above the umbilicus or at the smallest circumference to the nearest

0.2 in (0.50 cm). The measuring tape was applied to the site and pulled taut but not tight to avoid pinching of the skin. The participants were asked to stand up straight and relaxed with arms on their side at all times. Multiple measures were taken until two measures were within ¼ in (0.64 cm; Thompson et al., 2009). Waist circumference can be used to classify disease risk using a Classification of Disease Risk scale (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). Male participants assumed a standard push-up position, with hands pronated, flush with the floor, directly beneath the shoulder, with fingers pointed forward. The toes were the pivot-point, with feet together or up to 12 in apart. The participant's head was kept up and their back straight at all times. Female participants used the modified position, where hands are aimed forward, shoulder width apart, and the knees are the pivot-point, resting on a mat. The participant's legs were kept together at all times, keeping the lower legs in contact with the mat, and ankles plantar flexed. The head was kept up and the back straight. At the top of the range of motion, male and female participants were required to reach a straight-arm position. At the bottom of the range of motion, the chin had to have touched the mat, avoiding any contact of the stomach to the mat. All participants started in the down-position, chin touching the mat, and repetitions were counted upon each return to this position. 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). The participants were seated comfortably (not slouched) in a standard height chair. The grip bar was adjusted by the researcher to fit comfortably within the participant's hand with the second joint of the fingers fitting under the handle of the handgrip dynamometer. The participant's shoulder was adducted and neutrally rotated, elbow flexed at 90 degrees with the forearm and wrist in neutral position (Thompson et al., 2009). The researcher provided verbal instructions to begin squeezing the handgrip dynamometer as hard as possible and not to hold their breath. The instructor ensured no rapid jerking or wrenching motions. Grip strength of the dominant hand was recorded to the nearest 1.0 kg for each trial.

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). Participants were asked to remove their shoes and sit on a mat with both legs extended and feet flat against the sit and reach box (Figure Finder-Flex-Tester, Novel Products, Inc., Rockton, IL). Participants extended their arms and bent at the waist with their middle fingers overlapping one another without bending their knees. The participants performed three sub-maximal reaches followed by a fourth maximal reach. The fourth maximal reach was held for 2 seconds and recorded. (Thompson et al., 2009). 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). Appropriate seat height was determined by having the participant in a pedal-down position with their toes on the pedals, plantar flexed, displaying a flexed knee of about 30 degrees. The YSET began with a 2 min warm-up of free wheeling at the pedaling cadence of 50 revolutions per min (rpm) in order to familiarize participants to the cycle ergometer (Monark Ergometric 818, Stockholm, Sweden). Participants pedaled at a cadence of 50 rpm throughout the entire test. Upon completion of the warm-up, researchers applied 150kgm to begin the first stage. A minimum of three stages but up to four stages were completed. Each stage was a minimum of 3 min and the workload was increased in accordance to the YMCA cycle ergometry progression protocol (Table 3).

Table 3. *YMCA Cycle Ergometry Progression Protocol*

1st Stage	150 kgm/min (0.5 kg)

Stage	HR<80bpm	HR 80-89 bpm	HR 90-100 bpm	HR>100 bpm
2nd Stage	750 kgm/min (2.5 kg)	600 kgm/min (2.0 kg)	450 kgm/min (1.5 kg)	300 kgm/min (1.0 kg)
3rd Stage	900 kgm/min (3.0 kg)	750 kgm/min (2.5 kg)	600 kgm/min (2.0 kg)	450 kgm/min (1.5 kg)
4th Stage	1050 kgm/min (3.5 kg)	900 kgm/min (3.0 kg)	750 kgm/ min (2.5 kg)	600 kgm/min (2.0 kg)

Note. bpm = beats per minute; HR = heart rate kgm/min = kilogram-meter / minute.

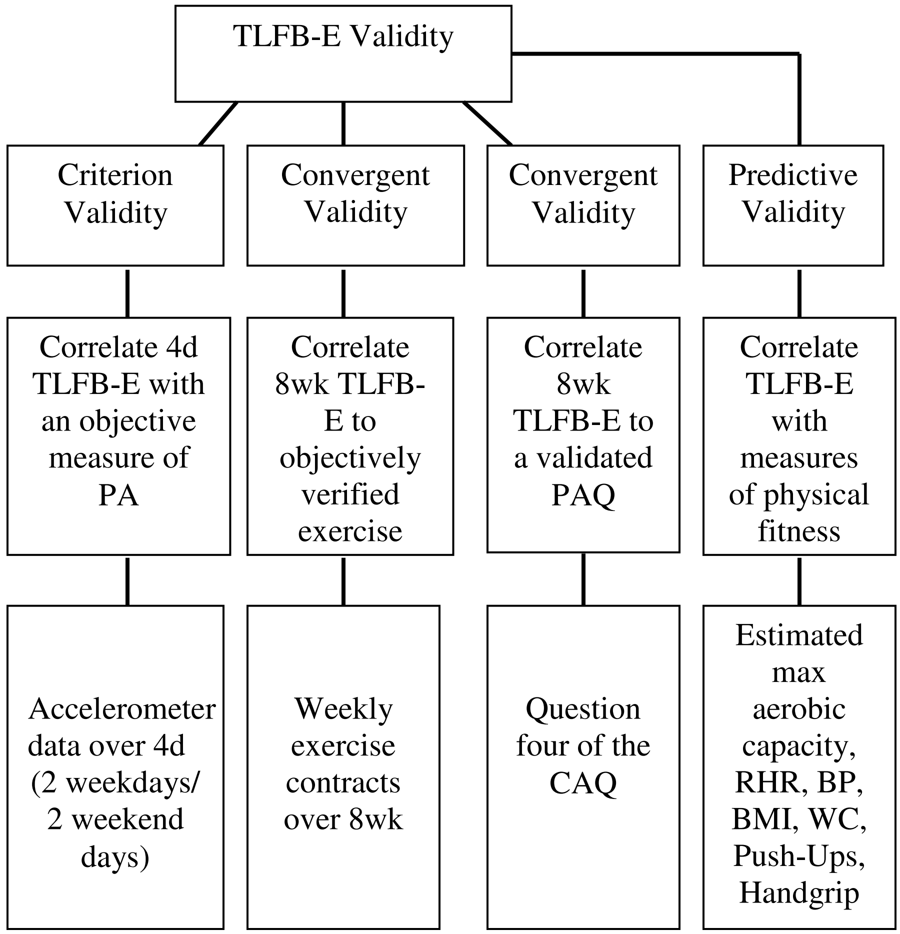
HR was monitored at least two times during each stage, near the end of the second and third minutes using a Polar T31-Coded Heart Rate Monitor (Polar Electro Oy, Kempele, Finland) and Polar Heart Rate Watch model F6 ceo537 (Polar Electro Oy, Kempele, Finland). The work rate was not increased when HR > 110 beats per min (bpm) until steady state HR (i.e., two HR within ± 5 bpm) was reached for each stage. Blood pressure (systolic / diastolic) was measured at approximately 2 min of each stage using an

American Diagnostic Corporation Sphygmomanometer (ADC, Hauppauge, NY) and a 3M™ Littmann® Stethoscope model Lightweight (3M Center, St. Paul, MN). Rate of perceived exertion was obtained near the end of the third minute of each stage using the Borg (6-20) scale (Borg, 1998). The test was terminated when the subject reached 85% of age-predicted maximal HR (70% of heart rate reserve), failed to conform to the exercise test protocol, experienced adverse signs or symptoms, requested to stop, or experienced an emergency situation. The YSET concluded with a 5 min cool-down/recovery period during which the participant continued to pedal at stage one intensity or lower, and all physiologic observations continued to be measured (e.g., HR, BP, signs and symptoms; Thompson et al., 2009). HR and workrates were used to predict cardiorespiratory maximal capacity using the YMCA plotting technique (Appendix D; Thompson et al., 2009). Maximum aerobic capacity was estimated and expressed in mL·kg·min.

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 product-moment correlation coefficients with $p < .05$ established as the level of significance for FITT reported on the TLFB-E compared to accelerometer (criterion validity, hypothesis 1), exercise contract (convergent validity, hypothesis 2), question four of the College Alumni Questionnaire (convergent validity, hypothesis 3), and health-related fitness assessments (predictive validity, hypothesis 4; See Figure 2). Paired t-tests examined the presence of under and over reporting on the TLFB-E.

Figure 2. *Methods Used to Examine Criterion, Convergent, and Predictive Validity of the Timeline Followback for Exercise*



Note. TLFB-E = Timeline Followback for Exercise; PA = physical activity; RHR = resting heart rate; BP = blood pressure; BMI = body mass index; WC = waist circumference; CAQ = College Alumni Questionnaire.

Several PA questionnaires have been validated using samples with a wide spectrum of PA 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). To ensure a range of exercise engagement, two month TLFB-E, 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). As part of the larger study's inclusion/exclusion criteria all participants were sedentary at baseline. Accelerometer data were only collected at baseline and two month, therefore accelerometer data were selected from these two time points. There were fewer assessments completed at six months than baseline and two month assessments. To ensure similar relative sample sizes at each assessment time frame, a time period selected with no available data was re-randomized between the two time points that data were collected. For baseline, 29 out of a possible 66 (44%) data time points were randomly selected for analysis. For the two month time point, 26 out of a possible 65 (40%) data time points were randomly selected for analysis. Finally, for the six month time point, 11 out of 30 (37%) data time points were randomly selected for analysis. All statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) version 14.0 (SPSS Inc., Chicago, IL).

Study 2 – Reliability

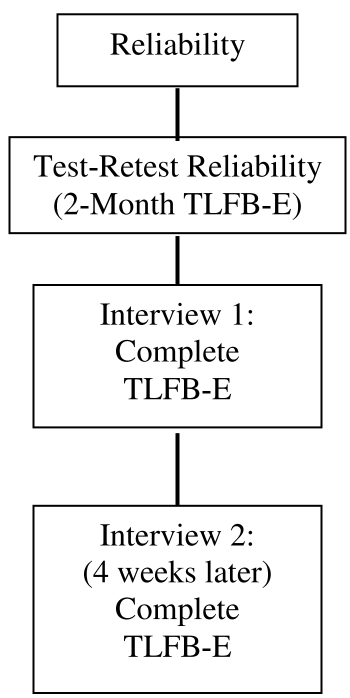
Participants

A separate sample of participants was recruited from the undergraduate subject pool in the communication sciences department at the University of Connecticut and received class research credit for completing the 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±1.0yr. Participant breakdown by racial 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, <18yr, and/or have previously filled out the TLFB-E.

Study Procedures

Participants met with study personnel two times. The first visit consisted of completion of a demographics questionnaire and a TLFB-E interview. Information obtained on the demographics questionnaire included: age, gender, ethnicity, marital status, GPA, and year in school. The TLFB-E collected information regarding PA/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 (See Figure 3).

Figure 3. *Methods Used to Examine Test – Retest Reliability of the Timeline Followback for Exercise*



Note. TLFB = Timeline Followback for Exercise

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 min (*time)* from interview one and two. Test-retest reliability criteria standards from Nunnally and Bernstein (1994) were used and included poor ($\leq .69$), modest ($\geq .70$) and adequate ($\geq .80$). 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). A kappa of one represents 100% agreement; a kappa of zero represents a chance agreement, while a negative kappa represents an agreement lower than expected by chance (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).

References

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. N. N. M., Strath, S. J., . . . Emplaincourt, P. O. (2000). Compendium of physical activities: An update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise, 32*(9), S498.

Ainsworth, B. E., Leon, A. S., Richardson, M. T., Jacobs, D. R., & Paffenbarger, R. S., Jr. (1993). Accuracy of the college alumnus physical activity questionnaire. *Journal of Clinical Epidemiology, 46*(12), 1403-1411.

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Human Kinetics Publishers.

Chung, P. K., & Yuen, C. K. (1999). Criterion-related validity of sit-and-reach tests in university men in Hong Kong. *Perceptual and Motor Skills, 88*(1), 304-316.

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Sallis, J. F. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise, 35*(8), 1381.

Dishman, R. K. & Steindhardt, M. (1988). Reliability and concurrent validity for a 7-d recall of physical activity in college students. *Medicine & Science in Sports & Exercise, 20*, 14-25.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. *New York: John Wiley & Sons.*

Hamilton, G. F., McDonald, C., & Chenier, T. C. (1992). Measurement of grip strength: Validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of Orthopedic and Sports Physical Therapy, 16*(5), 215-219.

Hsu, L. M., & Field, R. (2003). Interrater agreement measures: Comments on kappa n, cohen's kappa, scott's π, and aickin's α. *Understanding Statistics, 2*(3), 205-219.

Kriska, A., & Caspersen, C. (1997). A collection of physical activity questionnaires for health-related research. *Medicine and Science in Sports and Exercise, 29*(6), S203-5.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159.

McLellan, A. T., Alterman, A. I., Cacciola, J., Metzger, D., & O'Brien, C. P. (1992). A new measure of substance abuse treatment initial studies of the treatment services review. *Journal of Nervous and Mental Disease, 180*(2), 101-110.

Mozumdar, A., Liguori, G., & Baumgartner, T. A. (2010). Additional revised push-up test norms for college students. *Measurement in Physical Education and Exercise Science, 14*(1), 61-66.

Pickering, T. G., Hall, J. E., Appel, L. J., Falkner, B. E., Graves, J., Hill, M. N., . . . Roccella, E. J. (2005). 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, 111*(5), 697.

Rzewnicki, R., Auweele, Y. V., & Bourdeaudhuij, I. D. (2003). Addressing overreporting on the international physical activity questionnaire (IPAQ) telephone survey with a population sample. *Public Health Nutrition, 6*(03), 299-305.

Saunders, J. B., Aasland, O. G., Amundsen, A., & Grant, M. (2006). Alcohol consumption and related problems among primary health care patients: WHO collaborative project on early detection of persons with harmful alcohol consumption—I. *Addiction, 88*(3), 349-362.

Sobell, L. C., & Sobell, M. B. (1996). Timeline followback: A calendar method for assessing alcohol and drug use. User's guide. *Addiction Research Foundation,*

Strath, S. J., Bassett, D. R., & Swartz, A. M. (2004). Comparison of the college alumnus questionnaire physical activity index with objective monitoring. *Annals of Epidemiology, 14*(6), 409-415.

Thompson, W.R., PhD, FACSM, Gorden, N.F., PhD, FACSM, Pescatello, PhD, FACSM (Ed.). (2009). *ACSM's guidelines for exercise testing and prescription eighth edition.*

Chapter 4 – Discussion

The primary purpose of this thesis 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 on objective and subjective measures of PA/exercise. We sought to test reliability by using a test-retest method between two interviews separated by one month. This chapter serves as a synthesis and conclusion of the findings. It will be organized first by discussing the specific aims and hypotheses along with relevant findings. Then, the significance of the findings as related to the current literature will be explored by discussing the benefits of the TLFB-E as a self-report measure. Finally, future research pertaining to the findings will be discussed.

Specific Aims & Hypotheses – Study 1

Specific Aim 1: To assess criterion validity of the TLFB-E by examining the relationship among aerobic exercise reported on the TLFB-E and data obtained from accelerometers on matching days and time periods of 96 hours (i.e., 2 weekdays / 2 weekend days).

Hypothesis 1: Aerobic bouts of exercise recorded on the TLFB-E will correlate with aerobic bouts of exercise obtained with accelerometers over the four days. Correlations were significant for all variables assessed (rs = .35 to .39, $ps < .01$) displaying criterion validity supporting hypothesis one. The magnitude of correlations among the TLFB-E and accelerometers 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; rs = .51 to .97, ps < .05). However, the TLFB-E with a mean r of .37 has slightly higher criterion validity than other self-report questionnaires (mean $r = .30$) when compared to objective measures of PA (Sallis & Saelens, 2000).

Specific Aim 2: To evaluate convergent validity by examining the relationship among the TLFB-E and weekly exercise contracts administered over the same 8 week period. *Hypothesis 2:* There will be a positive correlation between exercise *frequency* (number of bouts), *intensity* (average MET hours, Kcal expended and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, combined aerobic and resistance) of exercise recorded on the TLFB-E and weekly exercise contracts. Validity coefficients between the TLFB-E and eight week exercise contract were significant for all variables assessed (rs = .65 to .80, $ps < .001$), displaying convergent validity of the TLFB-E supporting hypothesis two.

Specific Aim 3: To measure convergent validity by evaluating the association among FITT collected on the TLFB-E and FITT collected by question four of the College Alumni Questionnaire also known as the Paffenbarger Physical Activity Questionnaire (Bassett Jr. & Ainsworth, 2000). *Hypothesis 3:* There will be a positive correlation among exercise recorded on the TLFB-E and responses on the College Alumni Questionnaire for the following variables: *Frequency* (number of bouts), *intensity* (average MET hours, Kcal expended and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, combined aerobic and resistance) of exercise. Validity coefficients between the TLFB-E and question four of the College Alumni Questionnaire displayed a significant correlation for all variables assessed (rs = .49 to .75, $ps < .01$) except average MET hours per bout ($r = .06$, $p > .05$), displaying convergent validity of the TLFB-E supporting a majority of hypothesis three.

Correlations between the TLFB-E and weekly exercise contracts (*r*s = .47 to .80, *p*s < .001) were similar in magnitude to those found in convergent validity studies of the TLFB for gambling (Weinstock, Whelan, & Meyers, 2004) and panic frequency (Nelson & Clum, 2002; *r*s = .58 to .95, *p*s < .001). Correlations between the TLFB-E and question four of the CAQ for all variables assessed (*r*s = .49 to .75, *ps* < .01) except average MET hours per bout (*r* = .06, *p* >.05) were also similar in magnitude.

Specific Aim 4: The fourth aim of this study was to measure predictive validity by observing the relationship among exercise recorded on the TLFB-E and results from health-related physical fitness assessments including the YMCA submaximal bicycle ergometer test (YSET) (Thompson, Gordon, & Pescatello, 2009; Poldermans et al., 1993), handgrip dynamometer (Hamilton, McDonald, & Chenier, 1992; Rantanen et al., 1999), push-up test, sit-and-reach test, resting heart rate (RHR), resting blood pressure (BP), waist circumference (WC), and body mass index (BMI; Thompson et al., 2009).

Hypothesis 4: A positive or negative significant relationship will be displayed and will depend on the health-related fitness measure that the TLFB-E is being correlated with (e.g., negative relationship with BP, positive relationship with cardiorespiratory fitness). Variables correlated from the TLFB-E included: *Frequency* (number of bouts), *intensity* (average MET hours, Kcal expended, and rating of perceived exertion), *time* (total minutes), and *type* (aerobic, resistance, flexibility, or combined bout of aerobic and resistance) of exercise. Systolic BP displayed a negative relationship with total bouts (*p* = .044) and total aerobic bouts (*p* = .001) reported on the TLFB-E, indicating those who reported a higher total amount of exercise bouts as well as aerobic bouts reported on the TLFB-E tended to have lower systolic BP. Diastolic BP displayed a negative relationship

($p = .043$) with average MET hours per bout. This relationship indicates that those who reported a higher average MET hours per session on the TLFB-E tended to have lower diastolic BP. Relationships shown for systolic BP and diastolic BP with exercise reported on the TLFB-E are consistent with favorable changes from PA/exercise participation. Furthermore, these relationships between PA reported on the TLFB-E and BP agreed with our hypothesis and is supported by literature (Brandon & Elliot-Lloyd, 2006; King, Haskell, Taylor, Kraemer, & DeBusk, 1991; Krustrup et al., 2010). WC displayed a positive relationship with bouts of exercise on the TLFB-E that included aerobic and resistance ($p = .28$). This relationship indicates that those who participated in aerobic and resistance exercise tended to have a higher WC. This relationship was unexpected as the current literature suggests an increase in exercise / PA participation leads to a decrease in WC (Bigaard et al., 2005). Handgrip had a positive relationship with resistance bouts ($p = .32$) reported on the TLFB-E. This relationship indicates that those who reported higher rates of resistance training on the TLFB-E scored higher on the handgrip test. This relationship was expected and is supported by the literature (Peterson, Rhea, & Alvar, 2005). Sit and reach results displayed a positive relationship with total bouts of exercise ($p = .26$) and total aerobic bouts ($p = .26$) reported on the TLFB-E. This relationship indicates that those who reported higher rates of total bouts and aerobic bouts of exercise on the TLFB-E scored higher on the sit and reach test. This relationship was expected with flexibility exercise reported and not necessarily aerobic bouts (Garber et al., 2011). A possible reason why flexibility did not appear to have an association with sit & reach scores may have been because flexibility exercise reported on the TLFB-E was minimal. Pearson correlations among RHR, BMI, estimated maximal aerobic capacity, and push-up test data and the corresponding variables assessed on the TLFB-E were not statistically significant. Hypothesis four was modestly supported by correlations among the TLFB-E and health-related fitness assessments.

Specific Aims and Hypotheses – Study 2

Specific Aim 5: To determine test-retest reliability of the TLFB-E. *Hypothesis 5:* The TLFB-E will show modest test-retest reliability ($r \geq .70$) for *frequency* (number of bouts), *intensity* (Kcal expended and rating of perceived exertion), and *time* (total minutes). The TLFB-E demonstrated modest to adequate (rs = .79 to .97, $p < .001$) test-retest reliability (Nunnally and Bernstein, 1994) supporting hypothesis 5.

Specific Aim 6: To determine kappa statistic for *type* of exercise reported at interview one compared to *type* of exercise reported at interview two. *Hypothesis 6:* The TLFB-E will display a moderate kappa statistic (.41 – .60; Landis & Koch, 1977) for *type* of exercise reported at interview one and two. Kappa for *type* of exercise reported indicated a moderate classification agreement (k = .49, $p < .05$) rate between the two interviews (Landis & Koch, 1977) supporting hypothesis six.

Additional Findings

Discrepancy scores (i.e., mean differences) were calculated in order to determine discrepancies among FITT reported on the TLFB-E and FITT reported on the various subjective and objective measures of PA. Discrepancy scores indicated slight under reporting of total bouts and slight over reporting of Kcal expended over four days (p = .038) on the TLFB-E compared to accelerometer. Discrepancy scores for exercise contracts and TLFB-E indicated over reporting of total bouts, total aerobic bouts, total time, and displayed lower total Kcal expended on the TLFB-E over two months.

Compared to the College Alumni Questionnaire, participants under reported resistance bouts of exercise on the TLFB-E and displayed higher average MET hours per bout.

It is difficult to identify the direct cause of under and over reporting of the TLFB-E. Discrepancies may have been caused by: (1) Inaccurate reporting of exercise on the TLFB-E, (2) inaccurate reporting of exercise on the comparison measure, or (3) a combination of both. However, certain aspects of measures used in this thesis may help explain these under and over reporting discrepancies.

Under reporting of total aerobic bouts reported on the TLFB-E compared to total aerobic bouts recorded on accelerometers over four days may be explained by the method used to extract a bout of aerobic activity from accelerometer data. A bout of aerobic exercise was used as data when it met the following criteria: Moderate to vigorous intensity rating of ≥ 3 METs for ≥ 20 minutes. Event markers for accelerometers were not used in the larger study and therefore this extraction procedure was used as an alternative. Over reporting of Kcal expended as reported on the TLFB-E compared to the accelerometer may have been caused by inaccurate MET values given to activities reported on the TLFB-E. In many instances the small area that type of activity is recorded does not leave enough room for a fuller explanation of details of the exercise bout. If the administrator of the TLFB-E did not use the "notes" section to collect additional detail about the activity reported, the activity may not have been coded as accurate as possible when choosing MET values from the compendium of PA (Ainsworth et al., 2011).

Over reporting of the TLFB-E compared to weekly exercise contracts may be explained by exercise bouts that were not able to be verified on the exercise contracts due to lack of "proof of participation." Therefore, this unverified exercise was not counted as exercise on the weekly contracts. Additionally, exercise bouts may have been completed outside of the exercise contracts and not accounted for at weekly exercise contracting sessions. However, participants may have reported on the TLFB-E all exercise completed outside of verified and contracted exercise in addition to exercise that was contracted, leading to over reporting on the TLFB-E compared to weekly exercise contracts.

The discrepancy of under reporting on the TLFB-E may be due to the College Alumni Questionnaire using the quantity-frequency method to collect exercise. The average volume of exercise collected by the College Alumni Questionnaire does not account for bouts of exercise with more than one modality and instead includes them as multiple bouts. Therefore, it appears that participants are participating in more bouts of exercise, thus resulting in over reporting of exercise compared to the TLFB-E.

Impact of the Findings on the Current Literature

The TLFB has been psychometrically supported for use with multiple behaviors (Fals-Stewart, Birchler, & Kelley, 2003; Hodgins & Makarchuk, 2003; Weinstock, Whelan, & Meyers, 2004; Weinhardt et al., 1998; Brown et al., 1998; Nelson & Clum, 2002). The TLFB's international acceptance as a retrospective behavioral measure supports the purpose of its adaptation for use with exercise behavior (Weinstock et al., 2004; Hodgins & Makarchuk, 2003; Aalto, Tuunanen, Sillanaukee, & Seppa, 2006; Carlbring, Jonsson, Josephson, Forsberg, 2009; Collins, Eck, Torchalla, Schroter, Batra, 2009). This thesis is the first psychometric evaluation of the TLFB adapted for exercise and as a result may play a significant role in the potential addition of this version of the TLFB to the current literature of PA/exercise self-report assessments. However, in order to conclude that the TLFB-E is a well validated and reliable measurement tool, continuous research examining its psychometric properties is needed.

There are a multitude of self-report PA questionnaires available. Each self-report questionnaire displays its own unique advantages and disadvantages and are often chosen based on their ability to gather the desired information needed for participants, patients, or clients in research, clinical, or wellness settings. Presently, the International Physical Activity Questionnaire (IPAQ) has been the most widely used and most validated self-report PA questionnaire (van Poppel et al., 2010). Therefore, this may be considered the gold standard of PA self-report questionnaires.

However, van Poppel et al. (2010) conducted a meta-analysis of self-report PA questionnaires and found numerous limitations that surfaced in studies examining the psychometric properties of the IPAQ. Three studies included in the meta-analysis examined validity of the IPAQ long form through correlations with accelerometers. Two studies conducted in the United States showed r values ranging from -.02 to .36 and .23 to .47 between the IPAQ and accelerometers. Another study included in the meta-analysis conducted in Sweden displayed r values ranging from .12 to .63 between the IPAQ and accelerometers. Three studies included in the meta-analysis examined reliability of the IPAQ long form. Two conducted in the United States showed total intraclass correlation coefficients of .77 and .83. The third study included in the meta-analysis was conducted in Belgium and displayed a total intraclass correlation coefficient of .69.

Van Poppel and colleagues concluded that the IPAQ displayed inconsistencies among studies for its reliability and additional studies are needed to evaluate its validity because of inconsistent results when compared to objective measures of PA. More specifically, content validity of the IPAQ short form version is limited because it does not discriminate between settings, while the IPAQ long form has been explained as "too boring and repetitive" (van Poppel et al., 2010). In addition, the IPAQ is shown to be questionable in its discrimination between groups with differing PA levels measured with Doubly Labeled Water, the gold standard measure of PA (van Poppel et al., 2010). Lastly, the IPAQ utilizes the quantity-frequency method to collect information on PA, therefore only an average volume of PA rather than a specific amount is obtained.

Despite the advantages of the TLFB-E that have been outlined in this chapter, the TLFB-E has limitations. Nonetheless, this thesis has shown that the TLFB-E is reliable and valid in its ability to collect FITT when compared to multiple subjective and objective measures of PA. Thus, there are a number of ways that these data may be beneficial to those in research, clinical, and health and wellness settings when using the TLFB-E as a self-report assessment of past exercise behavior over two months.

Potential Benefits of the TLFB-E as a Self-Report Measure of Physical Activity

First, exercise patterns can be recognized on the TLFB-E by analyzing FITT daily, weekly, or monthly. This information can be used to determine an appropriate exercise prescription. The majority of retrospective PA questionnaires collect information on a general time frame only (e.g., 3 days, 1 week, or 2 months; Kriska & Casperson, 1997). This method, known as the quantity-frequency method, does not give the researcher the option to analyze patterns over differing time frames. On the TLFB-E, an individual may show a pattern of exercise of one to two (frequency) bouts per week for 30 minutes (time) at the end of the week (e.g., Thursday and Friday) of only resistance exercise (type; See Appendix A and *Timeline Followback for Exercise* description in

Chapter 3, Methods). These exercise components (FITT) give the personnel administering the TLFB-E a general idea of the individual's exercise habits. Thus, information gathered in the current example would indicate this person may need an intervention targeting exercise adherence for the beginning of the week and should most likely include cardiovascular exercise assuming an ultimate goal of meeting the ACSM's recommendations for healthy adults (Garber et al., 2011). If these exercise patterns only occur during the winter months of the TLFB-E, a seasonal effect may also be examined giving the researcher an idea of what months this individual may need additional exercise interventions.

Second, variability of exercise engagement reported on the TLFB-E can be recognized over time (i.e., ≤ 2 months). Many self-report PA questionnaires collect information over a short period of time (i.e., ≤ 1 week) such as the Bouchard Three-Day Physical Activity Record (Bouchard, Tremblay, LeBlanc, Lortie, Savard, & Theriault, 1983), Framingham Physical Activity Index (Kannel & Sorlie, 1979), KIHD Seven-Day Physical Activity Recall, KIHD 24-hour Physical Activity Record (Lakka & Sabren, 1992), Seven Day Physical Activity Recall (Sallis, Haskell, & Wood et al., 1985), Yale Physical Activity Survey (DiPietro, Casperson, Ostfeld, & Nadel, 1993), and IPAQ short form (Craig et al., 2003). Therefore, these PA questionnaires do not capture the variability of exercise that may occur over longer periods of time due to injury, changes in motivation, and other factors that affect exercise participation. The TLFB-E collects this information on a daily basis, asking the participant to include "special days", therefore capturing possible causes of the exercise variability as well as the exercise participated in.

Third, data can be analyzed in a number of ways. Variables analyzed in this thesis included: All bouts, aerobic bouts, resistance bouts, flexibility bouts, aerobic & resistance bouts, total time, Kcal expended, average MET hours per bout, and average RPE reported. However, in addition to these variables, the TLFB-E can produce variables to gather even greater amounts of information. For example, the researcher may be interested in looking at percentage of days the participant has exercised at different intensity levels and/or the pattern of weekday/weekend exercising. Examples of variables for this analysis may be percent of exercise or number of bouts reported in light exercise (< 3 METs), moderate exercise (3 – 5.9 METs), and vigorous exercise (≥ 6 METs). Any variables mentioned may also be split by week days and weekend days and may determine if an individual tends to be a "weekend warrior" who prefers to load their exercise on the weekends or prefers to exercise during the week.

Fourth, the TLFB has the ability to provide participants with feedback. Information gathered from the TLFB may be presented to participants to show current progress as well as problems that may still exist. With this information at hand, an individual who presents with serious health risks based on their behavior shown by the TLFB can be introduced to graphs and literature that exemplify evidence of probable future consequences of their current behavior. This strategy may be viewed as an "eye opener," having a positive impact on the participant's behavior and help increase motivation for change. These multiple uses of data collected on the TLFB-E are possible because of its unique layout that enables specific information to be gathered on a daily basis.

Future Research

This thesis provides evidence that the TLFB-E is a valid and reliable method to measure past exercise behavior among college students. However, the lifestyle and exercise habits of college students tend to be different than those of other populations (Behrens & Dinger, 2003). Thus, future studies should assess the psychometric properties of the TLFB-E in other populations to enhance its generalizability.

This thesis examined several types of validity of the TLFB-E including: criterion, convergent, and predictive validity. However, content validity was not examined. Content validity can be defined as the estimate of how much a measurement represents every element of a construct (Nunnally & Bernstein, 1994). Examining content validity will reinsure that the TLFB-E has been designed correctly and that the TLFB-E asks the appropriate questions to obtain the desired data (i.e., FITT). A feasible design to measure content validity of the TLFB-E would be to have an expert panel rate each question that is asked when administering the TLFB-E. The expert panel then would rate whether the question is essential, useful or irrelevant to measuring FITT of exercise (Nunnally & Bernstein, 1994). The results would then be analyzed and then the TLFB-E may be modified to improve its content.

Conclusion

The TLFB is a well validated and reliable retrospective behavioral measure for a variety of behaviors including: Alcohol and drug use (Ehrman et al., 1994; Sobell & Sobell, 1996) 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).

This thesis was the first study to examine the psychometric properties of the TLFB adapted for exercise. This thesis provides evidence that the TLFB-E is a valid and reliable self-report PA measure among college students. This research contributes to the current literature by adding a quality self-report PA questionnaire that may potentially improve upon commonly used quantity-frequency measures. However, additional research is still needed in order to consider the TLFB-E a well validated and reliable measurement of past PA. In the future, this thesis may provide exercise professionals with a cost-effective instrument to review past patterns of exercise. By obtaining this knowledge, exercise professionals will enhance their ability to prescribe appropriate interventions and exercise programs based on the client, participant, or patient's exercise patterns. Having the ability to provide more appropriate programs and interventions may lead to a decrease in sedentary behavior which may ultimately result in a more active and healthier life.

References

Aalto, M., Tuunanen, M., Sillanaukee, P., & Seppä, K. (2006). Effectiveness of structured questionnaires for screening heavy drinking in middle-aged women. *Alcoholism: Clinical and Experimental Research, 30*(11), 1884-1888.

Ainsworth B.E., Haskell W.L., Herrmann S.D., Meckes N., Bassett Jr. D.R., Tudor-Locke C., Greer J.L., Vezina J, Whitt-Glover M.C., Leon A.S. (2011). Compendium of Physical Activities: a second update of codes and MET values. *Medicine and Science in Sports and Exercise*, 43(8):1575-1581.

Bassett Jr., D. R., Cureton, A.L. & Ainsworth, B.E. (2000). Measurement of daily walking distance-questionnaire versus pedometer. *Medicine & Science in Sports & Exercise, 32*(5), 1018-1023.

Behrens, T., & Dinger, M. (2003). A preliminary investigation of college students' physical activity patterns. *American Journal of Health Studies, 18*(2/3), 169-172.

Bigaard J, Frederiksen K, Tjonneland A, et al. (2005). Waist circumference and body composition in relation to all-cause mortality in middle-aged men and women. *International Journal of Obesity*, 29(7):778–84.

Bouchard, C., Tremblay, A., LeBlanc, C., Lortie, G., Savard, R., Theriault, G. (1983). A method to assess energy expenditure in children and adults. *American Journal of Clinical Nutrition,* 37, 461-467.

Brandon, L. J., & Elliott-Lloyd, M. B. (2006). Walking, body composition, and blood pressure dose-response in african american and white women. *Ethnicity and Disease, 16*(3), 675.

Brown, R. A., Burgess, E. S., Sales, S. D., Whiteley, J. A., Evans, D. M., & Miller, I. W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviors, 12*(2), 101-112.

Carlbring, P., Jonsson, J., Josephson, H., & Forsberg, L. (2009). Motivational interviewing versus cognitive behavioral group therapy in the treatment of problem and pathological gambling: A randomized controlled trial. *Cognitive Behaviour Therapy,* 39, 92-103.

Collins, S. E., Eck, S., Torchalla, I., Schröter, M., & Batra, A. (2009). Validity of the timeline followback among treatment-seeking smokers in germany. *Drug and Alcohol Dependence,* 105 (2), 164-167.

DiPietro, L., Casperson, C.J., Ostfeld, A.M. Nadel, E.R. (1993). A survey for assessing physical activity among older adults. *Medicine & Science in Sports & Exercise,* 25, 628-642.

Ehrman, R. N., & Robbins, S. J. (1994). Reliability and validity of 6-month timeline reports of cocaine and heroin use in a methadone population. *Journal of Consulting and Clinical Psychology, 62*(4), 843-850.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2003). The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence, 18*(3), 131-142.

Garber, C. E., Bryan Blissmer, Michael R. Deschenes, Barry A. Franklin, Michael J. Lamonte, I-Min, . . . David P. Swain. (2011). 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. 43*(7), 1334-1359.

Hamilton, G. F., McDonald, C., & Chenier, T. C. (1992). Measurement of grip strength: Validity and reliability of the sphygmomanometer and jamar grip dynamometer. *The Journal of Orthopedic and Sports Physical Therapy, 16*(5), 215-219.

Hodgins, D. C., & Makarchuk, K. (2003). Trusting problem gamblers: Reliability and validity of self-reported gambling behavior. *Psychology of Addictive Behaviors, 17*(3), 244-248.

Kannel, W.B., Sorlie, P. (1979). Some health benefits of physical activity: The Framingham Heart Study. *Archives of Internal Medicine,* 139, 857-861.

King, A.C., Haskell, W.L., Taylor, C.B., Kraemer, H.C., DeBusk, R.F. (1991). Group-vs home-based exercise training in healthy older men and women. A community – based clinical trial. *The Journal of the American Medical Association,* 266(11), 1535-1542.

Kriska, A., & Caspersen, C. (1997). A collection of physical activity questionnaires for health-related research. *Medicine and Science in Sports and Exercise, 29*(6), S203-5.

Krustrup, P., Hansen, P., Randers, M. B., Nybo, L., Martone, D., Andersen, L. J., . . . Bangsbo, J. (2010). Beneficial effects of recreational football on the cardiovascular risk profile in untrained premenopausal women. *Scandinavian Journal of Medicine & Science in Sports, 20*, 40-49.

Lakka, T.A., Salonen, J.T. (1992). Intra-person variability of various physical activity assessments in the Kuopio ischaemic heart disease risk factor study. *International Journal of Epidemiology,* 21, 467-472.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159.

Nelson, W. A., & Clum, G. A. (2002). Assessment of panic frequency: Reliability and validity of a time-line follow-back method. *Journal of Psychopathology and Behavioral Assessment, 24*(1), 47-54.

Nunnally, J. C., & Bernstein, I. H. Psychometric Theory (1994). New York, NY: McCraw-Hill.

Peterson, M.D., Rhea, M.R., Alvar, B.A. (2005). Applications of the Dose-Response for Muscular Strength Development: A review of Meta-Analytic Efficacy and Reliability for Designing Training Prescription. *Journal of Strength and Conditioning Research,* 19(4).

Poldermans, D., Fioretti, P., Forster, T., Thomson, I., Boersma, E., El-Said, E., . . . van Urk, H. (1993). Dobutamine stress echocardiography for assessment of perioperative cardiac risk in patients undergoing major vascular surgery. *Circulation, 87*(5), 1506.

Rantanen, T., Guralnik, J. M., Foley, D., Masaki, K., Leveille, S., Curb, J. D., & White, L. (1999). Midlife hand grip strength as a predictor of old age disability. *The Journal of the American Medical Association, 281*(6), 558.

Sallis, J.F., Haskell, W., Wood, P. et al. (1985). Physical activity assessment methodology in the Five-City Project. *American Journal of Epidemiology*, 121, 91-106.

Sallis, J. F., Saelens, B. E. (2000). Assessment of physical activity by self-report: Status, limitations, and future directions. *Research Quarterly for Exercise and Sport.* 71(2), S1-14.

Sobell, L. C., & Sobell, M. B. (1996). 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.

Thompson, W.R., PhD, FACSM, Gorden, N.F., PhD, FACSM, Pescatello, L.S., (2009). *ACSM's guidelines for exercise testing and prescription eighth edition*.

van Poppel, M. N. M., Chinapaw, M. J. M., Mokkink, L. B., van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for adults: A systematic review of measurement properties. *Sports Medicine, 40*(7), 565-600.

Weinhardt, L. S., Carey, M. P., Maisto, S. A., Carey, K. B., Cohen, M. M., & Wickramasinghe, S. M. (1998). Reliability of the timeline follow-back sexual behavior interview. *Annals of Behavioral Medicine, 20*(1), 25-30.

Weinstock, J., Whelan, J. P., & Meyers, A. W. (2004). Behavioral assessment of gambling: An application of the timeline followback method. *Psychological Assessment, 16*(1), 72-79.

Appendix A - Example of the Timeline Followback for Exercise (TLFB-E)

January 2010

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Date	Notes:					1	2
						92	93
Exercise						Y N	Y N
Type							
Duration(time)						mins	mins
RPE (6-20)							
Special Day?						New Years Day	
	Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
Date	3	4	5	6	7	8	9
	94	95	96	97	98	99	100
Exercise	Y N	Y N	Y N	Y N	Y N	Y N	Y N
Type							
Duration(time)	mins	mins	mins	mins	mins	mins	mins
RPE (6-20)							
Special Day?			Cousin's Birthday				
	Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
Date	10	11	12	13	14	15	16
	101	102	103	104	105	106	107
Exercise	Y N	Y N	Y N	Y N	Y N	Y N	Y N
Type	Treadmill						
Duration(time)	30 mins	mins	mins	mins	mins	mins	mins
RPE (6-20)	12						
Special Day?				←---------------- On Vacation in North Carolina -------------→			
	Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
Date	17	18	19	20	21	22	23
	108	109	110	111	112	113	114
Exercise	Y N	Y N	Y N	Y N	Y N	Y N	Y N
Type							
Duration(time)	mins	mins	mins	mins	mins	mins	mins
RPE (6-20)							
Special Day?							
	Sunday	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**
Date	24	25	26	27	28	29	30
	115	116	117	118	119	120	121
Exercise	Y N	Y N	Y N	Y N	Y N	Y N	Y N
Type							
Duration(time)	mins	mins	mins	mins	mins	mins	mins
RPE (6-20)							
Special Day?							
	Sunday	Notes:					
Date	31						
	122						
Exercise	Y N						
Type							
Duration(time)	mins						
RPE (6-20)							
Special Day?							

Note. RPE = Rating of Perceived Exertion

In this example of the TLFB-E, the month of January is being used. The first column of each row includes a list of the variables the TLFB-E will obtain. “Date” will be determined by filling out information below the day in which the physical activity was completed, and “Exercise” will be determined with a circle around either “Y” representing “yes” or “N” representing “no” exercise done on that day. The blank area next to “Type” is where the Research Assistant will record the physical activity completed (i.e., running, stationary bicycle). “Duration” of the physical activity will be recorded in minutes and Rating of Perceived Exertion (RPE) recorded using the Borg 6-20 Scale (Borg, 1998).

January 10^{th} is an example of a day that exercise was reported by the participant. January 11^{th} is an example of a day that exercise was not reported.

Appendix B – Question four of the College Alumni Questionnaire

List any sports or recreational activities you have actively participated in during the past 2 months. Please remember seasonal sports or events.

Enter the average number of times per week you took part in these activities and the average duration of these sessions. Include only time you were physically active (that is, actual playing or activity time).

		Average Times per Episode	
Sport or Recreation	Times per 2 Months	Hours	Minutes
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______
___________________	____________	______	______

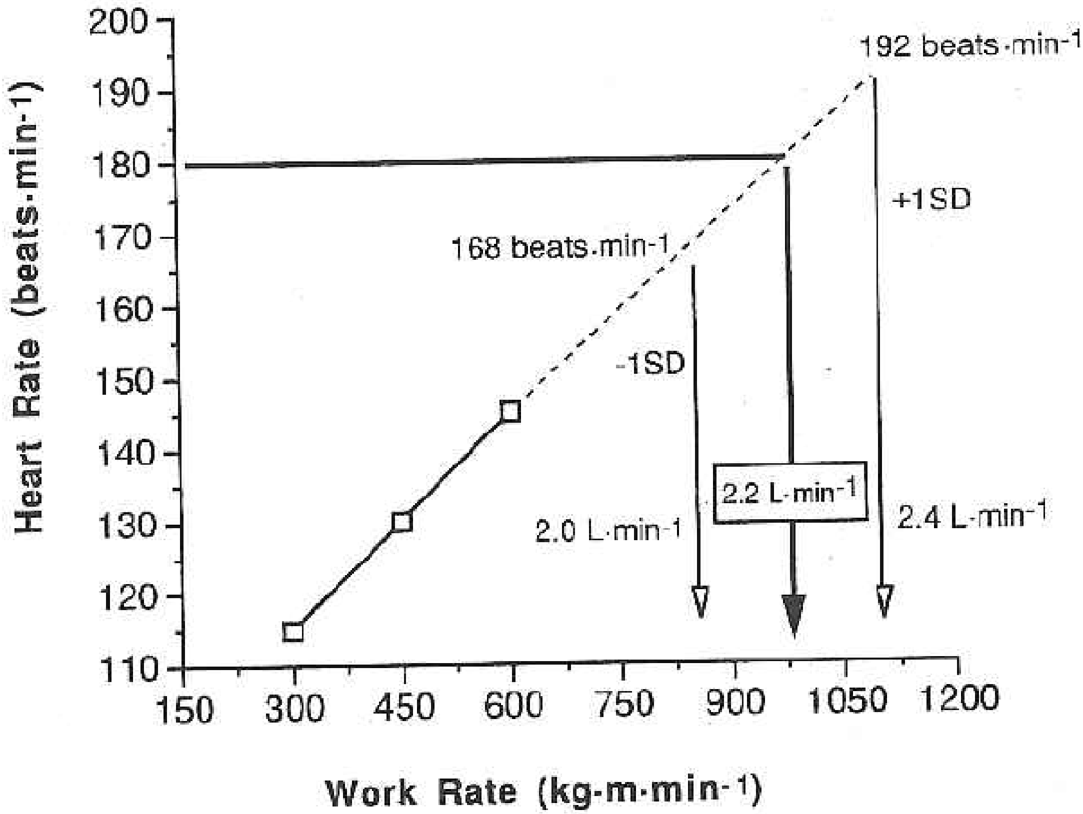
Appendix C - Example of a Physical Activity Contract

Physical Activity	Date & Time	Potential Problems	Verification	RPE	Done
Group Fitness Class: Spinning (60 min.)	Mon. @ 7pm	Class is Full	Picture of sign-in sheet	15	√
Play Basketball (60 min.)	Wed. @ 1pm	No courts open	Cell phone video of playing	14	√
Walk 10,000 Steps	Thurs	Raining	Pedometer	12	√
(Alternate Activity): Resistance Training: 2 sets of 8-12 reps 1.Chest Press Machine 2.Shoulder Press Machine 3. Leg Press 4.Seated Row	Sat. @ 2pm	Machines are taken	Cell phone video of one set per exercise	Ø	Ø

In the first column, three activities and one alternate activity are listed. An alternate activity may be completed if the subject is unable to complete one of the other contracted activities. Duration and intensity will be agreed upon by both the participant and Research Assistant. The second and third columns display date and time for these activities to be completed as well as any potential problems that may arise preventing the subjects from completing the activities. Verification for completion of each activity will also be documented. When reviewing a previous week’s contract, researchers will ask to see verification for each activity completed (e.g., cell phone video/photo, pedometer). If the participant fails to present adequate verification based on Research Assistant’s judgment, the participant will not receive credit for completing the activity. Researchers will also document Rating of Perceived Exertion for each activity completed based on the participant’s report.

Appendix D

Graphing of HR Response for Prediction of Maximal Aerobic Capacity



Appendix D displays an example from ACSM's Health Related Physical Fitness Assessment Manual 2nd edition (Dwyer, Davis, Pire, & Thompson, 2007) of submaximal cycle ergometer data being used to estimate maximal aerobic capacity. This example shows 3 submaximal work rates for a 40-year old, sedentary female weighing 64 kg.

A line is drawn connecting the three HR steady states (115bpm, 130bpm, 145bpm) and work output points (300kgm, 450kgm, 600kgm). The line is then extrapolated up to the age-predicted maximal HR of 180bpm (220 – age) and a perpendicular line is dropped down to work output (x-axis) to estimate the peak work rate (2.2 L·min) the subject would have achieved had they worked until peak capacity. This value was used as a determinant maximal oxygen uptake expressed in liters per minute (Dwyer et al., 2007).

References

Borg, G. (1998). *Borg's perceived exertion and pain scales.* Champaign, IL: Human Kinetics Publishers.

Dwyer, G.B., Davis, S.E. (2008). *ACSM's health-related physical fitness assessment manual 2nd edition.* Philadelphia, PA: Lippincott, Williams, & Wilkins.