Evaluation of the Feasibility of a Two-Method Measurement Design for the Assessment of Healthy Physical Activity Behavior in Youth

Amelia A. Miramonti
University of Nebraska - Lincoln, amelia.miramonti@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/cehsdiss
Part of the Other Education Commons

http://digitalcommons.unl.edu/cehsdiss/327

This Article is brought to you for free and open access by the Education and Human Sciences, College of (CEHS) at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Public Access Theses and Dissertations from the College of Education and Human Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
EVALUATION OF THE FEASIBILITY OF A TWO-METHOD MEASUREMENT DESIGN FOR THE ASSESSMENT OF HEALTHY PHYSICAL ACTIVITY BEHAVIOR IN YOUTH

by

Amelia A. Miramonti

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillments of Requirements
For the Degree of Master of Arts

Major: Educational Psychology

Under the Supervision of Professor James A. Bovaird

Lincoln, Nebraska

December, 2018
EVALUATION OF THE FEASIBILITY OF A TWO-METHOD MEASUREMENT DESIGN FOR THE ASSESSMENT OF HEALTHY PHYSICAL ACTIVITY BEHAVIOR IN YOUTH

Amelia A. Miramonti, M.A.
University of Nebraska, 2018

Advisor: James A. Bovaird

Purpose: Assess the reliability and validity of self- and parent-report survey responses regarding physical activity (PA), sedentary behaviors, and PA self-efficacy and determine if these data can be combined with objective physical activity monitor data to model the latent construct healthy physical activity behavior (HPAB).

Methods: 126 underserved 4th-5th grade students participated in a 12-week after-school nutrition, cooking, and physical activity program (WeCook: Fun with Food and Fitness). Participants and parents (n=103) completed surveys pre- and post-program and participants wore PA monitors for one week at PRE and POST. Unidimensionality and internal consistency reliability were assessed for survey measures and objective PA measures (ST=step counts, FL=floors climbed) and predictive validity of survey measures was assessed through correlation with ST and FL.

HPAB was modeled using z-scores (standardized across time, averaged) for youth self-reported physical activity (YPA), youth PA self-efficacy (YSE), parent-reported youth PA and sedentary behavior (AS), ST, and FL. Metric and intercept invariance were
established across time and between groups selected for assessing construct validity (gender, grade, socioeconomic status, weight status, school, minority, season, grant year).

Results: ST, FL, YSE, and AS were unidimensional. For ST coefficient $\alpha$ was 0.735 (PRE) and 0.805 (POST), for FL $\alpha$ was 0.686 and 0.684 (PRE and POST), for YSE, $\alpha$ was 0.829 (PRE) and 0.897 (POST), and for AS $\alpha$ was 0.545 (PRE) and 0.729 (POST). YPA was most predictive of the objective PA measures and was correlated with ST at PRE and POST and FL at PRE ($p<0.05$). YSE was predictive of ST at POST ($p<0.05$), but no other objective PA measures. AS was not predictive of any objective PA measures ($p>0.05$). HPAB exhibited measurement invariance across time and between groups of interest and some evidence for latent construct validity based on nomothetic span and construct representation was established.

Conclusions: This study establishes some evidence supporting the feasibility of modeling HPAB using survey and objective measures of PA in youth.
Copyright 2018, Amelia A. Miramonti
DEDICATION

To my family, and those friends who may as well be.
AUTHOR’S ACKNOWLEDGMENTS

First, I would like to thank my entire committee for taking me on under somewhat unusual circumstances, and for their flexibility and understanding throughout.

Specifically, I would like to thank Dr. Lisa Franzen-Castle and Dr. Michelle Krehbiel for allowing me the opportunity to work on the WeCook grant, and the latitude afforded me in managing and analyzing the data. In addition, I would like to thank Tara Dunker for her help and encouragement.

Finally, I would like to thank Dr. Jim Bovaird for his guidance through this process. I know I had many questions, so thank you for your patience! I cannot express how grateful I am.
FUNDING SOURCE

WeCook: Fun with Food and Fitness is supported by the Child, Youth, and Families at Risk grant funded through the National Institute of Food and Agriculture, U.S. Department of Agriculture award number 2014 – 41820 – 22207.
TABLE OF CONTENTS

Dedication ................................................................................................................................. 5
Author’s Acknowledgments .................................................................................................... 6
Funding Source .......................................................................................................................... 7
List of Figures ........................................................................................................................... 12
List of Tables ............................................................................................................................ 13
Chapter I: Introduction .......................................................................................................... 1
Chapter II: Literature Review ................................................................................................. 6
  Physical Activity Behaviors, Weight Status, and Health Across the Lifespan ................. 6
  Limited Evidence and Inadequate Evaluative Measures .................................................. 7
  Manifest and Latent Variables ............................................................................................ 12
  Structural Equation Modeling and the Two-Method Measurement Design ............... 14
  Reliability and Validity ....................................................................................................... 17
    Convergent and discriminant validity and construct representation ..................... 19
  Present Study ....................................................................................................................... 21
Chapter III: Methodology ..................................................................................................... 22
  Study Design ....................................................................................................................... 22
  Participants and Recruitment ............................................................................................... 22
  Program Curriculum ........................................................................................................... 23
  Data Collection Procedures ............................................................................................... 24
Survey Instruments ............................................................................................................. 24
Youth survey ......................................................................................................................... 24
Adult survey .......................................................................................................................... 25
Physical Activity Monitors ................................................................................................. 25
Data Processing ...................................................................................................................... 26
Statistical Analyses .............................................................................................................. 28
Unidimensionality ............................................................................................................... 28
Internal consistency reliability ............................................................................................ 28
Validity of survey responses ............................................................................................... 29
Latent healthy physical activity behavior ............................................................................ 29
Longitudinal measurement invariance ............................................................................... 29
Measurement invariance between groups .......................................................................... 30
Nomological network and contextual validity .................................................................. 30
Chapter IV: Results .............................................................................................................. 32
Unidimensionality ............................................................................................................... 33
Step counts and floors climbed .......................................................................................... 33
Youth physical activity self-efficacy items .......................................................................... 34
Adult survey items regarding child physical activity and sedentary habits ....................... 35
Reliability and Validity of Survey Measures ....................................................................... 35
Conclusions .............................................................................................................................................. 56
References .................................................................................................................................................. 57
Appendix A: Youth Survey ......................................................................................................................... 72
Appendix B: Adult Survey ......................................................................................................................... 77
LIST OF FIGURES

Figure 1. .................................................................................................................. 33
Figure 2. .................................................................................................................. 34
Figure 3. .................................................................................................................. 35
Figure 4. .................................................................................................................. 39
Figure 5. .................................................................................................................. 43
LIST OF TABLES

Table 1. .......................................................................................................................... 32
Table 2 ................................................................................................................................ 32
Table 3. ................................................................................................................................ 37
Table 4 ................................................................................................................................ 37
Table 5 ................................................................................................................................ 37
Table 6. ................................................................................................................................ 38
Table 7. ................................................................................................................................ 40
Table 8. ................................................................................................................................ 41
Table 9. ................................................................................................................................ 41
Table 10 .................................................................................................................................. 42
CHAPTER I: INTRODUCTION

The prevalence of overweight and obesity is a widely-acknowledged problem in the United States (US) and other developed countries (Pozza & Isidori, 2018; Ravussin & Ryan, 2018). Although significant efforts have been made to address the obesity epidemic in the US, evidence suggests that the current public health approach has shown meager success thus far (Ludwig, 2018). Overweight and obesity rates in the US have risen dramatically in recent decades among all age groups (Ogden, Fryar, Carroll, & Flegal, 2004; Ogden, Carroll, Kit, & Flegal, 2014; Ogden et al., 2006), but the increases in childhood obesity are particularly concerning, as obese children are more likely to remain obese in adulthood and are at greater risk of early onset of diseases such as Type II diabetes and heart disease (Cote, Harris, Panagiotopoulos, Sandor, & Devlin, 2013; Drenowatz et al., 2010; Hannon, Rao, & Arslanian, 2005; Singh, Mulder, Twisk, Van Mechelen, & Chinapaw, 2008). Moreover, children from low-income and minority families are at even greater risk of overweight and obesity than their white or Caucasian, higher socioeconomic status (SES) counterparts (Fahlman, Hall, & Gutuskey, 2015; Mulasi-Pokhriyal & Smith, 2010; Pan, 2016; Smith & Franzen-Castle, 2012).

Consequently, many health, nutrition, and obesity-prevention programs have been created to help at-risk children (Dobbins, DeCorby, Robeson, Husson, & Tirilis, 2009; Salmon, Booth, Phongsavan, Murphy, & Timperio, 2007; van Sluijs, McMinn, & Griffin, 2007). Obesity is a complex process involving both non-modifiable risk factors (e.g., genetics) and modifiable ones, (e.g., dietary patterns and physical activity) (Kipping, Jago, & Lawlor, 2008), which exacerbates the difficulties faced by public health advocates attempting to curb the growing trend. Despite the number of programs
implemented, their ability to effect behavioral changes is unclear, in part because of the broad range of program types, durations, and varied implementations and audiences. Therefore, there is a need for high-quality evaluative methods for assessing the effectiveness of current programs. More robust assessments would help guide future program development or help identify ineffective ones that should be modified or discontinued.

Obesity prevention programs generally target either dietary patterns and nutrition (energy intake) or physical activity and sedentary behaviors (energy expenditure), or a combination. Unfortunately, it is difficult to accurately assess free-living dietary intake, PA, and sedentary time in a cost-effective manner, and self-report measures are notoriously imprecise (Dyrstad, Hansen, Holme, & Anderssen, 2014). Despite their limitations, self-report measures are ubiquitous in program evaluation and research settings because they are inexpensive and easy to implement. For some programs, standardized versions (e.g., the Children, Youth, and Families at Risk [CYFAR] 4-H Common Measures) are mandated to provide consistent data collected nationally (University of Minnesota & USDA National Institute of Food and Agriculture, 2018). However, some of the measures used may not provide data that can be used as reliable and valid indicators of the constructs the measures are intended to assess, or simply may not be sensitive enough to detect changes after a behavioral intervention.

Fortunately, technological advances have decreased the costs and improved the accessibility of consumer-grade PA monitors, (Franzen-Castle, Dunker, Chai, & Krehbiel, 2017). Although the advent of lower-cost wearable physical activity monitors
coupled with software innovations has improved the feasibility of objective physical activity data collection for research and evaluation (Franzen-Castle et al., 2017), it may still be cost-prohibitive or too time-consuming for many researchers and program leaders. Therefore, there is a need to evaluate the association between objectively-measured physical activity data and self- and parent-reported responses to standard survey items regarding physical activity levels in youth in health-promotion and obesity-prevention programs.

Among at-risk youth, factors thought to influence obesity rates include poor eating habits, a lack of nutrition knowledge and food preparation skills, high levels of sedentary time, and low levels of physical activity (Heath, 2018; Walther, Dunker, Franzen-Castle, & Krehbiel, 2018). The WeCook: Fun with Food and Fitness (WeCook) afterschool program is funded by United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) CYFAR Grant Program and provided through Nebraska 4-H in cooperation with Community Learning Centers (CLCs) at two Title I elementary schools in a Midwestern city. The WeCook program aims to address some of these issues by teaching at-risk 4th and 5th grade youth how to make healthy food choices, prepare healthy foods, and increase their physical activity. Therefore, WeCook is a multi-faceted effort aimed at improving healthy nutrition and physical activity behaviors, which should theoretically contribute to decreased risk of overweight and obesity and chronic disease among participants.

In children, overweight and obesity are defined as having a body mass index (BMI, weight / height², [kg/ m²]) above the 85th and 95th percentiles for the child’s age.
and gender, respectively (Kuczmarski et al., 2002; Ogden & Flegal, 2010). However, weight change is gradual, and both height and weight are influenced by growth and development, so it may be unrealistic to expect measurable improvements in BMI or weight status within the time frame of the program (12 weeks). Yet, if WeCook successfully alters overall physical activity (HPAB) and nutrition habits, these behavioral changes may have a cumulative effect over much longer periods than program duration (Lawman & Wilson, 2012; Salmon et al., 2007). Therefore, it may be more informative to model healthy physical activity behavior as a latent construct measured by both subjective and objective measures of physical activity. If these disparate measures can reliably contribute unique information about physical activity behavior among youth, it may be a useful strategy for program evaluators to assess lifestyle changes with lasting benefits instead of only looking for changes in specific outcome variables (e.g., BMI).

Although the program encompasses both dietary and physical activity components, the present study is limited to the assessment of physical activity and related items, such as self-efficacy for physical activity. Therefore, the purposes of this study were to (a) assess the reliability and validity of the youth self-report and parent-report data regarding physical activity and sedentary behavior in youth participating in the WeCook afterschool program, and (b) determine whether the data from these measures, combined with free-living objective wearable physical activity monitor data, can be used to generate a model of the latent construct HPAB such that it might be a useful outcome measure for program evaluation.
The results of this work may demonstrate the reliability and validity of the youth self- and parent-report physical activity and sedentary behavior survey responses, supporting the continued usage of the instruments they came from. Alternatively, the results may indicate there is need for more reliable, valid, or sensitive measures for assessing potential physical activity behavior changes in this population. In addition, the latent HPAB construct modeled will be assessed for construct validity, which may support its use as an outcome measure when evaluating the WeCook program. If successful, these methods may also be adapted as an evaluative tool for similar programs in the future.
CHAPTER II: LITERATURE REVIEW

Physical Activity Behaviors, Weight Status, and Health Across the Lifespan

Referred to as an epidemic (Ng et al., 2014; Pozza & Isidori, 2018; Ravussin & Ryan, 2018), the growing prevalence of overweight and obesity has been observed in the US since the late 1980’s (Flegal, Carroll, Kuczmarski, & Johnson, 1998) and has been reported in 30 countries (Ng et al., 2014; Ravussin & Ryan, 2018). Obesity is associated with increases in chronic disease risks and healthcare costs, reduced quality of life, and decreases in productivity (Brownell, 1986; Finkelstein, Trogdon, Cohen, & Dietz, 2009; Luppino et al., 2010; Y. C. Wang, McPherson, Marsh, Gortmaker, & Brown, 2011; G. Wang & Dietz, 2002; Withrow &Alter, 2011). This growing trend affects people across racial, ethnic, and socioeconomic lines, but overweight and obesity and associated chronic diseases occur at elevated rates among minority and low socioeconomic status (SES) populations, a pattern which holds true for youth in those populations (Delva, Johnston, & O’Malley, 2007). Similarly, minority and low SES youth are more likely to exhibit less healthy dietary and exercise habits (Delva, O’Malley, & Johnston, 2006; Drenowatz et al., 2010) and have lower physical fitness (Fahlman et al., 2015), increasing their risk for overweight, obesity, and chronic diseases above and beyond any potential influence of genetics.

Fahlman and colleagues (2015) argue that behaviors established in childhood have cumulative effects across the lifespan and that interventions to influence behavioral changes in these high-risk populations are needed. Janz and colleagues (2000) studied the physical activity behaviors and physical fitness of children and adolescents over a 5-year period and concluded that programs aimed at maintaining physical fitness and physical
levels through puberty would have favorable health benefits over time. A follow-up study to the Harvard Growth Study published in 1992 demonstrated that overweight in adolescence predicted health risks independent of weight status in adulthood (Must, Jacques, Dallal, Bajema, & Dietz, 1992). Ruiz and colleagues (2009) conducted a systematic review of the predictive validity of health-related fitness in youth. The authors found strong evidence indicating that cardiorespiratory fitness in childhood and adolescence are associated with a healthier cardiovascular profile later in life, muscular strength improvements from childhood to adolescence are negatively associated with changes in overall adiposity. In addition, a healthier body composition in childhood and adolescence was associated with a healthier cardiovascular profile later in life and with a lower risk of death. Collectively, these studies emphasize the importance of interventions to improve the physical activity habits and weight status in youth, particularly those at high-risk of overweight and obesity.

**Limited Evidence and Inadequate Evaluative Measures**

Overall, only 42% of children aged 6-11 years meet the physical activity guidelines of at least 60 minutes of moderate-to-vigorous physical activity per day (Troiano et al., 2008). In regions reporting middle school data for the 2017 Youth Risk Behaviors Survey the proportion of students meeting the physical activity guidelines ranged from 22.8% in Boston, MA to 43.8% in Wyoming (Centers for Disease Control and Prevention (CDC), 2018). Although a large number and broad variety of behavioral interventions aimed at increasing physical activity have been implemented, many such programs have had small or negligible effects on the intended outcomes (Dobbins et al.,
 evaluated (Salmon et al., 2007; van Sluijs et al., 2007). Salmon and colleagues (2007) emphasized that, in the context of the growing obesity trend among children and adolescents, “there has never been a more urgent need for effective physical activity programs.” Unfortunately, the review by van Sluijs and colleagues (2007) found limited evidence of beneficial effects from physical activity interventions targeting minority children, girls, and those in low SES groups. However, there were few high-quality trials in these populations included in the review, highlighting the need for further research.

In particular, there is a need for more informative evaluation tools to assess program effectiveness. Many programs rely heavily, if not entirely, on self-report measures, because of their ease of use and low cost of implementation (Troiano, Gabriel, Welk, Owen, & Sternfeld, 2012). Unfortunately, these types of assessments often provide data that are unreliable or invalid indicators of the intended constructs, particularly in children (Chinapaw, Mokkink, van Poppel, van Mechelen, & Terwee, 2010; Patterson, 2000; Troiano et al., 2012; Zelener & Schneider, 2016). A systematic review of physical activity questionnaires in youth determined that none of the 64 questionnaires assessed in the studies reviewed met acceptable standards for both reliability and validity, although construct validity was higher among adolescents compared to younger children (Chinapaw et al., 2010).

The incongruity between self-reported and objectively-measured physical activity in children may be due to several factors. Studies in adults have shown that the disagreement between self-reported physical activity and accelerometry-based measures increases as exercise intensity and duration increase (Dyrstad et al., 2014), and this effect
could be similar or even greater in children. Janz and colleagues (2008) suggested that responses to physical activity questionnaires may be more reliable in older children because of age-related increases in reasoning skills. Similarly, Chinapaw and colleagues (2010) argued that because of developmental differences children may be less able to think abstractly about physical activity and perform detailed recall. In addition, the authors suggested that children’s physical activity patterns may be more intermittent than those of adults, increasing the difficulty of accurate recall. Zelener and Schneider (2016) also reported low accuracy for self-reported physical activity in adolescents and noted that many of the respondents failed to follow instructions or were double-counting physical activity in the moderate and vigorous categories. Walther and colleagues (2018) suggested students with learning disabilities, limited literacy, or who are English Language Learner (ELL) students may be more likely to misinterpret items on a survey and that objective physical activity measurements may be a way to detect change that might be obscured by error on a survey.

Because of some of the issues in assessing children’s physical activity levels through questionnaires, some studies have asked parents questions about their child’s physical activity levels. On one hand, adults may be better able to cognitively process the questions and provide information about the child’s physical activity habits and sedentary time. On the other hand, the parent is not always in the company of the child and may be unaware of the child’s activity levels at school or when the parent is at work. Corder and colleagues (2012) demonstrated that parent estimations of their child’s time spent being physically active are generally higher than when assessed using a wearable physical
activity monitor. Interestingly, Kesten and colleagues (2015) reported that while parents of children who were meeting the physical activity guidelines accurately perceived their child’s physical activity, but that the majority of parents of children not meeting the physical activity guidelines overestimated the child’s physical activity.

Compared to self-report measures of physical activity, many consider objective assessments preferable when feasible. While objective physical activity assessment using wearable physical activity monitors is not considered a true “gold standard” for physical activity, they have been considered one of the better tools for assessing free-living physical activity frequency and duration for research purposes (Chinapaw et al., 2010).

Historically, the research-grade devices have been prohibitively expensive for many researchers and most program evaluators. Gusmer and colleagues (2014) suggest that although physical activity monitoring is important for assessing population-level physical activity, there is a need for more affordable options for monitoring physical activity in the general population. Fortunately, recent technological advances and an increased consumer interest in fitness and activity tracking have catalyzed a proliferation of relatively inexpensive commercial wearable physical activity monitors (Franzen-Castle et al., 2017; Ridgers, McNarry, & Mackintosh, 2016). However, Gusmer and colleagues (2014) compared step counts between a consumer-grade (Fitbit® Ultra) and a research-grade physical activity monitor (ActiGraph™ GT1M), and energy expenditure estimates (kilocalories) from the two devices with energy expenditure measured via indirect calorimetry. The authors concluded that step counts from the Fitbit® Ultra and the ActiGraph™ GT1M may be used interchangeably but that energy expenditure estimates
were less accurate. Moreover, Fitbits do not provide any indication of wear time, and unless participants are required to keep detailed logs, there is no clear way to differentiate periods of little-to-no activity during which the participant was sedentary versus when the participant was not wearing it (Evenson, Goto, & Furberg, 2015). Although wearable physical activity monitors have several advantages over self-report measures, they are not without error.

In addition to self-reported physical activity, some studies have assessed self-reported physical activity self-efficacy. For many minority and low-income children, access to safe, affordable opportunities for physical activity may be limited. In some communities it may not be safe to walk to school or play outside, or there may be no public recreation spaces or programs nearby; private dance or sport programs may be cost-prohibitive, and many schools have limited extracurricular programs (Fahlman et al., 2015; Walther et al., 2018). Consequently, self-efficacy related to physical activity has been shown to be a contributing factor to physical activity levels in minority youth (Fahlman et al., 2015). Hausenblas and colleagues (2002) investigated determinants of physical activity in middle school students and noted higher self-efficacy for leisure time physical activity among students already exercising in other settings. In inner-city Hispanic American and African American children, physical activity self-efficacy was found to be a significant predictor of physical activity (Martin & McCaughtry, 2008a; Martin & McCaughtry, 2008b). Martin and colleagues (Martin, McCaughtry, Flory, Murphy, & Wisdom, 2011) also found that barrier self-efficacy was a significant predictor of physical activity in underserved (non-white, low-income) middle school
students. Trost and colleagues (2002) found that self-efficacy was a significant predictor of both intentions for physical activity and moderate-to-vigorous physical activity in white and African American 8th-grade girls.

**Manifest and Latent Variables**

Based on a review of the literature, the measurement of overall physical activity and sedentary habits is problematic. The WeCook program evaluation methods included survey measures of youth self-reported physical activity and physical activity self-efficacy, parent-report measures of youth physical activity and sedentary habits, and objectively-measured physical activity from a wearable physical activity monitor, each of which have strengths and limitations. However, data obtained from each of these assessments are likely caused (at least in part) by the general physical activity and sedentary behavior of the students and consequently, likely provide some information about the true physical activity behavior of the youth. This situation is exemplary of a more general one in which we wish to measure some cause or abstract concept that cannot be directly observed, but we can measure outcomes resulting from that unobserved cause or concept. This issue is common in social and psychological research (e.g., investigators assess constructs such as intelligence or depression) but is also seen across other fields (Bollen, 2002). Bollen states, “The idea that observable phenomena are influenced by underlying and unobserved causes is at least as old as religion…” (Bollen, 2002). In social and psychological research, the unobserved cause is referred to as a latent variable or latent construct, while the directly observed measures are referred to as manifest variables, observed variables, or indicators (Bollen, 2002). In the present
study, all of the aforementioned measures may be indicators of the latent construct of healthy physical activity behavior (HPAB).

In particular, one measurement method developed in psychological research that is relevant to the current study is the *two-method measurement design*, a type of planned missing data design (Graham, Taylor, Olchowski, & Cumsille, 2006). In this design there are two types of measures: “relatively cheap, noisy (e.g., less valid) measures of a construct and expensive, more valid measures of a construct” (Graham et al., 2006). With regard to the present study, the less valid measures would be the youth self- and parent-report survey responses, and the more valid measures would be the objective wearable physical activity monitor data. Graham and colleagues (2006) state that the benefits of using such a combination of cheap and expensive measures results in greater statistical power than with the less valid, inexpensive measures alone or the more valid, more expensive measures alone. This is achieved through improved modeling of potential response bias in the survey responses using information from the objective physical activity data. To our knowledge, no study has utilized this approach to assess physical activity behaviors in youth participating in a physical activity and nutrition intervention. Yet, Graham and colleagues specifically mention exercise and nutrition research as potential areas for application of the method (Graham et al., 2006). This model has some features in common with the multitrait-multimethod design (Campbell & Fiske, 1959; Graham & Collins, 1991; Kenny & Kashy, 1992). Specifically, the model allows for two sources of correlation among the cheap measures: the ‘real’, construct-related association and the method-related association or response bias. The benefits of this method include
smaller standard errors and larger effective Ns (i.e., increases in power that approximate what would be achieved with a larger sample), particularly in studies with small sample sizes and when the effect size is small (Graham et al., 2006). However, the authors note that these benefits are more likely to be realized when using structural equation modeling (SEM) than with other “more standard statistical methods” (Graham et al., 2006).

Consequently, this study uses an SEM approach.

**Structural Equation Modeling and the Two-Method Measurement Design**

Structural equation modeling (SEM) can be used to model a latent construct through the use of multiple indicators, reducing the effects of unreliability (Little, Bovaird, & Slegers, 2006). In addition, longitudinal SEM allows errors to be correlated, an assumption often violated in other longitudinal analysis techniques, and can be used to evaluate measurement invariance rather than assuming it.

Once a model is developed, confirmatory factor analysis can be used to test whether the model adequately reproduces the observed covariance structure of the observed data. Then, the model can be evaluated for several levels of longitudinal measurement invariance. The degree to which the model exhibits invariance determines the validity of various inferences that might be drawn from it. First, it is important that the pattern of relationships between indicators and constructs is consistent over time, although this is often assumed rather than tested directly (Little & Slegers, 2005). When this is true, it is referred to as *configural* or *pattern invariance*. By itself, configural invariance does not provide sufficient evidence that the construct is stable over time and comparisons between time points may not be meaningful.
Next, the configural invariance model is used as the comparison model for the subsequent level of invariance, which assesses whether the magnitude of the relationships between indicators and constructs is consistent over time. This is indicative of whether the manifest variables are equally good indicators of the latent construct over time. To test this, the loadings between the indicators and the latent construct are constrained to be equal across time. If model fit is not substantially worse than the configural invariance model, then the relationships are consistent, and inferences can be made regarding the latent variances and covariances over time. When this is the case, the model is said to have weak factorial, loading, or metric invariance. If the full metric invariance model (with all loadings constrained to be equal across time) fits worse than the configural model, it is possible to free some of the loadings and potentially establish what is referred to as partial metric invariance, wherein a majority of the loadings are constrained to be equal but at least one pair is freed. Partial metric invariance indicates that generally, most of the relationships between the indicators and the latent construct are consistent.

Next, if metric invariance (or at least partial metric invariance) is established, the full or partial metric invariance model retained in the previous step is used as the baseline for comparison for the next level of invariance, which assesses whether the origin of the relationships between indicators and the latent construct are consistent across time. To test this the intercepts of the relationships between the indicators and the latent construct are constrained to be equal across time. This level is referred to as strong factorial, scalar, or intercept invariance. If intercept invariance is established, latent construct
means can be tested for differences over time. Like metric invariance, it is possible that the model may only exhibit partial intercept invariance.

Lastly, if intercept invariance (or at least partial intercept invariance) is established, the full or partial intercept invariance model retained in the previous step is used as the baseline for comparison for the next level of invariance, which assesses whether all of the differences over time are captured by or attributable to differences in the latent construct over time. This level of invariance is tested by constraining the residual variances to be equal across time and compared to the intercept invariance model. If this level of invariance is established, it is referred to as strict factorial invariance. However, it is uncommon to find strict factorial invariance, because constraining the residual variances to be equal across time assumes that any measurement error is exactly equal across time, and the degree to which this is not the case may result in bias in the estimates of other parameters in the model (Little & Lee, 2014).

If interested in differences among groups as well as across time, it is important to assess measurement invariance between the groups. Like longitudinal measurement invariance, the inferences that can be made about differences in group variances, covariances, and means are based on the level of measurement variance established and are the same as those for longitudinal invariance. However, the meaningfulness of any differences between groups or over time is contingent upon validation of the latent construct.

However, SEM can require a fairly large sample size relative to the parameters being estimated in the model (Wolf, Harrington, Clark, & Miller, 2013). The WeCook
program had up to 30 participants per semester, with six semesters completed (including the pilot semester, which was excluded from these analyses). However, not all program participants consented to the research, which resulted in a moderately small sample size ($n = 126$, Table 1). Consequently, data reduction methods may be necessary. If the scales in the present study are unidimensional and the responses are reliable, a summary score (such as the mean, for items with consistent measurement scales) can be used as a single indicator in the SEM model, thereby reducing the number of parameters estimated.

**Reliability and Validity**

Reliability, or the accuracy or dependability of responses (Cronbach, 1951), is a precondition for validity (Bovaird & Embretson, 2008), while validity is vital in order to draw accurate conclusions from data. There are several types of reliability, such as test-retest reliability, inter-rater reliability, and internal consistency reliability. The latter will be used in the present study to assess how well a series of related items consistently represent an underlying construct (Cronbach, 1951; Tavakol & Dennick, 2011; Zinbarg, Revelle, Yovel, & Li, 2005). Cronbach’s coefficient $\alpha$, a generalized form of the Kuder-Richardson 20, (Cronbach, 1951) can be used to assess internal consistency reliability of sets of items that are unidimensional (i.e., represent only one underlying construct). Unidimensionality can be assessed several ways, but the present study will use principal axis factor analyses to confirm the unidimensionality of the latent construct a set of items is assessing (Lai, Crane, & Cella, 2006; Tavakol & Dennick, 2011). Alternatively, the analyses may provide evidence of scale multidimensionality; if a scale is
multidimensional, McDonald’s \( \omega_h \) would be a more appropriate assessment of internal consistency reliability (Zinbarg et al., 2005).

There are also several types of validity (Bovaird & Embretson, 2008; Cronbach & Meehl, 1955; Whitely, 1983), but this study will focus on predictive validity of the youth self- and parent-report survey measures and the construct validity of the modeled latent factor HPAB. Predictive validity is the degree to which a measure is predictive of a criterion, or ‘gold standard’ measure of the intended construct. Construct validity is sought when there is no clear criterion to predict, but it is desirable to determine what underlying construct(s) account for the results obtained from an assessment (Cronbach & Meehl, 1955).

Validation of a latent construct modeled in SEM can be achieved using a nomological network (or nomothetic span) approach and construct representation (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Whitely, 1983). Cronbach and Meehl’s original view of the nomological network idea was expanded upon by Campbell and Fiske when they introduced convergent and discriminant validity as subtypes of construct validity in their paper on the multitrait-multimethod matrix (Campbell & Fiske, 1959). This method is based on the idea that if the latent variable estimated in the model does indeed measure the intended construct, it should exhibit relationships with other measures (either directly observable or previously validated latent constructs) that are consistent with the theoretical basis of the intended construct. More specifically, convergent validity based on the premise that the modeled construct will be significantly correlated with other variables that should be, based on theory and previous research,
related to the intended construct. Conversely, discriminant validity is the premise that the modeled construct should have little or no relationship with variables which are not related to the intended construct (based on theory and previous research).

**Convergent and discriminant validity and construct representation.**

In the present study, the expected relationships for convergent validity include a positive relationship between the modeled healthy physical activity behavior (HPAB) construct and family income (Delva et al., 2006; Delva et al., 2007; Drenowatz et al., 2010; Fahlman et al., 2015) and a negative relationship with BMI (Delva et al., 2007; Drenowatz et al., 2010). In addition, it is expected that overweight or obese youth, minority youth, and those from lower SES families will have a lower HPAB than their non-overweight or obese, non-minority, and higher SES peers, respectively (Delva et al., 2006; Delva et al., 2007; Drenowatz et al., 2010; Fahlman et al., 2015; Walther et al., 2018). There may also be negative relationships between HPAB and grade and gender, with lower HPAB expected in 5\textsuperscript{th} graders compared to 4\textsuperscript{th} graders (Aaron, Storti, Robertson, Kriska, & LaPorte, 2002; Drenowatz et al., 2010) and lower HPAB in girls compared to boys, but these may be small or non-significant. Conversely, it is expected that there will be no difference between the schools or between the early and later cohorts, although there may be differences by season (Tucker & Gilliland, 2007).

If the data obtained from the survey measures are reliable and exhibit predictive validity with the objectively-measured physical activity data, they may be useful indicators when combined with the objective data in a longitudinal latent difference score model of healthy physical activity behavior. If measurement invariance is established and
the latent construct reflects the expected differences and is unrelated to theoretically independent constructs, the model may be useful in assessing change from the beginning to the end of the program or assessing potential differences in physical activity among groups of interest. If this approach results in a useful indicator of healthy physical activity behavior in diverse, underserved youth, it will be used as an evaluative tool for assessing grant outcomes. In addition, future projects could benefit from planning assessments in a manner that allows for the integration of both subjective and objective measures of physical activity and sedentary habits to get a better estimate of overall healthy physical activity behavior and any potential changes following interventions. If the measures do not provide reliable and valid data, or the data do not contribute unique, meaningful information to the model, it may be necessary to investigate ways to improve the assessments for future evaluative purposes. In either case, establishing more robust tools for the evaluation of current programs will help inform decisions regarding which ones should be continued or expanded and which should be discontinued or altered.

In summary, by partnering with the local Community Learning Centers at two Title I schools (defined as having at least 40% of students from low-income families), the WeCook program is addressing the need for healthy behavior programming focused on dietary and physical activity habits for underserved students. Through evaluation of the physical activity assessment tools currently in use and the potential development of an integrative assessment of healthy physical activity behavior using latent variable modeling this project may lay the groundwork for a component of the final evaluation of WeCook program outcomes.
Present Study

The purposes of the current study were to (a) assess the reliability and validity of youth self-report and parent-report data regarding physical activity and sedentary behavior in youth participating in the WeCook afterschool program and (b) determine whether data from these measures, combined with objective wearable physical activity monitor data, can be used to generate a model of the latent construct HPAB such that it might be a useful outcome measure for program evaluation. The current study was guided by the following research questions:

1. Are the youth and adult survey responses regarding physical activity and sedentary behavior reliable and valid measures of the constructs they are intended to assess?
   a. Do the youth and adult survey responses exhibit internal consistency reliability among related items?
   b. Are the youth and adult survey responses valid indicators of physical activity when compared with the objective measures from the physical activity monitors?

2. Can the latent construct HPAB be modeled using the youth and adult survey responses and physical activity monitor data?
   a. Does the model demonstrate construct validity, as assessed using a nomological network approach?
   b. Can the latent construct be used to generate an overall HPAB score that is sensitive to change and/or differences among groups?
CHAPTER III: METHODOLOGY

Study Design

The WeCook: Food with Fun and Fitness (WeCook) program is a 24-session after-school nutrition, cooking, and physical activity intervention program taking place through the Community Learning Centers (CLCs) at two Title I elementary schools in a Midwestern city as part of an ongoing grant. Title I allocates funding for schools to provide services to children who are failing, or most at risk of failing, to meet challenging state academic standards. For a school to be eligible to use Title I funds for school-wide programs, at least 40% of enrollment must be comprised by students from low-income families (U.S. Department of Education. Office of Elementary and Secondary Education, Office of State Support, 2015) Program sessions are 60 minutes long and occur twice per week (approximately 12 weeks, schedules permitting). In addition, there are three ‘Family Night’ sessions during which the participants prepare foods and parents and family members are invited to attend. Although the grant includes a research component, WeCook is a treatment-only program. This study is primarily a methodological analysis intended to lay the foundation for further research evaluating the outcomes of the WeCook program.

Participants and Recruitment

WeCook is offered to 4th and 5th grade youth who are recruited to participate in the WeCook program in the same manner through which they would normally sign up for other after-school programs through CLCs. Those who sign up to participate in the WeCook program are invited to be a part of the research study but are not required to do so. Youth who do not provide assent or whose parents do not provide consent to
participate in the research study are not excluded from any part of the programming as a result, but any surveys completed by these participants or their parents and any data collected from the Fitbit physical activity monitors are not included in the analyses. Up to 15 youth are allowed to sign up at each site per semester. This study protocol was approved by the University of Nebraska – Lincoln Institutional Review Board (approval #20150715356EP); copies of the youth assent and adult consent forms are found in Appendices C and D, respectively. Youth assent was obtained during the first week of programming and the adult consent form was sent home with the youth to be filled out by his or her parent/guardian. Parents/legal guardians were given the opportunity to ask questions during the Family Night activities. Since youth were not excluded from participating in any part of the program for non-participation in the research, the consent forms did not have to be completed prior to attending or participating in the program. Participants were excluded from analyses if they had participated in the WeCook program in a previous semester.

Program Curriculum

Each week had a central theme with one program day focused on nutrition and cooking and the other focused on physical activity. The nutrition and cooking day was dedicated to teaching food preparation skills and the importance of balanced nutrition using USDA Guidelines while the physical activity day was comprised of interactive games designed to increase physical activity while reiterating themes from the nutrition lessons.
Data Collection Procedures

Data collection occurs at the beginning and end of the 24-session program. In some cases, due to holidays or weather-related issues the sessions of the program may span more than 12 consecutive calendar weeks and occasionally lessons are unavoidably missed due to cancellations or scheduling conflicts.

Survey Instruments.

As reported previously (Walther et al., 2018), CYFAR grants require paper surveys to be administered to all participants at the beginning (PRE) and end (POST) of the program. These surveys include selected CYFAR 4-H Common Measures, which are identified, evaluated and vetted by the CYFARnet Evaluation Team led by the Universities of Arizona and Virginia Tech (University of Minnesota & USDA National Institute of Food and Agriculture, 2018). In addition, date of birth was collected at PRE and height (cm) and weight (kg) were measured at both time points.

Youth survey.

Youth PRE surveys were completed during the first day of the program before any educational instruction was given. POST surveys were completed during the final session. The youth surveys took approximately 15-20 minutes to complete. In addition to 13 questions related to physical activity, the youth survey included items regarding demographics, program participation, nutrition knowledge, and healthy eating and cooking self-efficacy (Appendix A). Of the 13 physical activity items, 11 address physical activity self-efficacy (Saunders et al., 1997), one addresses physical activity knowledge, and one addresses physical activity behavior (Walther et al., 2018). Due to
issues with how the physical activity knowledge question was worded and how responses were coded it was excluded from the analyses.

**Adult survey**

Adult PRE surveys were sent home with the child along with the consent forms and program information. The adult surveys contained items regarding demographics, program participation, parent-child interaction, parent perception of the child’s healthy eating self-efficacy, family eating habits, and the child’s physical activity and sedentary behavior. There were three items focused on physical activity and five items focused on sedentary behaviors (Appendix B).

**Physical Activity Monitors.**

Fitbit Charge (Fitbit, Inc.; San Francisco, CA) wearable physical activity monitors were used in conjunction with the Fitabase® platform (Small Steps Labs, LLC; San Diego, CA) to collect objective physical activity data for two 8-day time periods at the beginning and end of the program, (e.g., Tuesday to Tuesday). Fitabase® is a third-party data-aggregation system that integrates with the Fitbit servers and is designed for managing multiple participant profiles in a secure and confidential manner for purposes such as research. The participants are given the Fitbit during the afterschool hours on the first day of data collection and return it at the same time the following week. Consequently, the first and last days of data collection are incomplete and are excluded from the analyses. Therefore, four weekdays and two weekend days at PRE and at POST are include in the results. The measures collected from the Fitbits that were analyzed in
the present study include the average daily step totals and the average daily floors climbed at PRE and POST (ST1, ST2, FL1 and FL2).

**Data Processing**

BMI (kg/m²) was calculated from height (m) and weight (kg) and used in conjunction with date of birth and gender to categorize students by weight status (non-overweight/obese, overweight, or obese) according to the CDC BMI-for-age charts (Kuczmarski et al., 2002; US Department of Health and Human Services, 2000). Weight status (WS) was recoded into a dichotomous variable such that underweight or normal weight youth were coded as non-overweight or obese (WS = 0) and overweight and obese youth were coded as overweight or obese (WS = 1). Race and ethnicity variables were used to create a dichotomous variable indicating minority status such that white/Caucasian, non-Hispanic youth were coded as non-minority (MIN = 0) and youth who were either non-white, multiracial, and/or were Hispanic were coded as minority (MIN = 1). Parent-reported annual family income (FI) was used to create a dichotomous estimate of socioeconomic status (SES) such that incomes below $50,000 were coded as low SES (SES = 0) and incomes above $50,000 were coded as high SES (SES = 1). Grade was coded such that 4th graders, were the baseline (5TH = 0) and 5th-graders were the comparison group (5TH = 1). Semesters were coded such that spring cohorts were the baseline (FALL = 0) and fall cohorts were the comparison group (FALL = 1). Schools were coded such that the school where the pilot semester was conducted was the baseline (SITE = 0) and the second school was the comparison group (SITE = 1). To create a dichotomous variable (YEAR) for validity testing, the five cohorts included in
these analyses were split into early (first and second program years, YEAR = 0) and late cohorts (third program year, YEAR = 1). Objective physical activity data obtained from the Fitbit® wearable physical activity monitors (step counts [ST] and floors climbed [FL]) were aggregated via the Fitabase® platform, downloaded, and processed in the RStudio IDE v1.1.453 (RStudio, Inc.; Boston, MA) with Microsoft R Open v3.5.0 (Microsoft and R Core Team; Redmond, WA). Days with fewer than 500 steps recorded were excluded from analysis.
**Statistical Analyses**

Statistical analyses were conducted using SPSS v25 (IBM; Armonk, NY), RStudio IDE v1.1.453 (RStudio, Inc.; Boston, MA) with Microsoft R Open v3.5.0 (Microsoft and R Core Team; Redmond, WA), Mplus Version 7 (Muthén & Muthén; Los Angeles, CA), and Microsoft Excel 2016 (Microsoft Inc.; Redmond, WA). Figures were generated using SPSS v25, Mplus Version 7 or an open source scientific plotting software program (Veusz, v. 2.0.1, available at http://home.gna.org/veusz/). For the structural models, parameters were estimated using maximum likelihood (ML) in Mplus and $\chi^2$ difference tests were used for the evaluation of nested models. Because this is an exploratory study assessing potential feasibility, liberal criteria were used to assess model fit, with $RMSEA < 0.08$ (Browne & Cudeck, 1993) and $CFI > 0.90$ (Bentler, 1990).

**Unidimensionality.**

Principal axis factor analyses (SPSS v25, IBM; Armonk, NY), were used to confirm the unidimensionality of several subsets of variables: step counts (ST) and floors climbed (FL) across the six days at PRE and POST, the youth physical activity self-efficacy items, and the adult survey items regarding the child’s physical activity and sedentary habits. Number of factors retained was based on several considerations: visual inspection of the scree plots, magnitude of the first eigenvalue relative to the subsequent ones, interpretability of the factors, the Kaiser criterion (eigenvalues >1), and comparison with the results of a parallel analysis.

**Internal consistency reliability.**

Internal consistency reliability was calculated for the PA self-efficacy items from the youth survey and the PA and SB items from the adult survey using Cronbach’s
coefficient $\alpha$ (Cronbach, 1951). In addition, $\alpha$ if deleted was calculated and assessed. When deemed appropriate, items with higher $\alpha$ if deleted values were removed and the scale comprised by the remaining items was reassessed. When the responses associated with the variable sets exhibited both unidimensionality and reliability the scores were standardized ($z$-score) across the PRE and POST values. Reliability was rated as good, acceptable, questionable, or poor (Nunnally & Bernstein, 1994). Items within a scale were averaged for subsequent analyses.

**Validity of survey responses.**

Predictive validity of the survey responses (average of the youth physical activity self-efficacy items [YSE], youth self-reported days with at least 60 minutes of PA in previous week [YPA], and the average of the adult survey items regarding the child’s physical activity and sedentary habits [AS]) was assessed by testing whether they are significant predictors of the measures from the wearable physical activity monitors (step counts [ST] and floors climbed [FL]).

**Latent healthy physical activity behavior.**

**Longitudinal measurement invariance.**

Healthy physical activity behavior was modeled as a latent construct using the youth survey variables (YPA and YSE), adult survey variable (AS), and objective wearable physical activity monitor variables (ST and FL) as indicators. The configural model fit was assessed separately for PRE and POST measures. Then, the longitudinal configural model was fit and was used as the baseline model for longitudinal metric and
Intercept invariance testing; this model included correlated errors between time points for each individual indicator variable.

**Measurement invariance between groups.**

Once longitudinal invariance was established, the model was tested for group measurement invariance between grades (4TH and 5TH), males and females (M and F), schools (S1 and S2), weight status (NW and OW), minority versus non-minority (MIN and NM), SES (LOW and HIGH), and season (FALL and SPRING).

**Nomological network and contextual validity.**

Validity of the healthy physical activity behavior (HPAB) construct was assessed using a nomological network and contextual approach. Specifically, it was expected that at PRE, there would be a positive relationship between HPAB and family income (FI, an indicator of SES) and a negative relationship between HPAB and BMI. Overweight or obese students (OW) were expected to have a lower HPAB than non-overweight or obese students (NW). It was also expected that there would either be no difference in HPAB students or possibly a small decrease from 4th grade to 5th grade students (Sherar, Esliger, Baxter-Jones, & Tremblay, 2007). Minority students and girls were expected to have lower HPAB although the differences may be small (Trost, Pate, Sallis et al., 2002; Zelener & Schneider, 2016). However, it was not expected that there would be differences between the two schools or between cohorts from year to year, although there may be differences by season (Garcia, Pender, Antonakos, & Ronis, 1998). If the program is effective in altering physical activity habits, an increase in the overall mean HPAB would be expected. If program effectively addresses some of the reasons the target
population (low-income, overweight or obese, minority, and/or female students) is less active than their peers (higher income, healthy weight, white, non-Hispanic students) it would be expected that any relationships between HPAB and those variables would be weakened at POST compared to PRE. These effects were tested using a latent difference score model (Little et al., 2006) after identifying the measurement model and establishing measurement invariance.

After measurement invariance was established across groups and validity was confirmed via the nomological network relationships and construct representation, the latent means were assessed for differences between groups, (e.g., minority versus non-minority).
CHAPTER IV: RESULTS

Tables 1 and 2 contain participant demographics and anthropometric data, respectively. Participants were predominantly female and in 4th grade. Nearly half were white/Caucasian and non-Hispanic and nearly two-thirds were considered low-SES. Over 40% of students were overweight or obese at both PRE and POST.

Table 1.
Participant demographics.

<table>
<thead>
<tr>
<th></th>
<th>mean ± SD</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractional age†</td>
<td>years</td>
<td>10.1 ± 0.6</td>
<td></td>
</tr>
<tr>
<td>Attendance</td>
<td># sessions</td>
<td>20.0 ± 4.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>89.3 ± 16.5</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>n (%)</td>
<td>35 (27.8)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>91 (72.2)</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>n (%)</td>
<td>83 (65.9)</td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td></td>
<td>43 (34.1)</td>
<td></td>
</tr>
<tr>
<td>Season</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring</td>
<td>n (%)</td>
<td>51 (40.5)</td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td></td>
<td>75 (59.5)</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/Caucasian, non-</td>
<td>61 (48.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other race and/or</td>
<td>64 (50.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (0.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Income‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$25,000</td>
<td>36 (28.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$25,000-$50,000</td>
<td>45 (35.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$50,000-$75,000</td>
<td>16 (12.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$75,000-$100,000</td>
<td>10 (7.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;$100,000</td>
<td>5 (4.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>14 (11.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>126 (100)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

†Age was treated as a continuous variable for the purposes of assessing obesity status via the cutoffs from the Centers for Disease Control (CDC) Body Mass Index (BMI) charts for children (Kuczmarski et al., 2002). This slightly elevates the mean age compared to the result computed from traditional, categorical age: (mean ± SD) 9.6 ± 0.6 years. ‡ Family income <$50,000 was categorized as low-SES.

Table 2.
Anthropometric data.

<table>
<thead>
<tr>
<th></th>
<th>Pre mean ± SD</th>
<th>Post mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>cm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>141.1 ± 7.2</td>
<td>142.8 ± 7.2</td>
</tr>
<tr>
<td>Weight</td>
<td>kg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>41.5 ± 11.2</td>
<td>43.2 ± 12.0</td>
</tr>
<tr>
<td>BMI†</td>
<td>kg/m²</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20.6 ± 4.2</td>
<td>21.0 ± 4.7</td>
</tr>
<tr>
<td>Weight Status</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Non-obese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overweight</td>
<td>16 (12.7)</td>
<td>20 (15.9)</td>
</tr>
<tr>
<td>Obese</td>
<td>42 (33.3)</td>
<td>33 (26.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>6 (4.8)</td>
<td>17 (13.5)</td>
</tr>
<tr>
<td>Total</td>
<td>126 (100)</td>
<td>126 (100)</td>
</tr>
</tbody>
</table>

†Body Mass Index (BMI) cutoffs for overweight and obesity in children vary by gender and age in months. The cutoff for underweight is a BMI-for-age below the 5th percentile, overweight is above the 85th percentile, and obese is above the 95th percentile (Kuczmarski et al., 2002).
Unidimensionality

Step counts and floors climbed.

Figure 1 shows the scree plots from the principal axis factor analyses of step counts (ST, A. and B.) and floors climbed (FL, C. and D.) at PRE (A. and C.) and POST (B. and D.). ST and FL were determined to be unidimensional at PRE and POST. For ST and FL, the variance explained by the first factor was 32.9% and 30.2% at PRE and 42.0% and 31.5% at POST, respectively.

Figure 1. Scree plots from principal axis factor analyses of step counts (ST, A. and B.) and floors climbed (FL, C. and D.) at PRE (A. and C.) and POST (B. and D.). Open
circles are the eigenvalues from the initial, unrotated solution, black circles are the results of a parallel analysis, the dotted line indicates the Kaiser criterion (eigenvalues > 1.0), and the dashed line indicates the subjective cut point based on visual inspection.

**Youth physical activity self-efficacy items.**

Figure 2 shows the scree plots from the principal axis factor analyses of the youth physical activity self-efficacy responses (YSE) at PRE (A.) and POST (B.). At both time points the responses were determined to be unidimensional. The variance explained by the first factor was 37.1% at PRE and 49.9% at POST.

![Figure 2](image)

**Figure 2.** Scree plots from principal axis factor analyses of the youth physical activity self-efficacy (YSE) at PRE (A.) and POST (B.). Open circles are the eigenvalues from the initial, unrotated solution, black circles are the results of a parallel analysis, the dotted line indicates the Kaiser criterion (eigenvalues > 1.0), and the dashed line indicates the subjective cut point based on visual inspection.
Adult survey items regarding child physical activity and sedentary habits.

Figure 3 shows the scree plots from the principal axis factor analyses of the adult survey responses (AS) to the items regarding the child’s physical activity and sedentary behavior at PRE (A.) and POST (B.). At both timepoints the responses determined to be unidimensional based on visual inspection of the scree plots and the magnitude of the first eigenvalue relative to the subsequent ones, despite the lower variance explained at PRE. The variance explained by the first factor was 24.7% at PRE and 40.0% at POST.

![Scree plots from principal axis factor analyses of the adult survey (AS) items regarding the child's physical activity and sedentary habits items at PRE (A.) and POST (B.). Open circles are the eigenvalues from the initial, unrotated solution, black circles are the results of a parallel analysis, the dotted line indicates the Kaiser criterion (eigenvalues > 1.0), and the dashed line indicates the subjective cut point based on visual inspection.](image)

**Figure 3.** Scree plots from principal axis factor analyses of the adult survey (AS) items regarding the child's physical activity and sedentary habits items at PRE (A.) and POST (B.). Open circles are the eigenvalues from the initial, unrotated solution, black circles are the results of a parallel analysis, the dotted line indicates the Kaiser criterion (eigenvalues > 1.0), and the dashed line indicates the subjective cut point based on visual inspection.

**Reliability and Validity of Survey Measures**

Internal consistency reliability (Cronbach’s coefficient $\alpha$) for the youth survey physical activity self-efficacy items, the adult survey items regarding the youth’s physical
activity and sedentary habits, and step counts and floors climbed are shown in Tables 3-5. Where removing an item would both increase the reliability of the scale and was theoretically sound, items were removed. Specifically, the adult survey items regarding the child’s non-school-related computer time and cell phone/texting time were removed. At both PRE and POST, more than 50% of parents indicated that their child did not spend any time on these two activities in the preceding seven days. These activities may be minimal or not applicable for this sample of 4th and 5th grade students at a Title I school (e.g., limited access to a computer outside of school and/or no personal cell phone). For ST, reliability was acceptable at PRE and good at POST and for FL reliability was questionable at PRE and POST. For YSE reliability was good at PRE and POST but for AS reliability was poor at PRE and acceptable at POST. Table 6 contains the correlations between the survey measures and the measures from the wearable physical activity monitors.

YPA was the survey measure most consistently related to the wearable physical activity monitor data and was significantly correlated with average ST at PRE ($\tau = 0.233, p = 0.001$) and POST ($\tau = 0.178, p = 0.024$), and average FL at PRE ($\tau = 0.209, p = 0.018$) but not at POST ($\tau = -0.034, p = 0.737$). YSE had the highest internal consistency reliability of the survey measures ($\alpha = 0.829$-$0.897$). However, the YSE average score was only significantly correlated with ST at POST ($r = 0.239, p = 0.022$), and was not

\[\text{\footnotesize \ref{footnote}}\]

\footnote{Kendall’s $\tau$ was calculated for youth self-reported physical activity (YPA) instead of Pearson’s $r$ because the survey responses were from a single question on an ordinal scale (0-4).}
correlated with ST at PRE \(r = 0.070, p = 0.461\), FL at PRE \(r = 0.055, p = 0.640\), or FL at POST \(r = 0.190, p = 0.154\). After removing the computer and phone/text items from AS, the internal consistency reliability of the responses remained poor at PRE \(\alpha = 0.545\) although it improved at POST \(\alpha = 0.729\). However, the averaged AS scores did not correlate with any of the objective physical activity measures.

Table 3.
Internal consistency reliability of step count and floors climbed data.

<table>
<thead>
<tr>
<th></th>
<th>Step counts</th>
<th>Floors climbed†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>n responses</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td>Raw (\alpha)</td>
<td>0.735</td>
<td>0.804</td>
</tr>
<tr>
<td>Standardized (\alpha)</td>
<td>0.740</td>
<td>0.805</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weekday 1</td>
<td>0.691</td>
<td>0.813</td>
</tr>
<tr>
<td>weekday 2</td>
<td>0.700</td>
<td>0.772</td>
</tr>
<tr>
<td>weekday 3</td>
<td>0.721</td>
<td>0.764</td>
</tr>
<tr>
<td>weekday 4</td>
<td>0.686</td>
<td>0.767</td>
</tr>
<tr>
<td>weekend day 1</td>
<td>0.711</td>
<td>0.766</td>
</tr>
<tr>
<td>weekend day 2</td>
<td>0.673</td>
<td>0.753</td>
</tr>
</tbody>
</table>

† The lower sample size for floors climbed is a result of higher resolution data being overwritten if monitors were not synced soon enough after the data collection time period. This generally occurred because the monitor was not returned on time (e.g., monitor was forgotten at home or the child was absent the day they were collected).

Table 4.
Internal consistency reliability of youth physical activity self-efficacy responses.

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>n responses</td>
<td>119</td>
<td>109</td>
</tr>
<tr>
<td>Raw (\alpha)</td>
<td>0.829</td>
<td>0.897</td>
</tr>
<tr>
<td>Standardized (\alpha)</td>
<td>0.829</td>
<td>0.898</td>
</tr>
<tr>
<td>Item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.819</td>
<td>0.891</td>
</tr>
<tr>
<td>b</td>
<td>0.821</td>
<td>0.894</td>
</tr>
<tr>
<td>c</td>
<td>0.818</td>
<td>0.887</td>
</tr>
<tr>
<td>d</td>
<td>0.814</td>
<td>0.888</td>
</tr>
<tr>
<td>e</td>
<td>0.813</td>
<td>0.881</td>
</tr>
<tr>
<td>f</td>
<td>0.810</td>
<td>0.891</td>
</tr>
<tr>
<td>g</td>
<td>0.809</td>
<td>0.886</td>
</tr>
<tr>
<td>h</td>
<td>0.811</td>
<td>0.894</td>
</tr>
<tr>
<td>i</td>
<td>0.819</td>
<td>0.886</td>
</tr>
<tr>
<td>j</td>
<td>0.811</td>
<td>0.884</td>
</tr>
<tr>
<td>k</td>
<td>0.815</td>
<td>0.884</td>
</tr>
</tbody>
</table>

See Appendix A (Youth Survey) for item information.

Table 5.
Internal consistency reliability of adult survey responses regarding physical activity and sedentary time.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Computer removed</th>
<th>Phone/text removed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>n responses</td>
<td>103</td>
<td>72</td>
<td>103</td>
</tr>
<tr>
<td>Raw (\alpha)</td>
<td>0.521</td>
<td>0.761</td>
<td>0.538</td>
</tr>
<tr>
<td>Standardized (\alpha)</td>
<td>0.526</td>
<td>0.780</td>
<td>0.556</td>
</tr>
</tbody>
</table>
Table 6.

Validity of survey measures.

<table>
<thead>
<tr>
<th>Item</th>
<th>Step counts</th>
<th>Floors climbed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRE</td>
<td></td>
</tr>
<tr>
<td>Youth self-reported PA (YPA)†</td>
<td>0.233</td>
<td>0.001</td>
</tr>
<tr>
<td>Youth PA self-efficacy (YSE)</td>
<td>0.070</td>
<td>0.461</td>
</tr>
<tr>
<td>Adult survey PA/SB habits (AS)</td>
<td>0.058</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>POST</td>
<td></td>
</tr>
<tr>
<td>Youth self-reported PA (YPA)†</td>
<td>0.178</td>
<td>0.024</td>
</tr>
<tr>
<td>Youth PA self-efficacy (YSE)</td>
<td>0.239</td>
<td>0.022</td>
</tr>
<tr>
<td>Adult survey PA/SB habits (AS)</td>
<td>0.128</td>
<td>0.300</td>
</tr>
</tbody>
</table>

†Kendall’s τ was calculated for YPA instead of Pearson’s r because the survey responses were from a single question on an ordinal scale (0-4).

Latent Healthy Physical Activity Behavior

Model specification and invariance testing.

Longitudinal measurement invariance.

Table 7 shows the model fit indices for the configural models fit to the PRE data (Model 1), the POST data (Model 2), both PRE and POST with correlated errors (Model 3, Figure 4). In addition, Table 7 includes the model comparisons for metric (Model 4) and intercept (Model 5) invariance and the final model (Model 6), which includes
estimates of the latent means. The model exhibited both full metric and full intercept invariance.

**Figure 4.** Path diagram for confirmatory factor analysis. HPAB₁ and HPAB₂ are estimates of the latent construct of healthy physical activity behavior at PRE and POST, respectively. ST₁ and ST₂ are average daily step counts and FL₁, and FL₂ are average daily floors climbed. YPA₁ and YPA₂ are the youth responses from the self-reported physical activity item; YSE₁ and YSE₂ are mean scores from the physical activity self-efficacy scale (Table 3). AS₁ and AS₂ are mean scores from the adult survey items regarding the child’s physical activity and sedentary habits (Table 4). λ₁-λ₅ are the loadings fixed to be invariant across time. Not shown are the intercepts for the manifest variables, which were also constrained to be equal across time.
Table 7.

Configural models and model comparisons for longitudinal invariance testing.

<table>
<thead>
<tr>
<th>Model</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4 PRE-POST</th>
<th>5 PRE-POST</th>
<th>6 PRE-POST</th>
<th>metric, intercept, + latent means</th>
<th>( \Delta \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRE</td>
<td>configural</td>
<td>configural</td>
<td>configural</td>
<td>metric</td>
<td>( \Delta \chi^2 )</td>
<td>metric + intercept</td>
<td>( \Delta \chi^2 )</td>
<td></td>
</tr>
<tr>
<td>POST</td>
<td>3.08</td>
<td>10.41</td>
<td>40.24</td>
<td>46.44</td>
<td>6.20</td>
<td>56.39</td>
<td>9.95</td>
<td>51.82</td>
</tr>
<tr>
<td>( \chi^2 )</td>
<td>5</td>
<td>5</td>
<td>29</td>
<td>34</td>
<td>5</td>
<td>39</td>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>( df )</td>
<td>0.69</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
<td>0.287</td>
<td>0.04</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>( p )</td>
<td>0.00</td>
<td>0.10</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
</tbody>
</table>

**Multiple group measurement invariance.**

Table 8 shows the model comparisons used to test for differences between the final model (Model 6) and models using grade (5TH), gender (FEM), school (SITE), weight status (WS), minority status (MIN), SES, season (FALL), and grant year (YEAR) as binary indicators of group membership (Models 7-13). Full metric and intercept invariance were confirmed for each of these variables (i.e., model fit was not significantly worse than the baseline model [Model 6] when loadings and intercepts were constrained to be equal across groups, see Table 8 for statistical tests and fit indices).
Table 8.
Model comparisons for measurement invariance across groups. Baseline model for comparisons is Model 6 (Table 7).

<table>
<thead>
<tr>
<th>Model</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>5TH $\chi^2$</td>
<td>61.07</td>
<td>58.76</td>
<td>67.29</td>
<td>15.47</td>
<td>60.00</td>
<td>8.18</td>
<td>61.18</td>
<td>57.88</td>
</tr>
<tr>
<td>SITE $\chi^2$</td>
<td>58.76</td>
<td>67.29</td>
<td>70.78</td>
<td>55.18</td>
<td>45.8</td>
<td>15.47</td>
<td>61.18</td>
<td>9.36</td>
</tr>
<tr>
<td>WS $\chi^2$</td>
<td>61.07</td>
<td>58.76</td>
<td>67.29</td>
<td>15.47</td>
<td>60.00</td>
<td>8.18</td>
<td>61.18</td>
<td>9.36</td>
</tr>
<tr>
<td>SES $\chi^2$</td>
<td>61.07</td>
<td>58.76</td>
<td>67.29</td>
<td>15.47</td>
<td>60.00</td>
<td>8.18</td>
<td>61.18</td>
<td>9.36</td>
</tr>
<tr>
<td>YEAR $\chi^2$</td>
<td>58.76</td>
<td>67.29</td>
<td>70.78</td>
<td>55.18</td>
<td>45.8</td>
<td>15.47</td>
<td>61.18</td>
<td>9.36</td>
</tr>
<tr>
<td>$df$</td>
<td>45</td>
<td>8</td>
<td>45</td>
<td>8</td>
<td>45</td>
<td>8</td>
<td>45</td>
<td>8</td>
</tr>
<tr>
<td>$p$</td>
<td>0.06</td>
<td>0.32</td>
<td>0.08</td>
<td>0.54</td>
<td>0.02</td>
<td>0.39</td>
<td>0.39</td>
<td>0.05</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>CFI</td>
<td>0.89</td>
<td>0.90</td>
<td>0.85</td>
<td>0.88</td>
<td>0.85</td>
<td>0.88</td>
<td>0.89</td>
<td>0.91</td>
</tr>
<tr>
<td>SRMR</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
</tr>
</tbody>
</table>

**Nomological network and contextual validity and latent difference scores.**

Table 9 shows the path coefficients between the group indicators and HPAB at PRE and POST. Table 10 shows selected estimates for the latent difference score models, including the mean difference at PRE and interaction effect from PRE to POST for each variable for the nomological network and construct representation validity assessment.

Table 9.
Path coefficients for convergent and discriminant validity.

<table>
<thead>
<tr>
<th></th>
<th>PRE</th>
<th></th>
<th>POST</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>0.213</td>
<td>0.124</td>
<td>0.217</td>
<td>weak</td>
</tr>
<tr>
<td>FI</td>
<td>0.073</td>
<td>0.052</td>
<td>0.176</td>
<td>weak</td>
</tr>
<tr>
<td>WS</td>
<td>-0.054</td>
<td>0.134</td>
<td>-0.053</td>
<td>none</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.018</td>
<td>0.015</td>
<td>-0.150</td>
<td>weak</td>
</tr>
<tr>
<td>MIN</td>
<td>-0.073</td>
<td>0.108</td>
<td>-0.082</td>
<td>none</td>
</tr>
<tr>
<td>5TH</td>
<td>-0.176</td>
<td>0.109</td>
<td>-0.191</td>
<td>weak</td>
</tr>
<tr>
<td>FEM</td>
<td>0.010</td>
<td>0.118</td>
<td>0.010</td>
<td>none</td>
</tr>
<tr>
<td>FALL</td>
<td>0.128</td>
<td>0.112</td>
<td>0.137</td>
<td>weak</td>
</tr>
<tr>
<td>SITE</td>
<td>-0.164</td>
<td>0.103</td>
<td>-0.192</td>
<td>weak</td>
</tr>
<tr>
<td>YEAR</td>
<td>-0.159</td>
<td>0.114</td>
<td>-0.168</td>
<td>weak</td>
</tr>
</tbody>
</table>

* No paths were significant ($\alpha = 0.05$). Strength of relationships based on Cohen (1977) standards (none: 0.0-0.1, weak: 0.1-0.3, mild: 0.3-0.5, strong: >0.5).
Table 10

Selected parameters from latent difference score models.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.112 0.052 *</td>
<td>-0.517 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>-</td>
</tr>
<tr>
<td>LD (mean)</td>
<td>0.235 0.096 *</td>
<td>0.298 weak</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt;</td>
<td>-</td>
</tr>
<tr>
<td>SES</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.092 0.050</td>
<td>-0.481 mild</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; 0.213 0.124 0.217 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.153 0.078 *</td>
<td>0.343 mild</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; -0.073 0.145 -0.073 none</td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.098 0.052</td>
<td>-0.472 mild</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; 0.073 0.052 0.176 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.219 0.100 *</td>
<td>0.469 mild</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; -0.071 0.062 -0.168 weak</td>
<td></td>
</tr>
<tr>
<td>BS</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.109 0.050 *</td>
<td>-0.529 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.07 0.11 -0.084 none</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.147 0.089</td>
<td>0.311 mild</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.003 0.13 0.003 none</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.105 0.050 *</td>
<td>-0.515 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.02 0.01 -0.145 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>-0.116 0.312</td>
<td>-0.247 weak</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.013 0.13 0.114 weak</td>
<td></td>
</tr>
<tr>
<td>MIN</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.111 0.051 *</td>
<td>-0.518 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.073 0.108 -0.082 none</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.150 0.094</td>
<td>0.311 mild</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; -0.013 0.132 -0.013 none</td>
<td></td>
</tr>
<tr>
<td>5TH</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.102 0.048 *</td>
<td>-0.514 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.176 0.109 -0.191 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.086 0.076</td>
<td>0.184 weak</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.169 0.133 0.170 weak</td>
<td></td>
</tr>
<tr>
<td>FEM</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.107 0.050 *</td>
<td>-0.531 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; 0.010 0.118 0.010 none</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>-0.023 0.122</td>
<td>-0.049 none</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.224 0.144 0.213 weak</td>
<td></td>
</tr>
<tr>
<td>FALL</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.110 0.053 *</td>
<td>-0.530 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; 0.128 0.112 0.137 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.268 0.084 *</td>
<td>0.552 strong</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; -0.310 0.138 * -0.314 mild</td>
<td></td>
</tr>
<tr>
<td>SITE</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.097 0.044 *</td>
<td>-0.507 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.164 0.103 -0.192 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.052 0.087</td>
<td>0.111 weak</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.183 0.125 0.195 weak</td>
<td></td>
</tr>
<tr>
<td>YEAR</td>
<td>HPAB&lt;sub&gt;1&lt;/sub&gt; with LD</td>
<td>-0.117 0.053 *</td>
<td>-0.515 strong</td>
<td>γ&lt;sub&gt;1&lt;/sub&gt; -0.159 0.114 -0.168 weak</td>
</tr>
<tr>
<td>LD (intercept)</td>
<td>0.118 0.087</td>
<td>0.236 weak</td>
<td>γ&lt;sub&gt;2&lt;/sub&gt; 0.060 0.135 0.058 none</td>
<td></td>
</tr>
</tbody>
</table>

* Significant (α = 0.05). Strength of relationships based on Cohen (1977) standards (none: 0.0-0.1, weak: 0.1-0.3, mild: 0.3-0.5, strong: >0.5). LD is the latent difference score (change from PRE to POST) when the group covariate is zero, γ<sub>1</sub> is HPAB<sub>1</sub> regressed on the covariate (group difference at PRE), and γ<sub>2</sub> is LD regressed on the covariate (interaction between group membership and the change from PRE to POST).
**Figure 5.** Latent difference score model. HPAB\(_1\) and HPAB\(_2\) are estimates of the latent construct of healthy physical activity behavior at PRE and POST, respectively, and LD is the latent difference. ST\(_1\) and ST\(_2\) are average daily step counts and FL\(_1\), and FL\(_2\) are average daily floors climbed. YPA\(_1\) and YPA\(_2\) are the responses from the self-reported physical activity item; YSE\(_1\) and YSE\(_2\) are mean scores from the physical activity self-efficacy scale (Table 3). AS\(_1\) and AS\(_2\) are mean scores from the adult survey items regarding the child’s physical activity and sedentary habits (Table 4).

**Validity assessment**

The baseline latent difference score model showed a strong and statistically significant association between HPAB\(_1\) and the latent difference (LD) with a standardized regression coefficient (\(\beta\)) of -0.517 (\(p = 0.030\)). The mean for LD, (i.e., the change in HPAB from PRE to POST) reflected a significant increase in HPAB (mean = 0.235, SE = 0.096, \(p = 0.015\)) after the program.
**Socioeconomic status and family income.**

At PRE, there were weak positive associations between HPAB and high SES and family income (FI), although these were not significant (SES: $\gamma = 0.217, p = 0.085$; FI: $\gamma = 0.176, p = 0.154$). At POST there was a weak positive, non-significant relationship between HPAB and high SES ($\gamma = 0.140, p = 0.312$), but no relationship with FI ($\gamma = 0.005, p = 0.971$). In the latent difference score model, adding SES weakened the relationship between HPAB\textsubscript{1} and LD from strong to mild, and non-significant ($\gamma = -0.481, p = 0.068$), the mean for LD remained significant ($mean = 0.153, SE = 0.078, p = 0.049$), there was no group mean difference at PRE ($0.217, p = 0.085$), and no interaction for SES and LD ($-0.073, p = 0.616$). Adding FI to the LDS model also decreased the association between HPAB\textsubscript{1} and LD from strong to mild and non-significant ($\beta = -0.472, p = 0.061$), but the mean for LD remained significant ($mean = 0.219, SE = 0.100, p = 0.029$), and there was a weak positive, non-significant association between HPAB\textsubscript{1} and FI ($\gamma = 0.176, p = 0.154$), and a weak negative, non-significant interaction between LD and FI ($\gamma = 0.168, p = 0.250$). The relationship between SES and HPAB at PRE was weaker than expected but may be confounded by the relatively simplistic method used for determining SES (self-reported family income < $50,000$). Future studies may wish to consider other factors involved in SES, such as parent education. In general, these results only somewhat agree with the expectations of positive relationships between HPAB and SES and FI at PRE, and a weakening of the relationships at POST.
**Weight status and BMI.**

At PRE, there was no relationship between weight status (WS₁) and HPAB₁ (γ = -0.053, p = 0.688) but there was a weak negative non-significant association between BMI₁ and HPAB₁ (γ = -0.150, p = 0.238). At POST there was also no relationship between WS₂ and HPAB₂ (γ = 0.044, p = 0.741) and no relationship between BMI₂ and HPAB₂ (γ = 0.043, p = 0.745). In the latent difference score model, adding WS₁ to the model did not change the relationship between HPAB₁ and LD which remained strong and significant (β = -0.529, p = 0.031), the mean for LD was no longer significant (mean = 0.147, SE = 0.089, p = 0.099), and there was no association between HPAB₁ and WS₁ (γ = -0.084, p = 0.502), and no interaction between LD and WS₁ (γ = 0.003, p = 0.983). Adding BMI₁ to the model did not change the relationship between HPAB₁ and LD which remained strong and significant (β = -0.515, p = 0.037), the mean for LD was no longer significant (mean = -0.116, SE = 0.312, p = 0.110), and there was a weak negative non-significant association between HPAB₁ and BMI₁ (γ = -0.145, p = 0.231), and a weak positive non-significant interaction between LD and BMI₁ (γ = 0.114, p = 0.389). These results only somewhat agree with the expectation of negative relationships between HPAB and WS and BMI at PRE and a weakening of the relationships at POST.

**Minority.**

There was no relationship between minority status and HPAB₁ (γ = -0.082, p = 0.498) or HPAB₂ (γ = -0.094, p = 0.486). Adding minority to the latent difference score model did not change the relationship between HPAB₁ and LD which remained strong and significant (β = -0.518, p = 0.030), the mean for LD was no longer significant (mean = 0.147, SE = 0.089, p = 0.099), and there was no association between HPAB₁ and WS₁ (γ = -0.084, p = 0.502), and no interaction between LD and WS₁ (γ = 0.003, p = 0.983).
= 0.150, SE = 0.094, p = 0.110), and there was no mean difference between minority and non-minority students at HPAB\textsubscript{1} (\(\gamma = -0.082, p = 0.498\)), and no interaction between LD and minority (\(\gamma = -0.013, p = 0.922\)). These results do not generally agree with the expected small difference between minority and non-minority students at PRE and a weakening at POST. However, it is consistent with some studies which have not found differences in physical activity between minority and non-minority children (Whitt-Glover et al., 2009).

**Gender.**

There was no relationship between gender and HPAB\textsubscript{1} (\(\gamma = -0.010, p = 0.931\)), but there was a weak positive non-significant relationship between gender and HPAB\textsubscript{2} (\(\gamma = 0.235, p = 0.079\)). Adding gender to the latent difference score model did not change the relationship between HPAB\textsubscript{1} and LD which remained strong and significant (\(\beta = -0.531, p = 0.033\)), the mean for LD was no longer significant (\textit{mean} = -0.023, \(SE = 0.122, p = 0.850\)), there was no mean difference between male and female students at HPAB\textsubscript{1} (\(\gamma = 0.010, p = 0.931\)), but there was a weak positive non-significant interaction between LD and being female (\(\gamma = -0.213, p = 0.119\)). These results do not generally agree with the expected lower PA levels in girls compared to boys at PRE, but somewhat support the idea that the intervention might increase PA levels in girls. While much of the literature shows gender differences in activity levels (Magnússon, Sveinsson, Arngrímsson, & Johannsson, 2008; Trost et al., 1996), some of those effects were small (Trost et al., 2002) or could be attributed to maturational differences (Sherar et al., 2007; Thompson, Baxter-Jones, Mirwald, & Bailey, 2003). It is possible that the small sample size and high
proportion of girls in the study resulted in insufficient power to detect a gender difference at PRE. Alternatively, it is possible that more active boys simply chose to participate in other afterschool programs offered at the CLC, such as sports. This would be consistent with research showing higher sport participation in boys compared to girls and its relationship to PA levels (Barnett, Van Beurden, Morgan, Brooks, & Beard, 2009; Beets & Pitetti, 2005; Magnússon et al., 2008).

**Grade.**

Grade had a weak negative non-significant association with HPAB₁ (γ = -0.191, p = 0.104) but no association with HPAB₂ (γ = -0.008, p = 0.952). Adding grade to the latent difference score model did not change the relationship between HPAB₁ and LD which remained strong and significant (β = -0.514, p = 0.034), the mean for LD was no longer significant (mean= 0.086, SE = 0.076, p = 0.257), there was weak negative non-significant association between being in 5th grade and HPAB₁ (γ = -0.191, p = 0.931), and a weak positive non-significant interaction between being in 5th grade and LD (γ = 0.170, p = 0.206). These results somewhat support the expected lower PA in 5th grade students compared to 4th and the weakening of the relationship at POST by providing some evidence for a difference at PRE and an increase in PA among 5th graders to reduce the difference between grades at POST.

**Season.**

Season had a weak positive non-significant relationship with HPAB₁ (γ = 0.137, p = 0.252) and a weak negative non-significant association with HPAB₂ (γ = -0.198, p = 0.146). Adding season to the latent difference score model did not change the relationship
between HPAB\textsubscript{1} and LD which remained strong and significant ($\beta = -0.530, p = 0.037$), however (unlike the other latent difference models with covariates added) the mean for LD remained significant (mean = 0.268, SE = 0.084, $p = 0.002$), there was weak positive non-significant association between being in a fall cohort and HPAB\textsubscript{1} ($\gamma = 0.137, p = 0.252$), and a significant mild negative interaction between being in a fall cohort and LD ($\gamma = -0.314, p = 0.025$). These results partially support the expected difference between seasons and in particular agree with research showing a general decline in PA levels in youth during fall and an increase in the spring (cite). These results suggest that researchers and program evaluators should take into consideration the seasonality of PA levels in youth when establishing baseline PA levels or assessing change.

\textit{School.}

School had a weak negative, non-significant association with HPAB\textsubscript{1} ($\gamma = -0.192, p = 0.113$) and no association with HPAB\textsubscript{2} ($\gamma = 0.022, p = 0.870$). Adding school to the latent difference score model did not change the relationship between HPAB\textsubscript{1} and LD which remained strong and significant ($\beta = -0.507, p = 0.028$), the mean for LD was no longer significant (mean = 0.052, SE = 0.087, $p = 0.551$), there was weak negative non-significant association between being at the second site and HPAB\textsubscript{1} ($\gamma = -0.192, p = 0.113$), and a weak positive non-significant interaction between being at the second site and LD ($\gamma = 0.195, p = 0.154$). These results may either suggest that students at the two sites may differ in PA levels prior to the intervention for unknown reasons and that the program may have a differential impact on the students at the two schools. Alternatively, these results may also indicate that the model is detecting spurious
differences between the two schools, but more information is needed to determine whether this may be the case.

**Year.**

Grant year had a weak negative, non-significant association with HPAB₁ (\( \gamma = -0.168, p = 0.163 \)) and a weak negative non-significant association with HPAB₂ (\( \gamma = -0.102, p = 0.433 \)). Adding school to the latent difference score model did not change the relationship between HPAB₁ and LD which remained strong and significant (\( \beta = -0.515, p = 0.028 \)), the mean for LD was no longer significant (\( \text{mean} = 0.118, SE = 0.087, p = 0.175 \)), there was weak negative non-significant association between being in a later grant year and HPAB₁ (\( \gamma = -0.168, p = 0.163 \)), and no interaction between grant year and LD (\( \gamma = 0.058, p = 0.659 \)). These results only partially support the expected lack of relationship between grant year and PA levels in independent cohorts of students. While not significant, the weak relationships results suggest that between the early cohorts (program years 1 and 2) and the more recent cohorts (program year 3), there may have been a decrease in baseline PA levels among students but that the program may have counteracted it to some extent. Although this agrees with the national trend of decreases in PA levels among youth over time, it was not expected to detect a difference with a relatively small sample over a relatively short period of time. Conversely, these results may indicate that the model is detecting a spurious difference between grant years for unknown reasons.
CHAPTER V: DISCUSSION

Reliability and Validity of Survey Measures

The reliability and validity of the responses from the survey measures varied substantially. The survey responses from the youth physical activity self-efficacy scale reflected the expected unidimensionality (Figure 2) and good internal consistency reliability (Table 4, standardized $\alpha = 0.829$ at PRE and $0.898$ at POST), consistent with previous research (Motl et al., 2000; Saunders et al., 1997). However, contrary to previous findings (Fahlman et al., 2015; Martin, McCaughtry, & Shen, 2008; Martin & McCaughtry, 2008a; Martin & McCaughtry, 2008b; Martin et al., 2011; Trost et al., 2002), the physical activity self-efficacy score was not consistently a significant predictor of the objective physical activity measures (Table 6). This may indicate that in the present study, low physical activity levels may be more strongly influenced by other factors not related to self-efficacy. Alternatively, the small sample size in this study may have resulted in insufficient power to detect the relatively small relationship (cite).

Conversely, the responses from the youth self-reported physical activity item were significantly related to three of the four objective physical activity measures from the wearable physical activity monitors (Table 4) although the relationships are weaker than in some previous research (Brown, Hume, & Chinapaw, 2009; Chinapaw et al., 2010). Although younger children may struggle with accurate recall of physical activity behaviors, using a multiple-item scale may provide a more accurate assessment (Brown et al., 2009; Chinapaw et al., 2010; Terwee et al., 2010). The results from Zelener and Schneider (2016) however, highlight the importance of clear instructions and the inclusion of items asking about different types of physical activity (e.g., low, moderate,
and vigorous) and settings of physical activity (e.g.; PE class, sports, after-school) to prevent double-counting or non-reporting of different aspects of physical activity behavior. Further research is needed to establish a standardized questionnaire with acceptable reliability and validity for self-reported PA in 4th and 5th grade children (Brown et al., 2009; Chinapaw et al., 2010).

The parent survey responses regarding the child’s activity and sedentary behavior were generally unidimensional (Figure 3) and had poor to acceptable internal consistency reliability (Table 4, standardized α = 0.577 at PRE and 0.747 at POST). It is unclear why the reliability is lower at PRE, but it is possible that familiarizing the parents with the questionnaire may improve the reliability. Specifically, the factor analysis shows a notably larger first eigenvalue at POST compared to PRE. It is possible that the prior to completing the survey at PRE the parent respondents were not particularly attentive to their child’s physical activity and sedentary behaviors and that completing the PRE survey resulted in a priming effect, such that the parent became more observant of the amount of time the child spent engaging in physical activity or sedentary behaviors between the PRE and POST surveys. Consistent with previous findings (Corder et al., 2012), the scores from the parent survey exhibited poor predictive validity relative to the objective physical activity measures (Table 6). While parent estimation of child PA is poor in general, it is worse in parents of children not meeting the PA guidelines (Kesten et al., 2015). The poor agreement between parent-reported PA and objective measures may be exacerbated in single-parent families or families in which both parents work and do not generally observe much of the child’s free time physical activity and sedentary
behaviors. Given that the children in the present study attend Title I schools these issues may be more prevalent than in the general population. However, there was a number of missing parent surveys at both time points, which reduced the sample size available for testing the relationships. Future studies might benefit from including incentives or reminders to increase parent survey completion.

**Latent Healthy Physical Activity Behavior Models**

The results of the present study indicate that modeling healthy physical activity behavior as a latent construct using objective wearable PA monitor data together with youth self- and parent-report measures as indicators is feasible. The resulting model had configural, metric, and intercept invariance across time and between several groups of interest (Tables 7 and 8). In addition, there is some evidence for the construct validity of the model and potential sensitivity to change.

Overall, the results from the present study suggest that modeling healthy physical activity behavior using objective wearable physical activity monitor data in conjunction with survey measures could be a potentially useful method for assessing changes in youth physical activity and sedentary behaviors following an intervention. However, there are some caveats, as these results also point to some methodological and measurement issues which should be addressed in future studies and which may improve the reliability, validity, and sensitivity of the resulting model.

First, these results are in spite of the aforementioned issues of less-than-ideal reliability and predictive validity of some of the survey responses. Specifically, inclusion of a multiple-item PA questionnaire with clear instructions may improve the reliability
and validity of the self-reported PA. Future research should also investigate whether different questions for the parent survey that may be more applicable to the population being studied and/or familiarization of the parents with the parent survey items might help improve the reliability and validity of the parent assessment of the child’s PA.

Second, despite the data reduction methods used in the present study, the sample size was still relatively small for the model being estimated. Future studies may benefit from additional data reduction (such as combining the step counts and floors climbed data via factor analysis) and/or use of a larger sample size to increase power.

Although the latent difference score models provide some weak evidence for the expected differences at PRE, it also provides some weak evidence that the intervention may have impacted those relationships. However, the model detected a significant interaction effect for season, which is a factor that known to influence PA levels in youth, suggesting that the model may indeed be sensitive to differences but that the magnitude of some of the anticipated relationships was too small to be detected with the sample size in the present study, and because of the weak and/or non-significant relationships, there was little room for those relationships to be attenuated.

**Limitations and Future Directions**

This study was exploratory in nature and while it sheds some light on the feasibility of using multimethod SEM procedures in research and program evaluation settings involving physical activity assessment that can collect both survey measures and objective physical activity data, there are several limitations. Several of these stem from
the fact that this study was an exploratory and methodologically focused investigation using data from a pre-existing, non-experimental intervention.

First, several of the evaluative tools used in the WeCook program were mandated by the funding agency. Even with some issues raised by the reliability and validity analyses of the survey responses, the measures in use could not be changed (while it might be possible to include additional measures, it is unlikely as the program is in its final year and the potential burden of the addition likely outweighs the benefit). However, this highlights the need for further research investigating the reliability and validity of responses associated with some of the CYFAR 4-H Common Measures in use and the need for development of new ones. In particular, there appear to be limited options for some construct areas and age groups (i.e., physical activity in elementary school children).

There were also several limitations related to the adult surveys. Because only the PRE survey included demographic questions, it is possible that one parent may have responded at PRE and the other parent may have responded at POST, potentially confounding any differences. Second, a number of parents did not speak English, and the surveys were translated into several other languages, including Spanish and Arabic. However, literacy, in English or other languages may have been a limitation for parents in the diverse, low-income population in this study. In addition, future studies may also wish to investigate the potential for a priming effect of the PRE survey on the parent responses. Improving the reliability and validity of the parent-report physical activity and
sedentary behavior responses may be as simple as including an extra survey at or near the beginning of the program.

Future studies should also evaluate these methods in the context of an experimental intervention. The validity assessment of the latent construct in the present study relies on what amounts to correlational relationships. Without a control group or a way to approximate one, it is unclear whether the small effects observed were due to the program itself or outside factors. Ideally, future studies should aim to increase the sample size, even if the added participants only complete the survey measures and do not wear the wearable physical activity monitors.

Lastly, future studies might also investigate the application of the multimethod and planned missing design approach to other areas of nutrition and exercise research. Specifically, a next step might be to investigate the reliability and validity of the survey responses related to cooking and nutrition knowledge and behaviors from students in the WeCook program. Other potential areas include those suggested by Graham and colleagues (Graham et al., 2006), such as integrating dietary logs and in-depth interviews regarding dietary intake or BMI (cheap and widely reported) and hydrostatic weighing (more valid, but expensive assessment of adiposity).

Practical Applications

Overall, this study lays the groundwork for the potential implementation of an integrative assessment of physical activity in youth using multimethod indicators to model healthy physical activity behavior as a latent construct. To extend this method and provide an outcome score that could be used in outcome evaluations, the factor scores
from the latent difference model could be saved and analyzed for group comparisons or assessment of change over time. The structural model used in the present study allows for differential weighting of the indicators (e.g., the loadings for the objective measures are larger than the loadings for the adult survey responses). However, proper data treatment is necessary to prevent unintentional obfuscation of the results. Although z-score standardization was used in the present study, care was taken to avoid some of the issues that can occur when standardizing multiple variables over repeated measures. For example, when standardizing PRE and POST measures it is necessary to stack the data in a ‘long’ format (i.e., person-period) and standardize across all values; if PRE values are standardized separate from POST, they will both have a mean of zero and any mean differences across time will be obscured. Moeller (2015) identifies several potential pitfalls associated with standardization of data in longitudinal studies and identifies alternative methods.

Conclusions

Collectively, these results provide some evidence that the model may indeed be reflective of the latent construct of healthy physical activity behavior, but a larger sample size and possibly improvements in the quality of survey measures may be necessary to develop a more reliable, valid, and sensitive model. In addition, utilization of strategies to minimize missing data may also improve the model and its potential utility. Further research is needed to conclusively determine whether this approach is useful for program evaluation and research.
REFERENCES

Longitudinal study of the number and choice of leisure time physical activities from
mid to late adolescence: Implications for school curricula and community recreation
programs. Archives of Pediatrics & Adolescent Medicine, 156(11), 1075-1080.

Childhood motor skill proficiency as a predictor of adolescent physical activity.
Journal of Adolescent Health, 44(3), 252-259.


Bulletin, 107(2), 238.

Review of Psychology, 53(1), 605-634.

Sage Handbook of Social Research Methods, 269-289.


Heath, K. R. (2018). Assessing adult perceptions, outcomes and associations with youth nutrition and activity outcomes of participants in the WeCook: Fun with food and fitness program (M.S.).


Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach’s α, Revelle’s β, and McDonald’s ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123-133.
APPENDIX A: YOUTH SURVEY

WeCook Youth Survey

Participant ID # ___________________________ Date ____________

Please DO NOT write your name on this survey.
The answers you give will be kept private. This survey is voluntary.

DIRECTIONS: Please select the appropriate response for each item below.

1. I am: _______ Male _______ Female

2. How old are you? _______

3. What grade are you in school? _______

4. What is your ethnicity? (Select one)
   _______ Hispanic or Latino _______ Not Hispanic or Latino

5. What is your race? (Select one or more)
   _______ Asian _______ American Indian or Alaska Native
   _______ Black or African American _______ Native Hawaiian or Other Pacific Islander
   _______ White

6. Is your parent(s) involved in the military including the Guard or Reserve?
   _______ Yes _______ No

7. If yes, please specify
   _______ Air Force _______ Army _______ Guard
   _______ Marine Corps _______ Navy _______ Reserve

8. How many sessions of this club or activity have you participated in?
   _______

9. About how many hours per week do you participate in this club or activity?
   _______ Less than 1 hour _______ 6-7 hours
   _______ 1 hour _______ 8-9 hours
   _______ 2-3 hours _______ 10 or more hours
   _______ 4-5 hours

10. How long have you participated in 4-H?
    _______ Less than 1 year _______ 6-7 years
     _______ 1 year _______ 8-9 years
     _______ 2-3 years _______ 10 or more years
     _______ 4-5 years _______ Does not apply to me
WeCook Youth Survey

11. How long have you participated in any in-school activities like sports, student government, drama or dance, academic clubs, pep clubs, band or symphony?
   - Less than 1 year
   - 1 year
   - 2-3 years
   - 4-5 years
   - 6-7 years
   - 8-9 years
   - 10 or more years
   - Does not apply to me

12. How long have you participated in any other out-of-school activities like Boy Scouts, Girl Scouts, YMCA, Girls Inc., Junior Achievement, or youth groups at church, synagogue, or mosques?
   - Less than 1 year
   - 1 year
   - 2-3 years
   - 4-5 years
   - 6-7 years
   - 8-9 years
   - 10 or more years
   - Does not apply to me

DIRECTIONS: The following questions ask about your eating habits and how hard you think it would be for you to eat more of some foods and eat less of other foods. How hard would it be for you to...

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>0 Not hard at all</th>
<th>1 A little hard</th>
<th>2 Very hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Eat fruit for an after school snack?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Eat vegetables for an after school snack?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Choose water instead of soda pop or Kool-Aid when you are thirsty?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Drink 1% or skim milk instead of 2% or whole milk?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Choose a small instead of a large order of French fries?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Eat smaller servings of high fat foods like French fries, chips, snack cakes, cookies, or ice cream?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Eat a low-fat snack like pretzels instead of chips?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Drink less soda pop?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Drink less Kool-Aid?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**WeCook Youth Survey**

**DIRECTIONS:** The following questions ask you about being active. Being active can mean playing a sport, playing outside with friends, or doing an activity like riding a bike. Choose the answer which best shows how you feel about physical activity.

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>0 Not at all like me</th>
<th>1</th>
<th>2 A lot like me</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I can ask my friends to be active with me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I can ask my parents or another adult to do active things with me.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I have the skills I need to be active.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I can be active most days after school.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I can be active no matter how busy my day is.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I can be active no matter how tired I may feel.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I can be active even if it is hot or cold outside.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I can be active even if I have a lot of homework.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I can be active after school even if I could watch TV or play video games instead.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I can be active even if I have to stay at home.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I can be active even when I’d rather be doing something else.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WeCook Youth Survey

DIRECTIONS: The next 2 questions ask about physical activity. Place an “x” in the ONE box that represents your answer.

1. How often are you physically active for at least 60 minutes per day or more? (This includes activities such as exercise, sports, running, walking, dancing, etc.)
   - [ ] 7 days per week
   - [ ] 1-2 days per week
   - [ ] 3-4 days per week
   - [ ] 0 days per week

2. Why is physical activity good for kids?
   - [ ] Helps keep you from getting sick
   - [ ] Helps you pay attention in school
   - [ ] Builds healthy bones and muscles to keep you strong
   - [ ] Gives you energy
   - [ ] All of the above

DIRECTIONS: Circle the answer that best applies to you.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I eat vegetables...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>2. I eat fruit...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>3. I choose healthy snacks...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>4. I eat breakfast...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
</tbody>
</table>

DIRECTIONS: Place an “x” in the ONE box that represents your answer.

1. Yesterday, how many times did you eat French fries or chips? Chips are potato chips, tortilla chips, corn chips, or other snack chips.
   - [ ] None
   - [ ] 1-2 times
   - [ ] 3-4 times
   - [ ] 5 or more times

2. Yesterday, how many times did you eat doughnuts, cookies, brownies, cakes or candy?
   - [ ] None
   - [ ] 1-2 times
   - [ ] 3-4 times
   - [ ] 5 or more times
WeCook Youth Survey

3. Yesterday, how many times did you drink any regular sodas or soft drinks, punch, sports drinks, or other fruit-flavored drinks? (Do not count 100% juice or diet drinks)
   - None
   - 1-2 times
   - 3-4 times
   - 5 or more times

4. How many total cups of fruit and vegetables combined should you eat each day?
   - Less than 2 cups
   - At least 2 cups
   - At least 3 cups
   - At least 4 cups

5. How do you feel about cooking?
   - I really like to cook.
   - I kind of like to cook.
   - I don't like to cook.
   - I really don't like to cook.
   - I'm not sure if I like to cook.

6. How do you feel about making foods with your family?
   - I really like to make food with my family.
   - I kind of like to make food with my family.
   - I don't like to make food with my family.
   - I really don't like to make food with my family.
   - I'm not sure if I like to make food with my family.

7. Which of the following statements best describes you?
   - I can follow a recipe by myself.
   - I can follow a recipe with help from someone else.
   - I have never followed a recipe, and I do not feel I could make it by myself.

DIRECTIONS: Place an “x” in ALL boxes that represent ALL answers you think are correct.

1. Which of the following would be a healthy choice for a snack? Check ALL that apply.
   - Fruit and yogurt
   - Sports drink and cheese puffs
   - Whole grain crackers and cheese
   - Celery and peanut butter
   - Fruit juice and potato chips

2. Why is breakfast important? Check ALL that apply.
   - Helps you learn
   - Gives you energy
   - Makes you weaker
   - Helps keep you from getting sick
   - Helps you think and concentrate
Appendix B: ADULT SURVEY

WeCook Adult Survey
☐ Pre ☐ Post

Participant ID # ___________________________ Date ___________

Please DO NOT write your name on this survey. The answers you give will be kept private. This survey is voluntary.

DIRECTIONS: Please select the appropriate response for each item below.

1. I am a:
   ______ Male    ______ Female

2. How old are you?
   ______

3. What is your ethnicity? (Select one)
   ______ Hispanic or Latino    ______ Not Hispanic or Latino

4. What is your race? (Select one or more)
   ______ Asian    ______ American Indian or Alaska Native
   ______ Black or African American    ______ Native Hawaiian or Other Pacific Islander
   ______ White

5. What is your current level of employment?
   ______ Employed Full-time    ______ Unemployed, Stay-at-home parent
   ______ Employed Part-time    ______ Unemployed, Student
   ______ Unemployed    ______ Retired

6. What is your highest level of education completed?
   ______ Less than high school    ______ Associate’s degree
   ______ High school diploma/GED    ______ Bachelor’s degree
   ______ Post-secondary technical training    ______ Graduate degree
   ______ Some college

7. Have you ever served in the military including the Guard or Reserve?
   ______ Yes    ______ No

8. If yes, please specify
   ______ Air Force    ______ Army    ______ Guard
   ______ Marine Corps    ______ Navy    ______ Reserve

9. Are you currently active?
   ______ Yes    ______ No

10. Your family’s annual income:
    ______ <$25,000    ______ $25,000 - $50,000
     ______ $50,001 - $75,000    ______ >$100,000
WeCook Adult Survey

11. I feed my family using the following food resources: (Check all that apply)
   _____ WIC
   _____ Food Bank/Pantry
   _____ Family/Friends
   _____ Purchase at the grocery store
   _____ Purchase at the convenience store
   _____ SNAP/Food Stamps
   _____ Free/Reduced School Lunch
   _____ Commodity
   _____ Soup Kitchen/Church

12. How many children/adolescents under age 18 are in your household?  
   
13. How many people (including adults and children/adolescents) are in your household?  
   
14. How many sessions of this club or activity have you participated in?  
   
15. About how many hours per week do you participate in this club or activity?
   _____ Less than 1 hour
   _____ 1-2 hours
   _____ 2-3 hours
   _____ 3-4 hours
   _____ 4-5 hours
   _____ 5-6 hours
   _____ 6-7 hours
   _____ 7-8 hours
   _____ 8-9 hours
   _____ 9-10 hours
   _____ 10 or more hours

16. How long have you participated in 4-H?
   _____ Less than 1 year
   _____ 1 year
   _____ 1-2 years
   _____ 2-3 years
   _____ 3-4 years
   _____ 4-5 years
   _____ 5-6 years
   _____ 6-7 years
   _____ 7-8 years
   _____ 8-9 years
   _____ 9-10 years
   _____ 10 or more years
   _____ Does not apply to me

17. Are you involved in any other community/volunteer activities?
   _____ Yes
   _____ No

18. If yes, how many other activities are you involved in?  
   
2
WeCook Adult Survey

DIRECTIONS: Please select the appropriate response for each item below.

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>My child and I have warm, intimate moments together</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I encourage my child to talk about his/her troubles.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>I joke and play with my child.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>I make sure my child knows that I appreciate what he/she tries to accomplish.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I encourage my child to wonder and think about life.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>I feel that a child should have time to daydream, think, and even lose sometimes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>I express my affection by hugging, kissing, and holding my child.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>I talk it over and reason with my child when he/she misbehaves.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>I find it interesting and educational to be with my child for long periods.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>I encourage my child to be curious, to explore, and to question things.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I find some of my greatest satisfaction in my child.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>When I am angry with my child, I let him/her know about it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>I respect my child’s opinion and encourage him/her to express it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>I feel that a child should be given comfort and understanding when he/she is scared or upset.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>I am outgoing and relaxed with my child.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>I trust my child to behave as he/she should, even when I am not with him/her.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>I believe in praising a child when he/she is good and think it gets better results than punishing him/her when he/she is bad.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>I usually take into account my child’s preference when making plans for the family.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# WeCook Adult Survey

**DIRECTIONS:** The following questions ask about your child's eating habits and how hard you think it would be for him/her to eat more of some foods and eat less of other foods. How hard would it be for your child to...

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>0 Not hard at all</th>
<th>1 A little hard</th>
<th>2 Very hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fat fruit for an after school snack?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Fat vegetables for an after school snack?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Choose water instead of soda pop or Kool-Aid when you are thirsty?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Drink 1% or skim milk instead of 2% or whole milk?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Choose a small instead of a large order of French fries?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Eat smaller servings of high-fat foods like French fries, chips, snack cakes, cookies, or ice cream?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Eat a low-fat snack like pretzels instead of chips?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Drink less soda pop?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Drink less Kool-Aid?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DIRECTIONS:** Circle the answer that best applies to your family.

<table>
<thead>
<tr>
<th>#</th>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>We eat vegetables...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>2</td>
<td>We eat fruit...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>3</td>
<td>We choose healthy snacks...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
<tr>
<td>4</td>
<td>We eat breakfast...</td>
<td>Never or almost never</td>
<td>Some days</td>
<td>Most days</td>
<td>Every day</td>
</tr>
</tbody>
</table>
WeCook Adult Survey

DIRECTIONS: Place an “x” in the ONE box that represents your answer.

1. Do you think having frequent family meals would encourage your child to eat more healthy foods such as fruits and vegetables?
   - Yes
   - No

2. How sure are you that you could prepare a healthy meal at home?
   - Very sure
   - Somewhat sure
   - Not sure at all

3. How often do you plan family meals at home?
   - 7 days per week
   - 5-6 days per week
   - 1-2 days per week
   - 0 days per week
   - 3-4 days per week

DIRECTIONS: The following questions ask about your child’s overall activity outside of school. This would include structured exercise or sport activities, as well as playing with friends, dancing or doing work/horse. Answer the questions based on your child’s physical activity outside of school in the last 7 days:

1. Activity before School: How many days before school (between 6:00-8:00 am) did your child do some sort of physical activity?
   - 0 days
   - 1 day
   - 2 days
   - 3 days
   - 4 to 5 days

2. Activity after School: How many school afternoons/evenings (between 4:00-5:00pm) did your child do some sort of physical activity?
   - 0 days
   - 1 day
   - 2 days
   - 3 days
   - 4 to 5 days

3. Activity on Weekends: How much physical activity did your child do last weekend? (This could include exercise, work/horse, family outings, sports, dance, or playing with friends)
   - No activity (0 minutes each day)
   - Small amount of activity (1 to 30 minutes each day)
   - Small to Moderate amount of activity (31 to 60 minutes each day)
   - Moderate to Large amount of activity (1 to 2 hours each day)
   - Large amount of activity (more than 2 hours each day)
WeCook Adult Survey

DIRECTIONS: The following questions ask about your child’s time spent resting and sitting. Answer these questions based on the time your child spent sitting in the past 7 days.

4. TV Time: How much time did your child spend watching TV outside of school time?
   □ Didn't watch TV at all
   □ Watched less than 1 hour per day
   □ Watched 1 to 2 hours per day
   □ Watched 2 to 3 hours per day
   □ Watched more than 3 hours per day

5. Video Game Time: How much time did your child spend playing video games outside of school time?
   □ Didn't play at all
   □ Played less than 1 hour per day
   □ Played 1 to 2 hours per day
   □ Played 2 to 3 hours per day
   □ Played more than 3 hours per day

6. Computer Time: How much time did your child spend using computers outside of school? (This does not include homework time but does include Facebook, as well as time spent surfing the internet, instant messaging, and playing computer games)
   □ Didn't use the computer at all
   □ Used the computer less than 1 hour per day
   □ Used the computer 1 to 2 hours per day
   □ Used the computer 2 to 3 hours per day
   □ Used the computer more than 3 hours per day

7. Phone/Text Time: How much time did your child spend using a cell phone outside of school? (This includes time spent talking or texting)
   □ Didn't use a cell phone at all
   □ Used a cell phone less than 1 hour per day
   □ Used a cell phone 1 to 2 hours per day
   □ Used a cell phone 2 to 3 hours per day
   □ Used a cell phone more than 3 hours per day

8. Overall Sedentary Habits: Which of the following best describes your child’s overall sedentary habits at home this past week?
   □ Spent almost no free time sitting
   □ Spent a little time sitting during free time
   □ Spent a moderate amount of time sitting during free time
   □ Spent a lot of time sitting during free time
   □ Spent almost all free time sitting